Use of animal based measures for the assessment of dairy cow welfare ANIBAM

Bodil Højlund Nielsen, Alessandra Angelucci, Alessandra Scalvenzi, Björn Forkman, Francesca Fusi, Frank Tuyttens, Hans Houe, Harry Blokhuis, Jan Tind Sørensen, Janne Rothmann, et al.

EXTERNAL SCIENTIFIC REPORT

Use of animal based measures for the assessment of dairy cow welfare ANIBAM

Bodil Højlund Nielsen[1], Alessandra Angelucci[2], Alessandra Scalvenzi[2], Björn Forkman[3], Francesca Fusi[2], Frank Tuyttens[4], Hans Houe[3], Harry Blokhuis[5], Jan Tind Sørensen[1], Janne Rothmann[1], Lindsay Matthews[6], Luc Mounier[7,8], Luigi Bertocchi[2], Marie-Madeleine Richard[7,8], Matteo Donati[2], Per Peetz Nielsen[5], Romolo Salini[6], Sophie de Graaf[4], Sophie Hild[7,8], Stefano Messori[6], Søren Saxmose Nielsen[3], Valentina Lorenzi[2], Xavier Boivin[7,8] & Peter T. Thomsen[1]

1: Aarhus University, Department of Animal Science, P. O. Box 50, DK-8830 Tjele, Denmark; 2: Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini" – Centro di Referenza Nazionale per il Benessere Animale, Via Bianchi 9, 25124 Brescia, Italy; 3: University of Copenhagen, Department of large Animal Sciences, Faculty of Health and Medical Science, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark; 4: Institute for Agricultural and Fisheries Research, Animal Sciences Unit, Farm Animal Welfare & Behaviour Research Group, Scheldeweg 68, 9090 Melle, Belgium; 5: Swedish University of Agricultural Sciences, Department of Animal Environment and Health, Uppsala, Sweden; 6: Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Campo Boario 64100, Teramo (TE), Italy; 7: Université de Lyon, VetAgro Sup, UMR1213 Herbivores F-69280 Marcy L'Etoile, France; 8: INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France.

ABSTRACT

The overall aim of the project was to evaluate the use of routinely collected animal based measures (ABMs) for an evaluation of the overall animal welfare in dairy cow herds. ABMs being able to detect worst adverse effects in relation to animal welfare were identified based on the existing literature and expert opinion. The validity and robustness of these ABMs were evaluated and cow mortality, somatic cell count and lameness were selected for further study. A number of factors of variation were selected using expert opinion and used in a model to collate routinely collected data from Italy, Belgium and Denmark on selected ABMs. The routinely collected data was uploaded to the Data Collection Framework platform at EFSA and the data management in this process was evaluated. Five research datasets from Italy, Belgium, Denmark and France including information on ABMs as well as a measure of 'overall animal welfare' at herd level were analysed to evaluate the association between the ABMs (individually or in combination) and overall welfare. The measure of 'overall animal welfare' were not the same for all datasets. Except from the Italian data, the association between the ABMs and the different overall welfare measures were generally weak. Likewise, combining more than one ABM only improved the prediction of the overall welfare in the Italian dataset. Analyses of the other datasets could not confirm this finding. Finally, suggestions for future recordings of ABMs not routinely collected at the moment were given with a special focus on lameness. In conclusion, the relationship between selected ABMs and overall welfare at the herd level is complex and still not sufficiently studied. Therefore, a system using routinely collected ABMs to predict the overall welfare at herd level in dairy herds does not seem realistic based on the results from the present study.

Any enquiries related to this output should be addressed to AHAW@efsa.europa.eu

Available online: www.efsa.europa.eu/publications

KEY WORDS

Dairy cow; animal based measure; animal welfare

DISCLAIMER

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Question No EFSA-Q-2012-00724.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

SUMMARY

The European Food Safety Authority (EFSA) aims to establish a practical and validated basis for data collection of animal based measures (ABMs) for farmed animal species and consequent quantitative risk assessment of the welfare of target populations. Through a series of six integrated objectives this project evaluated the possibility to use routinely collected ABMs to predict the overall welfare status at herd level in dairy cow herds.

In Objective 1, a list of adverse effects was identified based on the results presented in previous EFSA reports. In a questionnaire survey, an expert panel was asked to rate the overall impact of the adverse effects considering severity as well as herd prevalence. Based on this, 'mortality – unassisted', 'mortality – euthanised', 'foot disorders', 'leg injuries', 'behavioural disruption – flooring', 'behavioural disruption – rest', 'behavioural disruption – feeding' (cubicles only) and 'exhaustion' were identified as the worst adverse effects (WAEs). A comprehensive list of ABMs that could potentially be used to detect the WAEs was then identified from the EFSA reports by selecting all ABMs reported to be associated with the WAEs. Another expert panel comprising researchers, producer organisation representatives, retailers, practising veterinarians, competent authorities, and NGOs from nine EU countries was then asked to provide information regarding whether these ABMs were routinely collected in their country. Finally, the resulting list of ABMs already collected in the field was subjected to an expert elicitation procedure where experts were asked to identify the linkages between the ABMs and the WAEs and also any linkages between the WAEs. Using this information together with the survey responses on the availability of ABMs routinely collected in the field and suitability for detecting WAEs, the consortium agreed to recommend the following ABMs for further investigations: 'Numbers of deaths (unassisted and euthanised)', 'evidence of mastitis' (somatic cell count (SCC)), 'numbers of foot lesions', 'measures of lameness', and 'numbers of leg lesions/swellings'.

In Objective 2, the validity and robustness of the ABMs selected in Objective 1 were evaluated. The evaluation was performed through a) a systematic literature review of validity and robustness of the ABMs, b) a description of databases available to the members of the project consortium including information on relevant ABMs and c) analyses of sensitivity and specificity of the ABMs in detecting the WAEs in these databases. In order to calculate the sensitivity and specificity, cut-offs were defined for each ABM and WAE. The description of the ABMs and the definition of the cut-offs were based on discussions among consortium members combined with information achieved from the description of the available data. Emphasis was put on feasibility in terms of data availability in the field. Three ABMs were recommended for further investigations: 1) 'Number of deaths' as a direct measure of 'mortality'. Though it was not significantly correlated to any of the other WAEs in the data analysis, it was recommended as useful because data on mortality was deemed as both valid and robust in the literature review and because data on mortality was routinely collected in most EU member states. 2) SCC was recommended as a measure as it was associated with important welfare measures such as overall welfare, mortality and lameness. In the data analysis, an association between SCC and mortality was found. 3) 'Measures of lameness'. In the literature, lameness is widely used as a measure of painful foot lesions in cows and it was deemed as being sensitive towards the most painful foot disorders. However, the validation of lameness scoring is difficult due to the lack of a gold standard for measuring pain in cows. Also, data on lameness might not currently be readily available in the field. Nonetheless, it was deemed an important ABM as it was correlated to the WAEs 'mortality, overall' and 'leg lesions'. The remaining ABMs had problems with robustness and were not recommended.

In Objective 3, a limited list of epidemiological parameters (called 'factors of variation') for the ABMs selected in Objective 2 was proposed. First, the main factors of variation associated with ABMs outside the range of acceptable welfare were identified in the literature. Next, a limited number of these factors of variation were selected by partners in the consortium based on 1) the association of the factors with the ABMs from Objective 2, 2) the feasibility of the collection and possibility of keeping information updated in the field and 3) their capacity to characterise a population. The factors parity, housing system, floor type,

days in milk, access to pasture, milk production, herd size, breed, geographical region and organic dairy production were identified as being the most important to collect. They were chosen for their relevance as risk factors/factors of variation for mortality, mastitis/elevated SCC and lameness, their feasibility for routine field collection and their capacity to characterise populations. Of these factors, housing system, floor type, access to pasture and the ABM lameness were not routinely collected in any of the countries included in this study. Collection of these parameters in the future may have the potential to improve the possibilities of an epidemiological surveillance of the welfare of dairy cows.

In Objective 4, a data model was developed to collate routinely collected data on the ABMs selected in Objective 2 and the factors of variation selected in Objective 3 from national or regional databases in different countries represented in the project consortium. A pilot test on this data model was performed using data from data providers from Italy, Denmark and Belgium. From this it was concluded that data from these countries were not collected in a uniform manner. The ABM lameness and the factors of variation housing, flooring, bedding, access to pasture and stocking density were not present in any of the provided datasets. Furthermore, data that were collected in different countries sometimes differed substantially in form. Consequently, several transformations were needed in order to merge the datasets. Regarding the ABMs, the annual mortality rate was the easiest of the three ABMs to merge: only a minor transformation was needed for the Italian database. SCC was harder to collate because three different types of SCC (bulk milk SCC, herd SCC (proportion of cows with high SCC) and individual SCC) were present in the three datasets provided. At the moment, bulk milk SCC seems most feasible to collect and collate. Three factors of variation needed transformation: region, breed and production type. The creation of breed code relevant across the EU was recommended. The overall conclusion drawn from Objective 4 was that the establishment of a Europe-wide database representing all partner countries was not possible at this point in time without having to perform many transformations and without accepting a lot of missing data.

In Objective 5, the usefulness of the ABMs selected in Objective 2 in combination with the factors of variation selected in Objective 3 for evaluation of the overall welfare in dairy cow herds was evaluated. Analyses were performed using five research datasets from four countries represented in the consortium. As no gold standard for the assessment of the overall welfare status existed, the ABMs and the factors of variation were tested against the overall welfare measures (13 measures in total) present in the research datasets. For each of these overall welfare measures three dichotomised outcomes (poorer versus better welfare) were constructed by using the median, the P25/75 and the P10/90 as thresholds. All three ABMs were identified in all the datasets, whereas seven out of the ten factors of variation were identified in at least one of the datasets. Days in milk and parity were not identified in any of the data. The factor geographical region was identified in the Italian and the Danish data but it was omitted from the analyses because the biological meaningfulness of the factor (as it was defined in the datasets) was deemed low. For each outcome, logistic regression models were used to evaluate the association of the ABMs (alone and in combination) with the dichotomised welfare score. Also, the effects of including the factors of variation in these models were investigated. Finally, conditional inference tree models were built in order to identify whether alternative data structures and relations between variables could be identified. The models and the predictive values of these were compared by BIC values and by constructing ROC curves and calculating the area under the curve (AUC) of these. Except from the Italian data (from IZSLER/CReNBA), the association between the ABMs and the overall welfare outcomes were generally low and the ROC AUCs were rarely significantly larger than 0.50. Equally, the combination of the ABMs seemed beneficial in the IZSLER/CReNBA data but results from the other datasets could not confirm this finding. When significant associations between the ABMs and the risk of classifying as a poorer welfare herd were found, results confirmed the expectation from the expert opinions in previous objectives: Increased levels of mortality, SCC and lameness were associated with poorer welfare. Where mortality and SCC were both associated with the poorer welfare outcome in five out of the thirteen different welfare measures, lameness seemed to be slightly more sensitive as it was found significant in seven out of the thirteen measures.

Finally, based on the results from the previous objectives suggestions were made regarding fine tuning of the approach of using routinely collected ABMs in the assessment of overall welfare status. First and foremost, results from Objective 4 and 5 indicate that the definition of the ABMs as well as the factors of variation and the way they are recorded need to be subjected to careful investigation and consideration in order to standardise these measures between countries for a possible future use of ABMs for evaluation of dairy cow welfare to be successful. Generally, it was recommended that the mortality and SCC should be recorded as aggregations over time in order to make the ABMs less sensitive to variation in point estimates. Also, substantial variation in the mortality rate between countries was evident. Therefore, the definition of one common threshold defining acceptable versus unacceptable levels of mortality seemed unrealistic. Instead, thresholds could be defined specifically for each country or alternatively, they could be data driven.

Based on the results from the other objectives, the final Objective 6 focussed on lameness and stocking density. These were ABMs/factors of variation that were not already collected on a routine basis but nevertheless considered important for an overall welfare assessment. The procedure for assessing lameness in the Welfare Quality® protocol could be used as a procedure for assessing lameness on a regular basis. Also, sample size considerations and recommendations were presented regarding assessment of the general level of lameness in a region (e.g. a country) as well as regarding the assessment of lameness at herd level. Stocking density was rated as highly relevant for dairy cow welfare in the expert elicitations in Objective 3. It is typically assessed by using a variety of different resource based measures. However, the usefulness of stocking density as a welfare indicator were deemed low because it seemed to be closely related to management routines and thus a measure of e.g. number of cows per feeding place could be interpreted very differently in different settings.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

An animal based measure (ABM) is a response of an animal - or an effect on an animal - used to assess its welfare. An animal based measure can be taken directly on the animal or indirectly and includes the use of animal records. It can result from a specific event, e.g. an injury, or be the cumulative outcome of many days, weeks or months, e.g. body condition.

The use of animal based measures (ABMs) to assess animal welfare has been the focus of several research projects over the past five years, and ABMs are now included in various schemes (e.g. Welfare Quality®) used on the field in order to evaluate the welfare status of animals. While assessments previously relied mainly on resource-based parameters, ABMs aim to measure the welfare status of the animal, including the effect of resource- and management-based factors.

Animal based measures have gradually been introduced into EU animal welfare legislation with the Directive on chickens kept for meat production (Council Directive 2007/43/EC[1]) and the Regulation on the killing of animals (Council Regulation EC No 1099/2009[2]). The EU Strategy for the protection and welfare of animals (2012-2015[3]) envisages a new EU legislative framework for animal welfare including the use of scientifically validated animal welfare outcome-based indicators to complement prescriptive requirements.

The European Commission requested EFSA to produce scientific opinions on the use of ABMs to assess the welfare of different farm animals. The AHAW Panel of EFSA has subsequently adopted and published scientific opinions for dairy cows[4], pigs[5] in 2011, and broilers[6] in 2012. These scientific opinions were prepared by a thorough scientific review and update of previous risk assessments, following which the AHAW Panel identified 1) how animal based measures could be used to ensure the fulfilment of the recommendations of EFSA scientific opinions on animal welfare, 2) how existing assessment protocols cover the main hazards identified in EFSA scientific opinions (and vice-versa), and 3) which relevant animal welfare issues cannot be assessed using animal based measures and what kind of alternative solutions are available to improve the situation. Last, the Panel identified the main factors, in the various husbandry systems, having been scientifically proven to have negative effects on the welfare of animals and to what extent these negative effects can be or not prevented through management.

In the course of preparing the scientific opinion on dairy cows, EFSA also contracted scientific studies to analyse the methodology applicable to the validation of ABMs. The results of these studies were published in two external scientific reports[7, 8].

Building on its experience, the EFSA AHAW Panel has also published a statement which provides guidance on the use of animal based measures to assess the welfare of animals[9]. The document particularly highlights that carefully selected combinations of ABMs would allow assessing and benchmarking the welfare of target populations of farmed animals in the EU. To achieve this goal, the Panel stresses the need to select “fit for purpose” and validated measures and to promote their systematic and harmonised collection, contributing to the establishment of a database for a quantitative risk assessment of animal welfare.

Bearing this in mind, in July 2012, EFSA organized a technical meeting with stakeholders and interested parties[10], aiming at reviewing ABMs that are already collected by different actors along the food chain (farmers, industry, retailers, etc.) and from different sources (such as, official controls, industry internal controls, farmers’ records, NGOs, scientists, etc.) and discussing whether such ABMs are considered to be robust, valid and feasible to collect. The meeting also gave an opportunity to exchange views and experience on main welfare issues (in dairy cows, pigs, and broilers) and the suitability of ABMs in these species.

Against this background, EFSA calls for a proof of concept on the use of animal based measures to assess the welfare of animals, preferably focused on one single species, and based on a pilot project involving several EU Member States.

The work of the AHAW Panel has been most extensive on dairy cows and this species is considered to be a good model for the pilot project. Also, several databases hosting information on dairy cows already exist and they could be considered as potential valuable source of ABMs data.

1 Council Directive 2007/43/EC, of 28 June 2007 laying down minimum rules for the protection of chickens kept for meat production. OJ L 182, 12.07.2007, p. 19-28.
2 Council Regulation (EC) No 1099/2009 of 24 September 2009 on the protection of animals at the time of killing. OJ L 303, 18.11.2009, p. 1-30.
3 European Commission, Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the European Union Strategy for the Protection and Welfare of Animals 2012-2015COM(2012) 6 final/2.
4 http://www.efsa.europa.eu/en/efsajournal/pub/2554.htm
5 http://www.efsa.europa.eu/en/efsajournal/pub/2512.htm
6 http://www.efsa.europa.eu/en/efsajournal/pub/2774.htm
7 http://www.efsa.europa.eu/en/supporting/pub/171e.htm
8 http://www.efsa.europa.eu/en/efsajournal/pub/253e.htm
9 http://www.efsa.europa.eu/en/efsajournal/pub/2767.htm
10 http://www.efsa.europa.eu/en/events/event/120704c.htm

TERMS OF REFERENCE AS PROVIDED BY EFSA

This contract/grant was awarded by EFSA to:

Aarhus University, Department of Animal Science, P. O. Box 50, DK-8830 Tjele, Denmark (co-ordinator)

Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Via Campo Boario, 64100 Teramo, Italy

Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini", 9 Via Antonio Bianchi, 25124 Brescia, Italy

Swedish University of Agricultural Science, PO Box 7070, SE 750 07 Uppsala, Sweden

Institute for Agricultural and Fisheries Research and it's Own Capital, 96 Burg. Van Gansberghelaan, 9820 Merelbeke, Belgium

Institut National de la Recherche Agronomique, Theix, F-63122 Saint Genes, Champanelle, France

Department of large Animal Sciences, Faculty of Health and Medical Science, University of Copenhagen, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark

Contract/grant title: Use of animal based measures for the assessment of dairy cow welfare

Contract/grant number: CFP/EFSA/AHAW/2012/01– GA02

INTRODUCTION AND OBJECTIVES

The European Food Safety Authority (EFSA) aims to establish a practical and validated basis for data collection of animal based measures (ABMs) for farmed animal species and consequent quantitative risk assessment of the welfare of target populations. Through a series of integrated objectives this project evaluates the possibility to use routinely collected ABMs to predict the overall welfare status at herd level in dairy cow herds.

Firstly, a background will be provided which briefly sum up definition of the concepts of animal welfare and how to measure it. Secondly, the individual objectives and the aims of these will be introduced. In the following chapters of the report, material and methods and results from the Objectives 1-5 will be presented and then, Objective 6 will be presented with suggestions on which data are missing today and how they can by routinely collected in the future. Finally, a general discussion, conclusions and recommandations will be given.

1.1. Background: Measure of overall welfare status

1.1.1. Definition of the concept of animal welfare

Animals have been kept by humans throughout history, but the views on animal rights and welfare have changed substantially over time. In the early years of animal protection legislation, wanton cruelty was considered the main problem but since the Second World War the focus has shifted towards animal welfare by setting limits on established and new uses of animals (Sandøe and Christiansen, 2008). A pioneer in this process was the Brambell Committee in the UK. The final report from this committee stated that farm animals are sentient beings and hence, requiring that animals should have some basic freedoms. These were later contextualised by the Farm Animal Welfare Council (FAWC) in 1979 as 'The Five Freedoms': 1) Freedom from hunger and thirst, 2) Freedom from discomfort, 3) Freedom from pain, injury or disease, 4) Freedom to express normal behaviour and 5) Freedom from fear and distress. 'The Five Freedoms' have been widely adapted into many welfare assessment protocols existing today. Although the recognition of animals as sentient brings was generally accepted during the late sixties and early seventies, it was not until 20-30 years later that legislation protecting animal welfare were implemented in many countries. In the Treaty of Amsterdam (EU, 1997), it is stated that member states 'shall pay full regard to the welfare requirements of animals' in the formulation and implementation of the Community's policies in order to 'ensure improved protection and respect for the welfare of animals as sentient beings'. These public concerns are depicted in the growing concern that choices regarding farming techniques should not be based solely on efficiency of the production system but also protect these animals from mistreatment and poor welfare. Consumer's awareness has been increasing and the implications of a production system on the animal welfare are considered in rendering the sustainability of different production system (Broom, 2010).

In the main, scientists fall into three different schools in thoughts in defining animal welfare (Fraser and Broom, 1990b, Fraser *et al*., 1997), 1) 'Natural life' focussing on the animal's ability to live a natural life, 2) 'Function', which concerns the animal's health and biological functioning, and 3) 'Feelings' concentrating on the animal's experience of aversive or positive feelings. Disagreements about animal welfare might primarily echo the divergence in the perception of animal welfare in different stakeholders. For example, non-producers tend to put high weight to 'natural living' whereas producers lean towards biological functioning by deeming good health and access to necessities as the most important issues (Sørensen and Fraser, 2010). Other central terms in defining animal welfare is the 'needs' and the 'wants' of the animals. According to these theories, animals have complex functional systems to control its state, for example body temperature, nutritional state and social behaviour, which are cooperating in order to control the animal's interactions with the environment thereby keeping each aspect of its state within a tolerable range (Fraser and Broom, 1990a). Motivational mechanisms control how the animals allocate time and resources to different physiological and behavioural activities. Thus, a 'need' can be defined as a deficiency in the animal that should be fulfilled to obtain homeostasis and can typically be alleviated by obtaining a particular resource or by responding to a particular stimuli from the environment or the body (Broom, 1996). A 'want' is a need that the animal is aware of – it is associated with feelings. Some needs are for particular resources like water, others may be needs to perform certain behaviours like rooting in pigs or dust-bathing in hens.

In a scientific context, the concept of animal welfare has to be defined in a way that allows us to assess and quantify good and poor welfare in a standardised way. The World Organisation for Animal Health (OIE) defines animal welfare as follows: 'Animal welfare means how an animal is coping with the conditions in which it lives. An animal is in a good state of welfare if (as indicated by scientific evidence) it is healthy, comfortable, well nourished, safe, able to express innate behaviour, and if it is not suffering from unpleasant states such as pain, fear, and distress'. Thus, animal welfare refers to the state of the animal, which is in the line with the following definition: 'the welfare of an animal is its state as regards its attempts to cope with its environment' (Broom, 1986).

Previous work in EFSA on the issue of animal welfare has recognised the multidimensional concept of animal welfare including the physical as well as the mental state of the animal (European Food Safety Authority, 2012a)). Also, despite the conceptual diversions between different definitions of animal welfare, it has been shown that scientists generally agree upon what they consider as the most serious risks to animal welfare and also when regarding dairy cows, scientists generally agree on the range of different needs (European Food Safety Authority, 2009). A variety of measures of particular welfare challenges have been presented over the years. As summarised in the EFSA report (European Food Safety Authority, 2009, measures of production, reproduction and health can provide useful information about animal welfare. Also, physiological measurements of stress and pain and behavioural observations can be valuable indicators of poor welfare. However, it is also agreed that multiple welfare indicators typically are needed in order to describe the overall welfare of the dairy cow.

1.1.2. Concepts of welfare assessment and the aggregation of welfare measures

There are two general approaches to the measurement of welfare: Either measuring indicators of poor welfare or recognition of good welfare (Fraser and Broom, 1990b), which can also be used in combination. As presented earlier, it is generally agreed that a welfare compromise arises if an animal is not able to satisfy one or more of its needs and wants. Depending on the motivational state of the animal some needs are of greater urgency than others. The importance and consequences of the different needs are often deduced from situations where some inadequacy in the environment is present. Alternatively, it can be assessed by using preference studies observing what the animal will do if given a choice (Fraser and Broom, 1990a).

In order to obtain an overall welfare measure, several welfare indicators can be aggregated into one common score depicting the level of welfare in the herd and typically, depending on the purpose of the welfare assessment, a predefined acceptance level is presented. The evaluation models can be descriptive, normative or prescriptive (Botreau et al., 2007). Descriptive models are describing a pre-existing, stable situation and are used to describe and compare observed situations. Normative models aim at checking the adequacy of observations against pre-defined norms and they can thereby be instructive regarding how people should act. Finally, in the prescriptive models no assumptions are made regarding the pre-existing situation. Rather, information are collected and collated in order to improve decisions taken and activities made in order to reach a certain goal.

The process of aggregating the welfare indicators has been approached in different ways in different studies. A non-explicit aggregation of several measures performed by an expert (or a group of experts) based on observations in one animal unit can be used to advice farmers. One example is the 'ethical account' presented by Sørensen et al. (2001). Here, animal based measures as well as environment based measures are considered and a report identifying welfare problems and their possible causes plus proposed strategies to improve the welfare status are prepared for the farmer. This kind of aggregation without using mathematical tools is simple. However, the aggregation process is less transparent and can be influenced by the expert(s)' subjective interpretation of the welfare measures (Botreau et al., 2007). Benchmarking systems (e.g. The Freedom Food Scheme (RSPCA, 2011)) or systems based on hazard analysis of critical control points (HACCP) are listing standardised, minimum requirements that farms, housing systems or equipment have to meet. The comparison of the measures at farm level is straightforward and transparent to the stakeholders which make these methods suitable for certification purposes. However, if no flexibility is induced in such a model it is rather conservative as no distinction is made between a farm that fails on only one aspect and a farm that fails on many aspects. Also, if strictly applied, the impact of each of the different measures included in these models is the same which not necessarily depict the level of suffering implied by the different elements (Botreau et al., 2007).

The 'Five Freedoms' protocol aims at a direct assessment of the animal's welfare. It is based on the logic of the 'Five Freedoms' first introduced by the Farm Animal Welfare Council (FAWC) in 1979 (FAWC, 1979). The concept of this system is to assess the welfare of the animal in terms of its state of nutrition, comfort, health, temperament and behaviour (Whay et al., 2003c). This system aims at a detailed evaluation of the animal welfare taking into account a wider variety of the essential 'needs' of the animal. The purpose is to evaluate the dairy cattle welfare at farm level and thereby provide a welfare ranking system of different herds (Whay et al., 2003c). The choice of measures included is based on results from a Delphi study, where opinions from experts on animal welfare were gathered (Whay et al., 2003a). Results are presented as a ranking of herds. Herds are initially ranked for each measurement and the overall rank of a herd is calculated as the mean rank of all measurements. Thus, some compensation is allowed.

Alternatively, scores can be summarised in different ways. When using the sum (or mean) of the ranks, an overall rank of each farm is found by calculating the sum (or the mean) of the partial ranking of each of the measures applied or alternatively, comparing the farm with the average value obtained over the population of inspected farms (Whay et al., 2003b, Whay et al., 2003c, Huxley, 2004). The main concern about using this kind of aggregation is that the overall ranking of a farm depends on the study population (Botreau et al., 2007). Instead, the weighted sum (or mean) of scores which provides absolute values can be used independent of the sample observed. This method is the most commonly applied and a common framework can be seen. Firstly, raw data are converted into partial scores on a commensurable scale. Next, weights are assigned to the values obtained for the different measures and finally, an overall score is calculated (Bartussek, 1999, Scott et al., 2001, Botreau et al., 2007). One example is the ANI 35 L/2000 (Bartussek et al., 2000). This method is generally easy to understand and while the overall score allows comparison between farms, the partial scores can be used actively to point out strong and weak areas within each assessed farm. However, to calculate a weighted sum on data assessed on interval or ratio scales it is assumed that intervals on the different scales are equidistanced – which might not always be true and therefore results can be confusing (Botreau et al., 2007). Also, sum of scores allows full compensation between the different welfare aspects, which might not be desirable (Spoolder et al., 2003).

In the assessment systems like the one presented by Capdeville and Veissier (2001), the aggregation process is performed in different stages. Initially, a few measures are grouped together; then group of measures are aggregated; and so on in a hierarchical procedure. After the aggregation of measures, Capdeville and Veissier (2001) applied logical rules that were designed in such a way that a good score could not fully balance a poor score. The rules valued the importance of the measures for the welfare of the animals and thus allowed less (or none) compensation for measures considered especially important. However, the aggregation process is not very transparent and thus, it can be hard to judge validity of the system (Botreau et al., 2007).

One of the most comprehensive and complex welfare assessment systems is the Welfare Quality® protocol (WQ®, Welfare Quality® (2009)). Here, amalgamation of measures and scores are done in three, hierarchical steps. Initially, 29 primarily animal based indicators are combined to calculate 12 criteria scores which are then combined into 4 principle scores. In this process, compensation is allowed within criteria but not between criteria or between principles. Finally, an animal unit is assigned to one of four welfare categories defined by minimum values (‘aspiration values’) for each welfare principle. As no compensation is allowed at this step all four principles should be above the ‘aspiration value’ to be assigned to a certain welfare category (Welfare Quality®, 2009). In the process of defining the aggregation of the measures in the WQ® protocol, expert opinion on the weighing of measure levels in the interpretation of scores was incorporated resulting in a non-linear model.

The multidimensional welfare index presented by Burow et al. (2013) is based on the ideas of the WQ® protocol but the number of indicators and criteria is reduced. Compared to the WQ® model, where a hierarchical aggregation procedure is applied, this multidimensional welfare index calculates the overall welfare index in a one step and single measure aggregation procedure resulting in a simpler, additive, linear model. This approach was taken to avoid the situation where a single measure can be related to more than one welfare aspect but due to the aggregation procedure, the measure is arbitrarily considered as part of only one welfare aspect (Botreau et al., 2007, Burow et al., 2013). The weighting of the measures and measure severity was defined using expert opinion. In contrast to the WQ® model, this system allows a single measure to affect the overall welfare index strongly depending on the prevalence and weight of the measure.

A final aspect is that the animal based indicators in on-farm animal welfare assessment for practical reasons often are used as group-level indicators providing information about the level of a certain welfare hazard in the herd (e.g. prevalence of lameness). However, welfare is a characteristic of the individual animal and it is relevant to assume that the welfare on a given farm is a sum of the welfare of the individuals in the farm. By using samples of animals the exact distribution of the indicator between the individuals is not known and therefore it is not possible to put a special weight on the individuals with the worst welfare (Jensen and Sandøe, 2013).

1.2. Animal based measures in welfare assessment

On-farm animal welfare assessment can be resource-based or animal based. Basically, resource-based indicators included in welfare assessment systems describe features of the environment and management, whereas animal based indicators rather measures how the animal copes with the given environment. An animal based indicator is a response of an animal – or an effect on an animal – that can be used to assess its welfare. Animal based indicators fall within the categories of behaviour, health and physiology. The animal based indicators can also be divided into direct indicators (measured by looking/measuring directly on the animals) and indirect indicators (based on register data about the animals like e.g. production or health records). Typically, assessment of resource-based indicators is fairly uncomplicated, easy and quick to perform. Also, it often serves as a strong basis for problem solving. The recording of animal based indicators is typically more cumbersome and the interpretation might be more difficult. However, they do provide more information about how animals are coping with the given environment (European Food Safety Agency, 2012a).

In order to be a valid and robust measure of welfare, an ABM should be closely related to aspects of the welfare (worst adverse effects). This means that the level of the ABM should change accordingly to changes in the animal’s perception of well-being and its attempt to cope with the given circumstances. Furthermore, the ABM should be robust in the sense that it can be reproduced over time and by different observers.

In the EFSA scientific opinions on the use of ABMs to assess the welfare of dairy cattle (European Food Safety Authority, 2009a-f), the severity and magnitude of the impacts of hazards in dairy production systems on the welfare of dairy cattle have been identified and scored. ABMs associated with the hazards have also been identified (European Food Safety Authority, 2009a-f; European Food Safety Authority, 2012a, b).

Results from measuring an ABM may vary according to a number of epidemiological parameters. Epidemiological parameters include risk factors but also factors that allow us to characterize a given population of farms. In epidemiological terminology, a risk factor may or may not imply causality. In the concepts of this project, we used the definition where a risk factor is "any factor associated with the increase of appearance or development of a phenomenon" (Toma et al., 1996). The phenomenon could be a welfare insult, i.e., in the context of this project: the score of an ABM outside a range of acceptability. Detecting such risk factors can help target the origin of the welfare insult more specifically.

1.2.1. General methodological considerations for describing validity and robustness of theABMs

This section aims to provide a brief overview of the diagnostic test terminology to illustrate how the methodology can be used to evaluate the validity and robustness of the ABMs as indicators of the worst adverse effects (WAEs). Here the WAEs are the 'target condition', i.e. what we wish to record. The description is thus fairly simple, but it may not be straightforward to record the WAEs objectively. The practical realization of the target condition is the case definition. A case definition for a WAE could be recording that an animal was healthy or more specifically "lame" or "not lame".

The performance of ABMs to correctly identify a WAE can be characterized by the sensitivity and the specificity, which are defined as:

- Sensitivity = (number of cows with ABM)/(number of cows with target condition)
- Specificity = (number of cows without ABM)/(number of cows without target condition)

Both include the term target condition. However, most often this is not available. Instead the relative sensitivity and specificity are estimated:

- Relative sensitivity = (number of cows with ABM)/(number of cows among cases)
- Relative specificity = (number of cows without ABM)/(number of cows among non-cases)

The relative sensitivities and specificities are quite often reported, but their validity strongly depends on how well the case definition reflects the target condition. This is often a very subjective assessment, and there are no measures to record this difference. Consequently, assessment of the performance measures is also subjective. The latent-class methodology can be used to identify unobservable traits within a population. Latent variables (as opposed to observable variables), are variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed (directly measured). This methodology can in some instances be an option, but it would require that we deviate from case definitions and let data determine the latent condition. This would instead require that we define different angles to our target condition. While the target condition has been defined via WAE, this is not relevant in the present project.

Sensitivity and specificity may here refer to the ABM's ability to measure its' specific case condition, or it may refer to an ABM's ability to measure a specified adverse effect. It can also be made on the recording level (i.e. how well is an ABM recorded to a database), which is often referred to as completeness in literature. In the present literature review and data analysis, all aspects are covered.

The robustness of the recording of ABMs can also vary due to the recording method, e.g. inter- and intra-recorder variability. The recorder can be a machine or a person, and the recordings can be made on different scales of the parameters, typically dichotomous, ordinal or continuous scales. For each scale, appropriate measures of intra- and inter-observer variability often include:

- Dichotomous or ordinal recordings
    - Cohen's Kappa coefficient, or agreement beyond chance
- Ordinal recordings
    - Spearman rank correlation coefficients
- Continuous variables
    - Pearson's correlation coefficients.

All these parameters are dependent on the prevalence of the target condition in the population, and care should thus be exerted in their interpretation. Furthermore, their interpretation is still subjective.

A guideline for Kappa suggests the following interpretation: >0.75: Excellent; 0.4-0.75: Fair to Good; and <0.4: Poor (Fleiss, 1981). Similarly, interpretation of the magnitude of correlation coefficients is subject to subjectivity. For example, correlation coefficients of 0.05 may be considered good in breeding programmes, whereas a correlation coefficient of 0.9 may be deemed insufficient for a device recording para-clinical parameters. However, in general the lower correlation coefficients the lower is the robustness and thus the reliability. Consequently, users may lose faith in recordings with low agreement or low correlation.

1.3. Objectives of the project

The overall aim of this project was to test whether routinely collected ABMs and associated factor of variation can be used to evaluate the welfare status of dairy herds across countries in the EU. In order to establish the basis for a European evaluation of the welfare status, prerequisites are the identification of valid, robust ABMs for dairy cows and relevant epidemiological risk factors (factors of variation) that are collectable and well correlated to the WAEs; the establishment of a platform for the data collection; and an evaluation of the selected ABMs and factors of variation with regard to their ability to predict the overall dairy cow welfare status in dairy cow herds across the EU.

This was pursued through a series of six specific objectives.

Objective 1 seeks to build on the EFSA developments by identifying the worst adverse effects (WAEs) of hazards in the EU member state (MS) intensive dairy production systems and, then, identifying ABMs already collected that can be utilised to detect such adverse effects. This was done by identifying from the EFSA Scientific publications on welfare of dairy cows the WAEs/consequences for dairy cow welfare in terms of prevalence and impact and those ABMs listed in the opinions that allow detection of these adverse effects and that are already being collected in the field. The outcome was a list of WAEs for dairy cattle welfare and an evaluation of ABMs already collected in the field and relevant to detecting such worst adverse effects.

From this list of WAEs and associated ABMs, Objective 2 seeked to identify which of the ABMs selected in Objective 1 would be relevant to collect for quantitative risk assessment of dairy cow welfare. For this purpose, we need to know the essential attributes of these measures. Thus, Objective 2 evaluated the ABMs selected in Objective 1 regarding validity and robustness. Initially, existing literature was reviewed regarding the validity and robustness of the ABMs toward their target condition and identifying studies that evaluate the ABMs as tests for the WAE. Secondly, there a description of the databases available to the consortium members was made and the sensitivity and specificity for selected ABMs as tests for selected adverse effects was calculated. The outcomes of Objective 2 comprised: 1) A description of relevant methodology for the evaluation of validity (sensitivity, specificity) and robustness (intra-observer agreement and inter-observer agreement) of selected ABMs and 2) estimates of sensitivity, specificity, intra-observer agreement and inter-observer agreement of selected ABMs based on results from the literature in combination with calculations performed on available data.

The overall goal of Objective 3 was to propose a limited list of epidemiological parameters (termed factors of variation) for the ABMs selected in Objective 2. Firstly, the main factors of variation associated with ABMs outside the range of acceptable welfare was identified in the literature and included in a list of factors leading to variation of the ABMs selected in Objective 2. Next, a limited number of these factors of variation were selected by partners in the consortium. This was done based on 1) the parameter's association with the ABMs from Objective 2, 2) the feasibility of the collection and possibility of keeping it updated in the field and 3) their capacity to characterise a population. Finally, the current availability in the field of the parameters selected was evaluated. Thus, the outcomes of Objective 3 were 1) a list of factors of variation associated with the ABMs selected in Objective 2, as found in a literature search, 2) a shortlist of final factors of variation selected by partners within the consortium and 3) information on the collection of the final factors of variation in existing national databases from the countries participating in the project.

The aim of Objective 4 was to develop a model to collate routinely collected data (RCD) from different countries represented in the ANIBAM consortium, based on the ABMs and factors of variation selected in Objective 3 and to perform a pilot test to collate these data. This enables the consortium to examine which data are routinely collected in a sample of EU countries, whether these could be collated in a single database and whether this database has potential use for predicting the risk of pertinent welfare problems of dairy cows.

The outcomes of Objective 4 were a pilot test of the sampling and merging of data on ABMs and factors of variation and an analysis of possible data management errors.

In Objective 5, the usefulness of selected ABMs for welfare risk assessment was evaluated. It was hypothesised that ABMs will be able to give relevant information about the 'overall welfare status' in a dairy herd. Also, it was expected that a combination of two or more ABMs will increase the precision of the prediction of the overall welfare level. The association between individual ABMs and the overall welfare status was evaluated using available research datasets. The 'marginal information value' of combining more ABMs was evaluated. Furthermore, the additive value of combining the ABMs with additional factors of variation was evaluated. The outcomes of Objective 5 were recommendations regarding the potential of using the selected ABMs and associated factors of variation for the prediction of herd level welfare status.

As part of the other activities in the project, ABMs and factors of variation may be identified where no routine recordings are made. The aim of Objective 6 was to discuss possibilities for future collection of a few selected ABMs/factors of variation in dairy herds.

MATERIALS AND METHODS

2. Objective 1

2.1. Identification of the WAEs on dairy cattle welfare

The WAEs for the welfare of dairy cows in intensive production systems (cubicle, straw yard and tie-stall) in Europe were derived from the full complement of adverse effects listed in the European Food Safety Authority (EFSA) publications (EFSA, 2009a-f; 2012a,b), Presi and Reist (2011), and Brenninkmeyer and Winkler (2012). Based on this full list and as explained in more detail below, a three-step process was used to identify the worst adverse effects:

a) Pre-selection of a list of worst adverse effects

b) Use of an expert elicitation procedure to identify and rank the pre-selected worst adverse effects

c) Selection of the highest ranked adverse effects (= worst adverse effects, WAE).

2.1.1. Pre-selection

The full list of adverse effects considered were those identified in the EFSA reports (2009a-f, 2012a,b), Presi and Reist (2011) and Brenninkmeyer and Winkler (2012) and are listed in Appendix A. A pre-selected list of WAEs was identified based on combination of ratings of severity score (3 or 4) and magnitude (10 or greater, based on a combination of prevalence and duration), as reported by EFSA (2009c-f). According to EFSA (2009c) severity score 3 is:

“substantial changes from normality indicative of pain, malaise, fear or anxiety. Strong change in adrenal or behavioural reactions, such as motor responses and vocalisations”

and severity score score 4 is:

“extreme changes from normality indicative of pain, malaise, fear or anxiety, usually on several measures that could be life-threatening if they persist”.

This pre-selection procedure resulted in a list of 16 WAEs as shown in Appendix B. These were cross-checked for consistency with the WAEs associated with animal based indicators identified in the EFSA report (2012a) and in the outputs from the EFSA technical meeting in 2012 (EFSA, 2012b). In order to ensure that other potential WAEs were taken into consideration during the expert elicitation procedure (see Section 2.1.2), respondents were asked to add any additional adverse effects (diseases, injuries and/or other) they considered rated as ‘worst adverse effects’.

2.1.2. Expert elicitation

The project consortium (Aarhus University (AU), University of Copenhagen (UCPH), Swedish University of Agricultural Sciences (SLU), Institut National de la Recherche Agronomique (INRA), Institute for Agricultural and Fisheries Research (ILVO), Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna “Bruno Ubertini” (IZSLER), Istituto Zooprofilattico Sperimentale del’Abruzzo e del Molise “G. Caporale” (ICT)) representing all member countries of the consortium (Belgium, Denmark, France, Italy and Sweden) contributed the experts (n=9). The professional skills of the respondents were:

veterinary epidemiological science (n=2), epidemiological health/welfare science (n=1), veterinary health/welfare research (n=3) and livestock welfare science (n=3).

Each expert was presented with a written questionnaire by email, with follow up by email or telephone, if required, for clarification or to ensure timely completion. The experts were provided with copies of the relevant EFSA reports on dairy cattle welfare (EFSA 2009a-f; EFSA 2012a, b); thus, they had access to published information about the severity and magnitude of all adverse effects (both the pre-selected list and other adverse effects).

The following instructions were provided to the respondents:

a) *"The worst adverse effects listed below (1-16) have been pre-selected from the EFSA documents. Please assess these by answering the questions in columns C to H. From your OWN reading of the EFSA documents, ADD any additional adverse effects that you consider should be rated as 'worst adverse effects' (provide details in the 'Other' categories (17-19)). State the adverse effect that you are referring to."*

b) *"Complete the questionnaire based on your personal knowledge together with that in the EFSA reports circulated. If you cannot answer confidently enter the letter N."*

The EFSA reports (2009 c, d, e, f) showed that the magnitude of most adverse effects varied with housing system. Thus, for these cases, the respondents were required to complete the questions separately for tie-stall, cubicle and straw-yard systems.

The key information sought from the questionnaire was a rating of the overall impact of an adverse effect (e.g. foot disorders and associated effects) by considering both its severity (0 to 4 scale) and herd level prevalence (0 to 4 scale). The severity and prevalence definitions were provided in the survey and are reproduced in Table 1: and Table 2: . The definitions for each level of severity were as described by EFSA (e.g., 2009c). The respondents were asked to use a matrix with severity on one axis and prevalence on the other to help in providing this information. In case the respondents considered that some important adverse effects had been omitted from those listed, the opportunity was provided for them to add and rate further adverse effects (diseases, injuries or other).

In addition, in order to ensure that particular aspects of adverse effects with severe consequences were not overlooked, the respondents were asked to identify a 'condition' (digital dermatitis would be an example from foot disorders) that they considered was the WAE. Then they were asked to rate the severity of this adverse effect. As it turned out this information on 'conditions' was not utilised, as the prevalence of the identified conditions was not known; thus, the overall impact based on severity and prevalence of these conditions could not be determined.

Table 1: Severity score criteria

Score	Description
Score 0	Negligible. No pain, malaise, frustration, fear or anxiety as evidenced by a range of behavioural, physiological and clinical measures
Score 1	Mild. Minor changes from normality indicative of pain, malaise, fear or anxiety
Score 2	Moderate. Moderate changes from normality indicative of pain, malaise, fear or anxiety. Clear change in adrenal or behavioural reactions, such as motor responses and vocalisations
Score 3	Severe. Substantial changes from normality indicative of pain, malaise, fear or anxiety. Strong change in adrenal or behavioural reactions, such as motor responses and vocalisations
Score 4	Very severe. Extreme changes from normality indicative of pain, malaise, fear or anxiety, usually on several measures that could be life-threatening if they persist

Table 2: Herd level prevalence definitions [(a)]

Score	Percentage (%)
Score 0	0-4
Score 1	5-9
Score 2	10-24
Score 3	25-49
Score 4	50+

(a): Although the occurrence for some conditions like mastitis is often reported as incidence risk or incidence rate, the respondents were required to give the herd prevalence (= number of animals affected on a given day/number of animals present on the same day). This can mean that a condition with e.g. an annual incidence risk of 30% or incidence rate of 50/100 cow years, may only have a prevalence of 1-2%.

2.1.3. Analysis of WAE expert elicitation

Each adverse effect was characterised by the median of the responses for severity and prevalence scores. Since the severity and prevalence measurements of the WAEs are not commensurate, the severity and prevalence data for each adverse effect were arranged in a matrix (severity x prevalence). A consensus agreement amongst the experts (amongst consortium partners) was used to rank the WAEs taking into account of the quantitative information displayed visually in the matrix and professional judgement of the trade-off between severity and prevalence. The consensus agreement was that: Severity 4 adverse effects were regarded as more extreme than those accorded Severity 3 (based on the extreme level of suffering likely to be experienced) and Severity 3 is more severe that Severity 2 and so on, and that adverse effects with the same severity score and a higher prevalence level would be rated as worse than those with a lower prevalence.

2.2. Identification of ABMs allowing detection of the WAEs and already collected in the field

The section comprised three components:

- Identification of ABMs allowing detection of the WAEs
- Identification of ABMs already collected in the field
- Identification of relevant and routinely collected ABMs for detecting the WAEs (and for further investigation in Objective 2).

2.2.1. Identification of ABMs allowing detection of the WAEs

The EFSA reports (EFSA, 2009a-f; 2012a, b) were searched to develop a comprehensive list of ABMs that could potentially be used to detect the worst adverse welfare effects identified in Section 2.1. All ABMs reported in these EFSA documents as being associated with these WAEs were selected. The list was submitted to the consortium partners for their contributions to the list. The final list of ABMs identified was:

- Numbers of hock, knee, skin lesions and swellings
- Measures of lameness (e.g. locomotion score)
- Evidence of discomfort when standing (e.g. time resting a foot)

- Numbers of foot lesions or infectious foot conditions
- Evidence of mastitis (e.g. somatic cell count)
- Measures of overgrown/misshapen hooves
- Number of dead animals (unassisted death)
- Number of dead animals (euthanised)
- Measures of abnormal movement (e.g. number of slips)
- Frequency of agonistic behaviour (e.g. numbers of chasing up from lying)
- Measures of lying in passage (e.g. % animals)
- Measures of abnormal standing-up or sitting-down behaviour (e.g. rising with front legs first)
- Measures of standing in water/slurry (e.g. number of animals)
- Measures of posture at rest (e.g. number of cows lying diagonally)
- Cleanliness score
- Number of collisions with equipment
- Time spent resting
- Time spent standing
- Measure of nutritional status (e.g. body condition score)
- Measure of hind legs in cubicle passage (e.g. number)
- Measures of feed intake (e.g. feeding time, rumen fill)
- Measures of behaviour at feeding (e.g. number of displacements)

To allow for the possibility that the experts may have identified other appropriate ABMs, an additional category was created i.e. Other ABMs (any additional ABMs used/suggested by the respondent).

2.2.2. Identification of ABMs already collected in the field

The list of appropriate ABMs was submitted to experts with knowledge of the use of ABMs in intensive dairy cattle production systems in the EU. The nine MS that produce 80% of the EU's milk supply (Germany, France, UK, Netherlands, Italy, Poland, Denmark, Belgium and Sweden) were included in the survey to determine which of the ABMs identified are already collected in the field. The aim was to have the survey completed by up to three experts from each country. All consortium partners were requested to provide the names, contact details and relevant expertise of experts in these nine MS. Other experts were recommended by professionals known to the consortium and active in the area of ABM measurement or collection (EFSA, Bristol University and Copa-cogeca). The criteria agreed by the consortium for selecting experts were: a good practical knowledge of the use/recording of ABMs in the field and a good knowledge of their relevance to assessing overall welfare. Experts meeting the criteria included practising veterinarians, producers/farmers, retailer organisations with welfare assessment protocols, members of NGOs with experience of practical welfare assessment protocols, animal welfare researchers, and competent authorities. This process generated a total of 44 respondents, with a minimum of three for each of the nine MS.

The survey was sent to a total of 44 experts: 15 researchers; 12 producer organisation representatives; 7 retailers; 4 practising veterinarians; 4 competent authorities; and 2 NGOs. The experts were asked to complete a written questionnaire, with telephone follow-up to facilitate completion (if required). Email and telephone follow-up, both with the experts themselves and colleagues of the experts, was used to increase the response rate. The questionnaire provided the following instructions:

"This questionnaire is part of a pilot project proposed by EFSA. The objective of the questionnaire is to determine if there are animal based measures (ABMs) for assessing the welfare of dairy cows that are already collected in the field with quantitative data. The ABMs (e.g. hock lesions) listed below are useful for identifying/detecting/measuring the worst adverse welfare effects (e.g. leg disorders) for dairy cows. This survey is asking only about the existence of the ABMs and quantitative data and is not seeking to obtain the quantitative data. The aim of the pilot project is to provide EFSA with tools for moving towards a quantitative risk assessment of the welfare of the dairy cows by using quantitative data for ABMs. There are two situations for which we are seeking information:

1. *ABMs that are collected routinely on most farms.*
2. *ABMs collected on herds reasonably representative of the population."*

The respondents were asked to provide their name, the country and organisation they represented, and characteristics of the country's farms. It was made clear that the anonymity of the respondents would be protected. The respondents reported that all intensive production systems were present in the countries surveyed (cubicle, straw yard, tie stall in Belgium, Germany, Italy and Sweden; and cubicle and straw yard in Denmark, France Netherlands and the UK). The average MS herd size varied from 20 to 700.

The respondents were provided with the complete list of selected ABMs (shown in Section 2.2.1).

They were asked to provide information on five topics about the ABMs 'collected in the field' in their country. As agreed by the consortium, 'Collected in the field' was defined in two different ways: (1) registered data (e.g. routinely collected on most farms by farmers/veterinarians/competent authorities (denoted as *ABMs collected routinely on most farms by qualified personnel* in the remainder of this document) (2) research or assurance assessment data on widely-used protocols that may be routinely collected on herds over defined periods that are reasonably representative of the population (but not collected routinely in all herds) (denoted as *ABMs routinely collected on herds reasonably representative of the population* in the remainder of the document). Thus, the five questions about ABMs were requested for each of these two 'routinely collected' situations.

The key information was sought in the first question. The respondents were asked (Yes/No) if each ABM was recorded in their country. In an attempt to facilitate the work in subsequent objectives of the project, if an ABM was routinely recorded the respondents were also requested to provide:

1. A brief description of the ABM collected;
2. The type of person (veterinarian on-farm, farmer, competent authority, researcher, assurance assessor, other) who recorded the ABM;
3. The percentage prevalence of the ABM at the herd level;

4. The percentage of animals with the ABM recorded in the database.

Taken together, the two pieces of information requested in (3) and (4) were designed to provide Objective 2 with a measure of the usefulness of the ABMs recorded for identifying WAE. As it turned out, there was too little information provided on this topic to be useful (see Results Sections 7.2.1and 7.2.2).

2.2.3. Analysis of survey of ABMs collected in the field

The frequencies of recording ABMs in the two situations (*ABMs collected routinely on most farms by qualified personnel* and *ABMs routinely collected on herds reasonably representative of the population*), the frequencies of the different recorders used and the frequencies of the definitions of ABMs were calculated. In addition, the medians of the prevalence of each ABM in each situation and the percentage of animals with the ABM recorded in the databases, together with measures of variability (minimum and maximum, 1st and 3rd quartiles) were calculated.

2.2.4. Identification of relevant, routinely collected ABMs for detecting the WAEs

This was a multi-step process:

1. An expert elicitation procedure utilising the consortium partners was used to help identify the appropriate ABMs. Firstly, the consortium members were presented with a diagram (see Appendix C. (a)) showing both the identified WAEs and the ABMs already routinely-collected in the field. They were asked to identify the ABMs that allow detection of adverse effects by placing an arrow on the diagram linking the ABMs with specific adverse effects. The instructions to the respondents highlighted that each ABM may link to more than one adverse effect. For analytical purposes, the responses were amalgamated, further discussed by email and agreed between the partners. Further, particular adverse effect may serve as an animal based measure of other adverse effects. Thus, the consortium partners were presented with a second diagram (see Appendix C. (b)) and requested to identify the adverse effects that allow detection of other adverse effects by placing an arrow on the diagram linking the adverse effects. All seven consortium partners responded, with multiple responses from University of Copenhagen and IZSAM giving a total response number of nine. For analytical purposes, the responses were amalgamated, further discussed by email, agreed between the partners and used to assist with identification of the appropriate ABMs as part of the process detailed in parts (2) – (5) below.

2. ABMs recorded on the most farms were accorded high priority;

3. Highest priority was accorded to ABMs associated with the most severe WAEs. The consortium's professional judgement was that: Severity 4 adverse effects were regarded as more extreme than those accorded Severity 3 (based on the extreme level of suffering likely to be experienced); and that adverse effects with the same severity score and a higher prevalence level would be rated as worse than those with a lower prevalence;

4. Lowest priority was accorded to ABMs that were associated with WAEs that did not apply to all housing systems and had lower prevalence;

5. If a single ABM had the potential for use in detecting both highly-ranked and lower-ranked adverse effects, then it was recommended over one that could detect just a lower-ranked effect.

3. Objective 2

3.1. Literature review

The validity of the ABMs selected in objective 1 was considered from two perspectives: First, the ability of the measure to correctly identify animals or herds with the target condition (e.g. lameness scoring for identifying lame cows or herds with a specified level of lameness). At this level also robustness estimates are provided. Second, for this project we were also interested in evaluating the validity of the different ABMs in the prediction of specified adverse effects (e.g. lameness scoring for identifying herds with high cow mortality). In order for an ABM to be a 'useful' ABM, it should be valid and robust from the first perspective as well as valid from the second perspective. Validity is described by sensitivity (Se) and specificity (Sp) and robustness covers intra- and inter-observer agreement. Low validity can also be due to recording failure (low completeness).

For the literature review, the definitions of the ABMs and the WAEs were kept in general terms, because we needed to use the definitions described in the relevant studies. Later on, detailed and specific definitions of the ABMs and the WAE were needed to perform calculations. These definitions and the process of describing them are presented in Section 8.2 and 8.4.

Relevant literature was identified in two ways. Firstly, systematic searches in databases as described below were used. Secondly, the consortium members' immediate knowledge of both original publications and review literature from recently finalised and on-going projects was used to identify a number of relevant publications.

The review was performed following the EFSA guidance: "GUIDANCE OF EFSA - Application of systematic review methodology to food and feed safety assessments to support decision making" (European Food Safety Authority, 2010).

The guidance outlines the following key steps in the review: preparing a review, searching for studies, selecting studies for inclusion, collecting data from included studies, assessing the methodological quality of included studies, synthesising data from the studies, presenting data and results, interpreting the results and drawing conclusions.

The preparation of the review was provided by the formulation of the objectives of the EFSA project as well as the identified list of ABMs and WAEs. In the search for studies, search words were identified within 3 different areas or so-called key components:

1. Relevant animal species
2. Terms for validity and robustness
3. Terms describing the individual ABMs

The searches were performed in Web of Science looking up the terms describing the key components in the search field 'Topic' (search was also done using 'Title' but this provided very limited literature and therefore a broader search was used).The specific search strings used for each of the areas are presented in Table 3: . Within the three different areas the Boolean operator OR was used and between the different areas or key component the Boolean operator AND was used.

Table 3: Search strings for each area in the literature review

Areas	Name	Search string
1	Relevant animal species	Dairy cow* or cow* or cattle
2	Validity and robustness	sensitivity or specificity or reliability or repeatability or reproducibility or agreement
3	The individual ABMs	
	Mortality (unassisted and euthanised)	Mortality or dead or death or deaths or survival or euthan*
	Evidence of mastitis	Somatic cell count or SCC
	Measures of lameness	Lameness or lame or locomotion or locomotory or gait score
	Number of foot disorders	Foot lesion*' or foot disorder*' or claw lesion*' or digital lesion*' or hoof trim*
	Number of leg lesions	Tarsal lesion* or tarsal swelling*' or carpal swelling* or carpal lesion* or hock lesion* or hock swelling*
4	Behavioural disruptions	Lying down or time to lie down or collision or aggression or agonistic behaviour

To limit the results to dairy cows, the following terms were used for all searches: 'dairy cow*' or 'cow*' or 'cattle'. This was used in combination with terms describing the ABMs. The results were restricted to validity and robustness by using the terms: 'sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement'. "Herd sensitivity" and similar herd associated terms are not terms for medical subject headings (MeSH terms) and were consequently not included.

The search words can be combined and reduced in numerous ways. The search strategy was optimised in different ways. Firstly, the strategy was made sensitive by including many synonyms (such as cows or cattle). Hereafter AND was used between the areas to narrow it down to the key components. Examples of the sensitivity to different combinations of search words are provided in the Appendix F. F-H).

After searching, the publications were selected for inclusion by screening for relevance. For example, despite the detailed search words it could happen that the article dealt with beef cattle. In the collection of data it was sought if they used the same measure that could be compiled in a table. These included first of all the measures on validity (sensitivity and specificity) and robustness (correlations, kappa etc.), which were compiled in tables where applicable. Other available and important epidemiological information such as sample size or confidence intervals was also considered. In cases where a measure was covered by very few publications, or where the results from the publications could not be compared, the results were summarized in the text of the result section.

The methodological quality of the studies was assessed according to the general state of the art of conducting epidemiological studies including relevant study population, adequate study design, sufficient sample size and avoidance of too much bias such as selection bias, information bias and confounding bias.

The data were synthesised and presented in summary tables by elaborating on the gross tables made in the phase of collecting the data.

3.2. Description of databases

For the analyses of the association between the ABMs and the WAEs, three datasets were available to the consortium:

1. Data from IZSLER/CReNBA, Italy

2. Data from ILVO, Belgium

3. Data from INRA, France

The database owners of relevant databases filled out a 'database protocol' including information about:

- Purpose of the database
- Responsible person(s)/contact person
- Data ownership and availability
- Which variables are recorded according to legislation
- Who does the recording
- Overview of variables (types and coding)

This 'database protocol' together with detailed information about the variables in the datasets can be found in Appendix J. and Appendix L.

3.3. Description and definitions of the WAE and the ABMs

The descriptions of the ABMs and the definition of cut-offs of the ABMs in this project were initially approached by expert discussions within the consortium. At ScienceNet, different definitions were presented and initially consortium members agreed on how to measure a given ABM. The next step was to define cut-offs for the selected measures. This discussion also took its' beginning at ScienceNet and was followed up with a teleconference. At this teleconference, the participants agreed on how to measure the ABMs and different cut-offs were suggested based on experts' opinions and experience with data availability (see Section 8.2 and Table 23:).

The available data were exploited by creating a matrix (Table 24:) with all possible combinations of the WAE and the ABMs. For each of these combinations data were consulted to see whether relevant data components could be identified in the data (Section 8.3). Finally, one specific cut-off per parameter was chosen based on the discussion within the consortium and investigation of data (Section 8.4, Table 25: (WAEs), and Table 26: (ABMs)).

3.4. Data analysis

3.4.1. **Analysis of the data**

Data management and data analyses were done in SAS version 9.3 (SAS Institute, Cary, North Carolina, USA).

3.4.1.1. Descriptive analysis

The defined WAE and ABMs were identified in the available data and if necessary, new variables were created to obtain dichotomised variables according to the definition of the WAE, the ABMs and their respective cut-offs. Then, the distributions of the dichotomised variables were evaluated for each dataset using the Freq procedure in SAS.

3.4.1.2. Sensitivity and specificity

Estimates of the relative sensitivity and specificity of the defined ABMs used to detect a given WAE were all estimated at the herd-level using the Freq procedure in SAS. Pearson's $\chi 2$-test was used to determine if a given ABM provided information about a WAE or not, although for data with less than 5 counts in the resulting contingency tables, the Fisher exact test was used. An ABM was considered informative if the P-value was <0.05. The relative sensitivity and the relative specificity of the defined ABMs for detecting the WAE were all estimated at herd-level. The matrix in Table 24: shows the possible combinations of the WAE and the ABMs. Also, it is stated for which WAE/ABM combinations relevant variables could be found in the available data.

3.4.1.3. Inter- and intra-observer agreement

No data were available to illustrate agreement estimates.

4. Objective 3

4.1. Literature search for relevant factors of variation

The collection of epidemiological parameters/risk factors (hereafter termed factors of variation) associated with ABMs outside the range of acceptable welfare was carried out for the following ABMs selected in Objective 2:

1. Mortality on farm: unassisted
2. Mortality on farm: euthanized
3. Somatic cell count (SCC)
4. Lameness

In literature, some ABMs and worse adverse effects (WAE) are overlapping, e.g. risk factors for an elevated SCC (ABM) and clinical mastitis (WAE) might not be differentiated. In addition, risk factors for clinical mastitis can be the same as those associated with variations in somatic cell counts. Thus, three categories of ABM/WAE were used as study variables: '***mortality***', '***mastitis/elevated SCC***' and '***lameness***'. This reduces the risk of missing important information by being too specific: overlooking such information could decrease the ability of the literature search to detect relevant epidemiological parameters for impaired welfare. The parameters were searched independantly for each category (one category at a time).

4.1.1. **Inclusion criteria**

The study aimed at identifying a link or association (or the absence of one) between an epidemiological parameter and each or all of the above categories. The search terms included those relevant to (1) mortality (on-farm, i.e. unassisted death and euthanasia), and/or (2) clinical mastitis, somatic cell count (SCC), bulk tank SCC, udder/teat/intramammary infection (IMI) and/or (3) lameness or leg/foot/hoof/claw/hock lesions/diseases or gait or locomotion (see Table 4: for specifications). 'Factors' were any variable presumed to be associated with the outcomes of the variables of interest. The population of interest was dairy cows including heifers.

Table 4: Keywords used in the primary literature search in Google Scholar.

Parameter	Search strings		
Study population	cow*, cattle, dairy		
Connector	risk, factor*, effect, associat		
Study variable	(1) mortality, death, died, euthan*, culling	(2) mastitis, somatic cell count, SCC, BMSCC, udder, intramammary, infection, IMI, health	(3) lameness, gait, locomotion, score, lesion*, hoof, claw, leg, foot

The search was done using Google Scholar as primary web search engine in addition to the files from the EFSA reports and opinions of 2009: '*Identification, validation and collection of data on animal based measures to create a database for quantitative assessment of the welfare of dairy cows*'. The EFSA files were used as review papers.

4.1.2. Exclusion criteria

Review papers were preferentially not included. Review papers and EFSA files were used to detect original papers of interest, which were the ones preferably included in order to avoid potential reporting errors in reviews. References of interest cited in an article as well as the articles citing this article were checked.

The expansive list of risk factors and other epidemiological parameters produced by the literature search was then refined by regrouping similar factors (e.g. synonyms), eliminating those that were only anecdotal, not relevant and/or not supported by sufficiently solid statistical evidence (see Table 30:).

4.2. Selection of the final parameters based on partners' opinion in the consortium

Further refinement of the list was carried out in order to retain only parameters being 1) associated with the highest impact or variability on the ABMs/WAEs, 2) easy to collect (criteria 1 and 2 should both be fulfilled), and 3) able to characterise major aspects of a population. Each partner was provided with a spreadsheet listing all parameters in one column and asked for each parameter (see Appendix N.):

a. To score the parameter according to its degree of relevance, i.e. the strength of the association, with the 3 categories of ABMs/WAEs, ranking from 1 (low relevance) to 5 (high relevance).

b. To indicate whether the parameter is/would be easy to record on farm by a "non-specialist" as well as being easy to collect routinely and to keep updated in a database (feasibility).

c. To indicate whether a parameter is useful or needed to characterise/define a specific population, e.g. 'country' helps characterise French cattle farming compared to another country, or 'stall type' helps characterise tie-stalls operations compared to loose housing operations.

Eleven partners from the consortium participated in this process (response rate 100%) and their answers were recorded. For question 'a', parameters were ranked according to their mean score. Only parameteres with a mean relevance score ≥ 3.5 ('a') AND selected as easily collectable ('b') by at least 8 out of the 11 partners were retained. For question 'c', parameters indicated as characterising a population were kept if selected by at least 8 out of 11 partners. Final parameters were those responding to criteria 'a' and 'b', plus those answering to criterion 'c'. This method of selection was adopted within the consortium during online discussions.

4.3. Availability of factors of variation in routinely collected databases

Data providers of Routinely Collected Databases in the 5 countries of the consortium were asked to fill a table to provide information on the routine collection of the final parameters and the three ABMs (from Objective 2) (Appendix O.).

Data providers were asked to add details on the availability of the parameters, the level of collection (herd/individual), in how many farms/animals a year, the annual number of sampling moments, the scale used, as well as the method of collection. If necessary, the table was translated by project partners in the language of the concerned country and data providers responses were translated back to English. Since the different databases might have had different ways of recording the same information (e.g. parity or days in milk), it was decided to leave to some extent the definition open and allow data providers to give their own definition.

5. Objective 4

5.1. Creating a data model

5.1.1. Data access and data ownership issues

The ambition in Objective 4 was to collate RCD from all member states represented in the consortium (Italy, Belgium, Denmark, France and Sweden). First, an inventory was made of relevant and potential data providers. This was performed by the national consortium partners. Subsequently, the first contact with the data providers was made by the consortium partners who asked permission to access the RCD for the purpose of this study. For Italy, two official letters explaining the background and purpose of the project and asking permission to get access to data from the ANIBAM consortium were necessary. Once permission was granted, the data were collected in two steps:

1 The first step was in overlap with Objective 3: the data providers were asked which of the in Objective 3 selected factors of variation are present in their database (Objective 3, Appendix P. specifies which data were requested)

2 For the second step, the data providers were officially asked by mail to send us their data on the selected ABMs (mortality, lameness and somatic cell count) and factors of variation/geographical information (country, region, animalID, herdID, holdingID, year, month, day, breed, production type, housing, flooring, bedding, herd size, access to pasture, milk yield, parity, days in milk, SCC, density, annual mortality rate (AMR) and lameness) for 2012

It was agreed with EFSA to collect data for a single calendar year as an appropriate compromise between a sufficiently large amount of data and feasibility of managing these data. We opted for asking data for 2012 as this was the most recent calendar year for which the data were complete in all countries. Data were asked without further transformations/aggregations, to make it easier for the data providers to send the data.

Table 5: The organisations that were contacted for the required RCD of the ABMs and factors of variation in the partner countries

	Belgium	Denmark	France	Italy	Sweden
Somatic cell count	CRV[(a)]	Danish Cattle Federation	Institute of Animal Husbandry	IZSLER[(b)]	Växa Sverige
Mortality	Rendac	Danish Cattle Federation	Bureau of Animal Protection	IZSAM[(c)]	Växa Sverige
Lameness	-	-	-	-	-

(a) CRV: Coöperatie Rundveeverbetering
(b) IZSLER: Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini" (which is also National Reference Centre for Bovine Milk Quality)
(c) IZSAM: Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale"

In all partner countries, data on two of the three selected ABMs, somatic cell count and mortality, are collected routinely (Table 5:). However, for Belgium it was not possible to provide us with the data on dairy cow mortality as no distinction is made between beef cattle and dairy cattle in the national database that collects this data. For the Belgian SCC data, it was necessary to sign a contract stating the goal of the use of the data and the period during which the data will stay in the DCF database. For Sweden, it was not possible to collect any data. This was due to their privacy legislation which prohibited the data provider to give us the data unless each individual farmer had given consent for doing so. Due to time restraints this was not possible during this project.

None of the countries routinely collects data on the third ABM, namely lameness (Table 5:).

5.1.2. Creating a model forDCF input

The first step in creating the data-model consisted of identifying which elements present in the RCD are readily available in the EFSA SSD and which need to be added. To illustrate how these elements are defined, an example is given in Table 6: . In Table 7: the data elements that were readily available are shown.

Table 6: Example of an element available in the SSD (EFSA, 2013)

Element Code[(a)]	Section Code[(b)]	Section [(c)]	Element Name[(d)]	Element Label[(e)]	Type[(f)]	S/R/C[(g)]	M[(h)]	Controlled terminology [(i)]	Description [(j)]
D.02	D	Sample taken	repCountry	Reporting country	xs:string (2)	S		COUNTRY	The country the reported data refer to (ISO 3166-1-alpha-2).

(a) An alphanumeric code providing a unique identifier for the data element. The element code is made of the section identifier code plus a progressive number.
(b) The section code identifies the entity of the SSD data model.
(c) The section describes the key entity of the SSD data model.
(d) Unique element name is provided; this is to be used for column names, field names and tags depending on the software programs, files or databases implementing the SSD.
(e) The data elements are described also by a label to be used in reports, print outs or in the graphical interfaces of the software programs that will manage the SSD.
(f) A data type is associated to each data element and it defines the values that it can contain.

(g) Single, repeatable or compound data element. It can contain S (Single) if the data element can be reported only once (generic structure: value1), R (Repeatable) if one or more values can be reported within the data element. C (Compound) is used for those data elements that are made from one optional base term plus facets or from many attributes.
(h) Mandatory elements are flagged in this column with the value M.
(i) Provides the acronym of the catalogues that can be used to populate the data element. A catalogue is a finite and enumerated set of terms intended to convey information unambiguously.
(j)Provides a short description on what the data element should contain. (EFSA, 2013)

Table 7: Elements included in the SSD that are relevant for the ANIBAM project

Element Name	Description
localOrgId	Unique identification of the local or regional or national organisation that provided the information
localOrgCountry	Country where the local organisation is placed. (ISO 3166-1-alpha-2).
progId	Unique identification code of the programme or project for which the sampling unit was taken.
progType	Type of programme recording the indicators
sampPoint	Point, in the food chain, where the indicator was recorded
progInfo	Additional specific information and comments on the sampling programme depending on specific requirements of the different data collection domains such as if the programme is used for the verification of the Salmonella reduction target, number of animal under the control program, total number of samples tested, etc.
sampUnitType	Define the level at which the reported indicator is reported
sampId	Identification code of the sample taken.
sampCountry	Country where the holding is located
sampArea	Area where the holding is located (Nomenclature of territorial units for statistics - NUTS - coding system valid only for EEA and Switzerland).
sampY	Year of sampling. In case the sampling has been performed over a period of time the start date (as year) of sampling should be reported.
sampM	Month of sampling. In case the sampling has been performed over a period of time the start date (as month) of sampling should be reported.
sampD	Day of sampling. In case the sampling has been performed over a period of time the start date (as day) of sampling should be reported.
sampMatType	Type of sample taken (e.g. food, food stimulants, animal, feed, environment; food contact material), identifying the sub-domain of the matrix catalogue to be used.
sampMatCode	Description of the sample taken characteristics using the FoodEx2 catalogue.
sampMatText	Description of the sample taken characteristics using free text.
analysisY	Year when the analysis was completed.
analysisM	Month when the analysis was completed.
analysisD	Day when the analysis was completed.
anMatCode	Encoding of the matrix only required in case of somatic cell count
paramCode	Indicate type of numerical value reported
paramText	Additional information on indicator or herd measurement
anMethText	Method of measuring indicators for example type of lameness scoring
anMethInfo	Additional specific information and comments on the analytical method depending on specific requirements of the different data collection domains such as disk concentration and diameter for antimicrobial resistance diffusion method, method sensitivity and method specificity, migration time, migration temperature, etc.
resId	Identification code result a row of the data table in the transmitted file. The result identification code must be maintained at organisation level and it will be used in further updated/deletion operation from the senders
resUnit	Unit of measurement for the values reported in Result value
resVal	Numerical value for specific measurement as categorised in paramCode and expressed in the unit specified by the element Result unit.
resQualValue	This field should be completed only if the result value is qualitative e.g. positive/ present or negative/ absent. In this case the element Result value should be left blank.

There were elements in the routinely collected data from one or more countries that were not present in the current version of the SSD. First, these elements were identified by examining the data that were asked from the data providers. The elements that were not yet included in the SSD consisted of specific housing or management elements (breed, production system, housing, flooring, bedding, pasture) and identification elements (animalID, herdID, holdingID). These elements needed to be created in cooperation with EFSA. This means that in the DCF, each element was given a specific code, name, label, type, and description as shown in Table 8: . From the files sent by the data providers, categories and terms were derived. Subsequently, an XSD file was created to accommodate uploading into the DCF. The final set of elements used for this project can be found in Appendix U.

Table 8: Elements that were added to the SSD for the ANIBAM project

Element Name	Description
animalId	Report animal ID where indicator is reported at animal level to allow all indicators for an animal to be linked
herdId	Report herd ID to allow herd-level indicators to be linked
sampHoldingId	Report holding ID to allow holding-level indicators to be linked
breed	Breed of dairy cows in herd
prod	Type of production system
housing	Housing system used on the holding
flooring	Flooring used in housing
bedding	Bedding used in housing
pasture	Dairy cows have access to pasture

5.2. Transformation methods

To be able to upload all data into a single model, some of the data elements in the data providers' files needed to be transformed. These transformations were performed in Access. Information on this process for the specific partner countries can be found underneath. Please note that the variables bedding and density were not included based on the results from Objective 3 but due to a specific request from EFSA.

Data were provided in different formats. For Italy and Belgium it was provided in Excel 2007, for Denmark in XLS and SAS (Sas7bdat) format and for France in .txt-format. All data were imported into Access 2007-2010. Using an 'Update query', data from the databases were copied into the right cells from the XSD file that was made. It was analysed which differences existed between datasets and decided what was the best possible method to merge the data. The option that provided most detail was always chosen. Since ID's were made by automatically numbering the herds, animals and holdings, no linking between datasets from different data providers was possible.

All transformations were performed using queries. The only exception is the date, for which the 'Design view' was used, and the field 'notation' was changed in yyyy, mm or dd for Year, Month and Day respectively.

Table 9: Transformations that were performed on the data

Element	Query type	Description	Used for
Region	Select & Update	Select region name/postal code, Update into NUTS code	Italy, Denmark
Breed	Select & Update	Select breed name, Update into the categories used (Holstein, Otherspecdairy, Dualpurpose and Other)	Italy, Denmark
Production type	Select & Update	Select production type, Update into the categories used (Z0216, A0C6Y and A0C6Q)	Italy
BMSCC	Update	Update by using the query: [resVal] * 1000, which multiplies the number for BMSCC in resVal with 1000	Italy
SCC	Update	Update by using the query: [resVal] * 1000, which multiplies the number for HSCC in resVal with 1000	Denmark
AMR	Update	Update by using the query: [resVal] * 100, which changes the ratio to a percentage	Italy

Regions were coded using the international Nomenclature of Territorial Units for Statistics

(NUTS) as prescribed by EFSA (Table 9:).

As regard to breed, it was chosen to categorize in the following manner:

- Holstein: all Holstein dairy cows (including Red Holsteins)
- Other specialised dairy breeds: all dairy breeds except Holstein
- Dual purpose: all breeds used for meat and milk
- Other: all breeds that could not be categorized using the above categories (for instance mixed breeds of which the breeds were not clear)

The rationale for this categorisation relates to the huge differences between breeds in the different countries. In the Italian database 79 breeds were found, while in the Danish database only 5 breeds were found. This would mean that breed would overlap greatly with country. Because dual purpose breeds were expected by the consortium to differ greatly from specialised dairy breeds, this was added as extra separate category. Any meat breeds that were found were deleted from the database.

For production type, the coding that was already present in the SSD was used, being: A0C6Q = (Intensive production) and A0C6Y (Conventional non-intensive production).

Both BMSCC and ISCC were expressed as cells/ml. The definitions of BMSCC and HSCC from Objective 2 both include measures taken during three months. However, this was not applicable to the data received from the data providers. Data was given for one sampling moment, and transformation was not possible due to the facts that the months of sampling were not always consecutive. Therefore the definitions used were for HSCC: percentage of cows with somatic cell count > 400,000 / No cows tested, and for BMSCC: bulk tank cell count in one sampling event.

Even though sample year was by default 2012, since data from that year was asked, in some datasets there were data from other years. These data were deleted.

Underneath, each database is described in detail. For each database the original column names and definitions can be found in Appendix U.

5.2.1. Italian database (IZSAM + IZSLER)

The Italian database was derived from two different sources: Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise (IZSAM) and Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini" (IZSLER). Both are Public Health Service Bodies whose task is to, amongst other tasks, perform epidemiologic monitoring of animal health. IZSAM collects mortality data from dairy herds in Italy. Somatic cell count in Italy is not collected in a centralised database, but different regions have their own entities collecting these data. For the ANIBAM project, IZSLER (which is also the Italian National Reference Centre for Bovine Milk Quality) provided SCC data, for the largest milk-producing area in Italy (mostly Lombardia and Emilia Romagna regions) where it performs the milk tests.

The SCC database consisted of 8 columns with the following column names: Date of record, Laboratory, Farm code, Bulk Tank SCC, BTSCC Geometric mean, Region, Province and Country. The database consisted of 128,328 lines in total.

In the AMR database, 22 columns were found containing all column names from Table 10: . Before transformation this database consisted of 103,208 lines in total. Access to pasture was present in the Italian AMR database, but since it was the same as the production type and did not fit the definition proposed in Objective 3, it was chosen not to upload these data.

Table 10: Availability of the data elements in the Italian databases (IZSAM + IZSLER)

Element	Availability mortality	Availability SCC	Transformation
Region	Available, textual	Available, textual	NUTS-code
AnimalID	Available, but not for the ANIBAM project since Herd-level was provided	Not available	
HerdID	Available	Available	
HoldingID	Available	Not available	
Year	Available	Available, in one cell with month and day	Put into different cells
Month	Available	Available, in one cell with year and day	Put into different cells
Day	Available	Available, in one cell with year and month	Put into different cells
Breed	Available, the most prevalent breed	Not available	Removed all meat breeds from the file, categorized the dairy breeds
Production type	Not compulsory; not available in some cases. If available three options: transhumance, intensive, extensive.	Not available	Categorised differently: A0C6Y and A0C6Q
Housing	Not available	Not available	
Flooring	Not available	Not available	
Bedding	Not available	Not available	

Element	Availability mortality	Availability SCC	Transformation
Herd size	Available, average number of cows on farm in 2012	Not available	
Access to pasture	Not available	Not available	
Milk yield	Not available	Not available	
Parity	Available, sum of overall calvings in the life of each cow present in the farm in 2012	Not available	
DIM	Not available	Not available	
Density	Not available	Not available	
SCC	Not available	Not available	
HSSC[(a)]	Not available	Not available	
BMSCC[(b)]	Not available	Available	
AMR[(c)]	Ratio between number of cows dead on farm in 2012 and Herd size	Not available	
HLAME[(d)]	Not available	Not available	

(a): High Somatic Cell Count (Percentage of cows with Somatic cell count > 400000 / No cows tested)
(b): Bulk Milk Somatic Cell Count (bulk tank cell count in one sampling event)
(c): Annual Mortality Rate
(d): Lameness at herd level (Percentage of lame animals in the herd per year)

A total of 79 breeds were represented in the Italian mortality database, of which 4 were categorised as 'other', 27 as 'dual purpose', 25 as 'other specialised dairy', 4 as 'Holstein' and 19 as 'meat'. Entries for beef cattle were removed. Breeds that are 'dual purpose' (meaning that they are farmed both for meat and milk production) were not removed. Breed was defined as the most prevalent breed on the farm, determined on the basis of the parents of every newborn animal. The complete list of breeds and how they were categorized can be found in Table 11: .

Production type in the original database was categorized as 'Extensive', 'Intensive' and 'Transhumance' (last category meaning that the animals are moved to the mountains for a part of the year). These categories were transformed to the codes present in the DCF, for the exception of transhumance. The definition of transhumance is 'the animals are moved (mostly to the mountains) for some part of the year'. This did not fit into any of the existing production codes available in the SSD. Therefore the transhumance entries were left blank. All data, both for the AMR and SCC database, were provided at Herd-level.

Table 11: Breeds found in the Italian Mortality database and their categorisation for the DCF

Breeds	Category
METICCIO/INCROCIO	Other
NON INDICATA	Other
METICCIO/INCROCIO	Other
ALTRE RAZZE	Other
ALTRE PODOLICH	Dualpurpose
ABONDANCE	Dualpurpose
AUBRAC	Dualpurpose
BELGIAN RED	Dualpurpose
BURLINA	Dualpurpose
CABANNINA	Dualpurpose
CINISARA	Dualpurpose
DEXTER	Dualpurpose
GARFAGNINA	Dualpurpose
PEZZATA NERA POLACCA	Dualpurpose
PEZZATA ROSSA D'OROPA	Dualpurpose
PEZZATA ROSSA FRIULANA	Dualpurpose
PEZZATA ROSSA DELLA MOSA	Dualpurpose
PIE ROUGE DES PLAINES	Dualpurpose
PINZGAUER	Dualpurpose
PISANA	Dualpurpose
PODOLICA	Dualpurpose
PODOLICA PUGLIESE	Dualpurpose
PONTREMOLESE	Dualpurpose
SARDO BRUNA	Dualpurpose
SARDO-MODICANA	Dualpurpose
TARENTAISE	Dualpurpose
VALDOSTANA CASTANA	Dualpurpose
VALDOSTANA PEZZATA NERA	Dualpurpose
VALDOSTANA PEZZATA ROSSA	Dualpurpose
VARZESE	Dualpurpose
VARZESE / TORTONESE / OTTONESE	Dualpurpose
ALTRE RAZZE PEZZATE ROSSE	Otherspecdairy
BRUNA ALPINA: Brown Swiss	Otherspecdairy
BRUNA	Otherspecdairy
GRIGIA ALPINA	Otherspecdairy
PEZZATA ROSSA ITALIANA SIMMENTAL	Otherspecdairy
AGEROLESE	Otherspecdairy
ALTRE RAZZE PEZZATE NERE	Otherspecdairy
ANGLER	Otherspecdairy
BIANCA VAL PADANA / MODENESE	Otherspecdairy

Breeds	Category
BRETONNE PIE-NOIRE	Otherspecdairy
DANISH RED	Otherspecdairy
GRIGIA ALPINA	Otherspecdairy
GRIGIA ALPINA / GRIGIA VAL D'ADIGE	Otherspecdairy
GRIGIA VAL D'ADIGE	Otherspecdairy
JERSEY	Otherspecdairy
MODICANA	Otherspecdairy
MONTBELIARD	Otherspecdairy
PEZZATA ROSSA CECOSLOVACCA	Otherspecdairy
PEZZATA ROSSA ITALIANA SIMMENTAL	Otherspecdairy
REGGIANA	Otherspecdairy
RENDENA	Otherspecdairy
SARDA	Otherspecdairy
SICILIANA	Otherspecdairy
TARINA	Otherspecdairy
SWEDISH RED AND WHITE	Otherspecdairy
FRISONA	Holstein
FRISONA ITALIANA (PEZZATA NERA)	Holstein
FRISONA OLANDESE	Holstein
FRISONA PEZZATA ROSSA	Holstein
ANGUS	Meat (removed)
BLONDE D'AQUITAINE/GARONNESE	Meat (removed)
BLU BELGA	Meat (removed)
BRAHMAN	Meat (removed)
BRUNA SARDA	Meat (removed)
CALVANA	Meat (removed)
CHAROLAIS	Meat (removed)
CHIANINA	Meat (removed)
GALLOWAY	Meat (removed)
GUASCONE	Meat (removed)
HIGHLAND	Meat (removed)
LIMOUSINE	Meat (removed)
LONGHORN	Meat (removed)
MARCHIGIANA	Meat (removed)
MAREMMANA	Meat (removed)
PIEMONTESE	Meat (removed)
ROMAGNOLA	Meat (removed)
SALERS	Meat (removed)
SPRINZEN PUSTERTALER	Meat (removed)

5.2.2. Danish database

The Danish data were provided by the Danish Cattle Federation (Knowledge Centre for Agriculture). The majority of Danish dairy farms (~95%) is represented in the SCC database and all Danish herds are in the mortality database. Data were sent in two different datasets: one containing mortality and one containing SCC. Also the data necessary to calculate DIM were sent separately. Calculation of DIM was supposed to be performed by using the last calving date. This calculation was not possible due to the fact that it could not be determined (based on the information supplied by the data provider) whether the animals were still in milk, or slaughtered.

In the original SCC database the following column names were found: ID, Cow ID, Herd ID, Date of milk control, and Milk control number per farm, kg of milk on the test day and the individual SCC for a total number of 3,920,127 lines. Kg of milk on test day could not be used for the element Milk yield, since this did not meet the definition set in Objective 3. The SCC dataset was inserted in two ways: one on animal level with the Individual SCC and one on farm level with HSCC.

The AMR file contained Herd ID, Country, Postal code, breed code, breed name, number of cows, number of cows euthanized, number of cows dead and sum of the cows euthanized and dead and mortality. The total number of lines was 3,670.

Table 12: Availability of the data elements in the Danish databases

Element	Availability mortality	Availability SCC	Transformation
Region	Available, Postal code	Not available	
AnimalID	Available	Available	
HerdID	Available	Available	
HoldingID	Not available	Not available	
Year	Not available	Available, in one cell with month and day	Put into different cells
Month	Not available	Available, in one cell with year and day	Put into different cells
Day	Not available	Available, in one cell with year and month	Put into different cells
Breed	Available – names and codes	Not available	Categorise
Production type	Not available	Not available	
Housing	Not available	Not available	
Flooring	Not available	Not available	
Bedding	Not available	Not available	
Herd size	Available	Not available	
Access to pasture	Not available	Not available	

Element	Availability mortality	Availability SCC	Transformation
Milk yield	Not available	Available for most herds and cows (~95%)	
Parity	Not available	Not available	
DIM	Available in separate file – to calculate	Not available	Calculation was not possible
Density	Not available	Not available	
SCC	Not available	Available, animal level, per date per farm	Multiply with 1000
HSSC[(a)]	Not available	Not available	
BMSCC[(b)]	Not available	Not available	
AMR[(c)]	Available	Not available	
HLAME[(d)]	Not available	Not available	

(a): High Somatic Cell Count (Percentage of cows with Somatic cell count > 400000 / No cows tested)
(b): Bulk Milk Somatic Cell Count (bulk tank cell count in one sampling event)
(c): Annual Mortality Rate
(d): Lameness at herd level (Percentage of lame animals in the herd per year)

Table 13: Breeds found in the Danish database

Breed name in database	Category
Holstein	Holstein
Red Holstein	Holstein
Crossbreed	Other
More than one breed	Other
Red Danish dairy breed	Otherspecdairy
Jersey	Otherspecdairy

The Danish SCC data was provided at animal level, with each line representing an animal at a certain date in a certain farm. Because we chose for as much detail as possible and because of the limited time, we decided to only insert the animal-level data in the DCF. This also made sure that SCC, HSCC and BMSCC data were all inserted in the DCF.

5.2.3. French database

The French data were provided by the Institute of Animal Husbandry. Only the mortality data were sent. There were 42 columns in the original dataset, of which the ones applicable for the ANIBAM project were country code, animal ID, breed code, birth date, date of death and date of slaughter. The total number of lines was 1,142,277.

Table 14: Availability of the data elements in the French database

Element	Availability	Transformation
Region	Not available	
AnimalID	Available, numerical	Add country code and put into different cells
HerdID	Available, the first four digits of the AnimalID	Put into different cells
HoldingID	Not available	
Year sampling	Available, in one cell with month and day	Put into different cells
Month sampling	Available, in one cell with year and day	Put into different cells
Day sampling	Available, in one cell with year and month	Put into different cells
Breed	Available, in code	
Production type	Not available	
Housing	Not available	
Flooring	Not available	
Bedding	Not available	
Herd size	Not available	
Access to pasture	Not available	
Milk yield	Not available	
Parity	Not available	
DIM	Not available	
Density	Not available	
SCC	Not available	
HSSC[(a)]	Not available	
BMSCC[(b)]	Not available	
AMR[(c)]	Date of death available	Calculate AMR
HLAME[(d)]	Not available	

(a): High Somatic Cell Count (Percentage of cows with Somatic cell count > 400000 / No cows tested)
(b): Bulk Milk Somatic Cell Count (bulk tank cell count in one sampling event)
(c): Annual Mortality Rate
(d): Lameness at herd level (Percentage of lame animals in the herd per year)

Calculation of AMR was not possible, due to the fact that herd size was not available for the French data. Therefore this dataset was not inserted into the DCF.

5.2.4. Belgian database

The Belgian database was provided by CRV, which is a company that, amongst other tasks, performs the 'milk production registration' (MPR) in Belgium. In MPR, data are collected on for instance fertility, milk yield and individual SCC. A total of 2215 Belgian dairy farms (21% of the total number of farms) are included in the CRV database; all are located in Flanders. The majority of the Belgian dairy farms are located in Flanders (58%). In the original database that was used, the column names were: date the sample was taken, number of animals tested on SCC, % of animals with a SCC higher than 400000 and % of new infections based on the average SCC. The total number of lines was 24946.

Table 15: Availability of the data elements in the Belgian database

Element	Availability	Transformation
Region	Not available	
AnimalID	Not available	
HerdID	Available	
HoldingID	Not available	
Year sampling	Available, in one cell with month and day	Put into different cells
Month sampling	Available, in one cell with year and day	Put into different cells
Day sampling	Available, in one cell with year and month	Put into different cells
Breed	Not available	
Production type	Not available	
Housing	Not available	
Flooring	Not available	
Bedding	Not available	
Herd size	Not available	
Access to pasture	Not available	
Milk yield	Not available	
Parity	Not available	
DIM	Not available	
Density	Not available	
HSSC(a)	Available	
BMSCC(b)	Not available	
AMR(c)	Not available	
HLAME(d)	Not available	

(a): High Somatic Cell Count (Percentage of cows with Somatic cell count > 400000 / No cows tested during)
(b): Bulk Milk Somatic Cell Count (Percentage of bulk tank measures with Somatic cell count > 400000)

(c): Annual Mortality Rate
(d): Lameness at herd level (Percentage of lame animals in the herd per year)

5.2.5. Insertion into the DCF

After the transformation for all datasets was complete, the datafile was transformed into an XML-file. This xml file was called 'result' in all cases, and was opened using Notepad. In all XML-files dataroot was replaced with dataset, and 'xmlns:od="urn:schemas-microsoft-com:officedata" generated="2014-06-04T10:48:17"' was removed behind dataroot. Once this was done, the dataset was ready to insert into the DCF via https://dcf.efsa.europa.eu/dcf-war/. In the 'transmissions' tab, the progress of the uploading is shown (processing, rejected or validated). When the file was rejected, the details were shown in the 'ack message'. Here the first error that the DCF encounters is shown. Most errors were to do with the XSD file (mostly that certain elements were set as mandatory while they were not present in the data base), with elements not being present in the database or with typo's. These were all solved either by adjusting the XSD, filling in the missing elements or correcting typo's.

6. Objective 5

This section will describe the general approach to the analyses of the ability of the selected ABMs and factors of variation to predict the overall welfare status. A description of the measures of the overall welfare and the available data and the statistical methods will be given.

6.1. Data description

Five research datasets were available to the consortium for the analyses of the ability of the chosen ABMs to predict the overall welfare status. The IZSLER/CReNBA dataset stems from an Italian study; the Burow and the Otten data are data from two Dansih studies and finally, data collected in the Welfare Quality® from France and Belgium were available.

In Table 16: the identified ABMs and factors of variation in the different datasets are listed.

Table 16: ABMs and factors of variation found in the available datasets. Xs in brackets are variables in the dataset, but either very skewed data (e.g. breed) or no biological meaning (e.g. region in IZSLER/CReNBA, Otten/Burow)

	IZSLER/CReNBA	Burow	Otten	WQ® France + Belgium
Mortality	X	X	X	X
SCC	X	X	X	X
Lameness	X	X	X	X
Region	(X)	(X)	(X)	X
Production type (Organic/Conventionel)		X	X	
Herd size	X	X	X	X
Housing	X			X

	IZSLER/CReNBA	Burow	Otten	WQ® France + Belgium
Floor	X			
Pasture			X	X
Milk production	X	X	X	
Breed		(X)	(X)	X
Parity				
DIM				

6.1.1. The IZSLER/CReNBA data

The dataset from IZSLER/CReNBA contains data from an Italian welfare assessment protocol for dairy cows kept in loose housing systems. The protocol was developed by Dr. Luigi Bertocchi and his co-workers. CReNBA is the Italian acronym that stands for National Reference Centre for Animal Welfare, which is located in Istituto Zooprofilattico Sperimentale della Lombardia ed Emilia Romagna "Bruno Ubertini" (IZSLER, Brescia). The protocol and scores are not published yet, so we can only refer to the overall principles in the score calculation. For the current analyses, we will not use the score values but the percentage, obtained by each herd, calculated on the available range score (between the minimum and the maximum possible scores of the system). The protocol is based on ABMs and risk factors identified in the EFSA recommendations, in the WQ® system, in the EU Draft revised Recommendations concerning Cattle (8th revision, 24th September 2009) and in the European and Italian legislations. It consists of 74 indicators, divided into 4 different areas (A, B, C and D): Management and personnel (area A), Structures and equipment (area B), Animal based measures (area C) and Control of environmental conditions and alarms (area D).

Area A contains variables describing herd management and personnel, e.g. handling, feeding ration, availability of water for all animals, cleaning of structures and number of inspections. Area B contains variables related to resources, e.g. housing type, floor and bedding, number of cubicles, feeding places and water points per cow. Area C comprises animal based welfare indicators such as lameness, udder health and skin lesions. Finally, area D contains variables related to the herd environment, e.g. temperature and alarm system. For each of these 4 areas, a weighted sum of each indicator is calculated to obtain an area score.The final overall welfare measure is calculated as the simple sums of the 4 area scores. A handbook in Italian about the IZSLER/CReNBA welfare assessment protocol for dairy cows in loose housing systems is available; early it will be ready a consultation text in English, free on IZSLER website (www.izsler.it).

The dataset includes the results of the welfare assessment of 608 randomly selected Northern Italy dairy herds. The analysed herds were very different in herd size, from a minimum of 12 animals to a maximum of 1,413 animals.The percentages of the 4 area scores and the overall welfare score percentage are on a continuous scale. The three ABMs (mortality, SCC and lameness) are on a categorical scale with three levels each. Furthermore, the dataset includes two ordinal factors of variation, "Housing system" and "Floor type" and two continues factors of variation, "Milk production" and "Herd size" (see Appendix M. for further information and definitions of the variables).

6.1.2. The Otten data

Three different animal welfare indexes are calculated in the Otten data. Each of these is analysed separately in the analyses. Details can be found in Otten (2014), however, the indexes are in the process of publication (peer review) and thus we can only refer to the principles in the calculations of the three indexes. The register-based and the resource-based indexes are based on non-graded measures and calculated as weighted, additive indexes. The resource-based index comprises indicators describing the system and the resources in the herd. A total of 127 indicators is recorded and combined into 16 main indicators which are then combined into the overall resource-based index using weighted sums. The register-based index comprises welfare indicators regarding mortality, milk yield and quality, reproduction, disease and abattoir data found in the Danish Cattle Database. Here, the register-based index is calculated as a weighted sum of 24 selected indicators. The animal based welfare index comprises clinical, animal based measures. It is calculated as a weighted sum, however, the model is extended, so that the severity of the ABMs is taken into account and herd prevalences are adjusted to decrease the influence of extreme values. The protocol used for the animal based index is modified from the WQ® protocol and fitted to Danish conditions. It consists of ten clinical and two behavioural measures. For all three indexes, weights are assigned by expert opinion and a larger value of the indexes represents a poorer welfare.

The dataset includes 72 random selected Danish dairy herds with more than 100 cows, all housed in loose housing systems with cubicles. The three ABMs (mortality (annual mortality rate), SCC (mean bulk milk SCC last 12 months) and lameness (percentage of moderately and severely lame cows at visit)) are all on a continuous scale, however for the current analysis the ABM variables are transformed into categorical scales. Regarding the ABMs mortality and SCC the thresholds defined in objective 2 were used. Hence, mortality was dichotomised at the level 5% and SCC at 300,000. No herds had a SCC of more than 400,000 and therefore this threshold was not relevant in these data. For the ABM lameness, using the threshold suggested in Objective 2 was not meaningful due to the distribution of the data. Instead, the ABMs were dichotomised based on the median (median = 28%). Furthermore, the dataset includes the following four factors of variation: "Access to pasture" and "Production type" that were both dichotomous variables, "Breed" includes 4 levels and is on an ordinal scale and "Region" is given on a nominal scale (see Appendix M. for further information and definitions of the variables). However, the variables "Access to pasture" and "Production type" were confounded as all organic farms in Denmark are legally obliged to use pasture grazing. Therefore, only "Production type" was used in the analyses. Due to the size of Denmark and the relatively homogeneity of the climate and production types acrros the country, the variable "Region" was omitted from the statistical analyses.

6.1.3. The Burow data

The overall welfare measure in the Burow data is a modification of the Welfare Quality® assessment and is calculated as a multidimensional animal welfare index based on weighted sums, however, here the severity of the ABMs are taken into account and herd prevalences are adjusted to decrease the influence of extreme values. The final welfare assessment protocol comprises 17 resources- and animal based measures. As the study is published and the details regarding the index can be found in Burow et al. (2013). In the current analyses we used what is referred to as the welfare index during summer.

All the explanatory variables used in these analyses are defined equivalently to variables in the Otten data and detailed descriptions can be found in the section above or in appendix. However, as all the herds used by have access to pasture, this variable is omitted. Hence, the analysis on the Burow data only contains 31 herds.

6.1.4. The Welfare Quality® data

The Welfare Quality® data includes the criteria and principle scores as well as the overall classification. The measures included in the calculations of these scores and the aggregation procedure are described in Welfare Quality® (2009). Briefly, a long list of measures (typically scored on a three-point scaling ranging from 0 to

2) is aggregated into twelve criteria scores which are then aggregated into four principle scores. Based on the principle scores an overall classification is assigned to the herd.

Two different WQ® datasets were available. The first data stemmed from a French study of 131 commercial dairy herds. The other dataset was from Belgium and comprised data from 64 Belgian commercial dairy herds. All the herds in both datasets were classified as either 'Acceptable' or 'Enhanced'. Using the overall classification as the outcome was tested (data not shown) but none of the variables were able to detect any differences between the two levels. It was therefore decided to analyse at the level of the principles instead.

The four principle scores Health, Housing, Behaviour and Feeding all were on a continuous scale. The three AMBs (mortality (annual mortality rate), SCC (% cows with SCC >400,000 during the last 3 months) and lameness (percentage of moderately and severely lame cows at visit)) were all on a continuous scale, however for the current analysis the ABM variables were transformed into categorical scales. The thresholds defined in objective 2 were used. Hence, mortality was dichotomised at the level 5% and the % cows with high SCC at 10% and lameness at 8%. Furthermore, the dataset included the following four factors of variation: breed (binary, milk/double purpose), pasture access (binary, yes/no), housing (binary, cubicle/strawyard) and herd size (continuous, number of cows).

6.2. Overall descriptive comparison of the available welfare assessment systems

Even though the four assessment systems use some of the same measures they differentiate substantially in other ways. The IZSLER/CReNBA welfare assessment protocol is composed of a variety of resource and animal based measures to be collected during the farm visit, whereas the Welfare Quality® programme primarily collect animal based indicators. The 'Modified Welfare Quality' protocol is based on the Welfare Quality® protocol but uses fewer measures and a simpler aggregation of measures. The Otten animal welfare indexes (AWI) was developed to compare the value of three different sources of welfare indicators. Therefore, the AWI has three different overall welfare measures based on either resource based, animal based or register data based measures.

The assessment systems differ also in the way indicators are aggregated. The AWI based on register- and resource-based measures are both calculated as simple additive and weighted indexes. The Burow index and the AWI based on animal based measures are also weighted sums, however, here the severity of the ABMs are taken into account and herd prevalences are adjusted to decrease the influence of extreme values. In the AWI and the Burow indexes a larger value of the index depicts a poorer welfare. The IZSLER/CReNBA system (more details can be found in Section 6.1.1) comprises 74 indicators allocated into four main topics, referred to as "areas"; one area consists of animal based measures and three areas comprises different resource based measures (related to management risk factors, housing risk factors and general - mostly environmental- risk factors). Each resource-based indicator is multiplied by a weight depending on the severity and on its relevance in terms of welfare consequences. Each animal based measure is multiplied by a weight depending on its relevance in identifying a welfare issue. A final weighted sum is calculated in order to obtain a score for each area. Then, the overall welfare score is calculated as the sum of the four area scores with no weighting applied to this step. In the IZSLER/CReNBA system the highest score is given to the herds with the best welfare. While the overall score allows comparison between farms, the 4 partial scores can be used actively to point out strong and weak areas within each assessed farm. As previously described, the WQ® scores are calculated in a more complex hierarchical process allowing moderate compensation within the criteria level but not between criteria or principles.

6.3. Statistical analyses

Analyses were performed using the software package R (R version 3.0.3 (2014-03-06)).

In each dataset, the overall welfare measure were identified and likewise, as many as possible of the selected ABMs (SCC, mortality and lameness) and the selected factors of variation (parity, housing system, floor type, DIM, access to pasture, milk production, herd size, breed, geographical region, organic dairy production) were identified. If the ABMs were given at a continuous scale, they were dichotomised. If it was biologically meaningful, this was done according to the thresholds suggested in objective 2 (see Table 23:). Otherwise, the dichotomisation was data driven using the median as the threshold. Initially, the overall welfare measure was described regarding its mean, SD and distribution and a frequency distribution were presented. The distribution of the overall welfare measure within the levels of the ABMs and the categorical factors of variation was illustrated by boxplots. The continuous factors of variation were plotted against the overall welfare measure to illustrate any correlation between them.

For benchmarking of the herds' welfare status in the different datasets, binary outcomes were defined as binary measures of the overall welfare. Benchmarking were done at the level of the median, the 25th (or 75th) and the 10th (or 90th) percentiles. For example, for an overall welfare measure where a low score indicates poor welfare, the outcome was constructed by splitting the overall welfare measure at P25. The resulting binary variable thus indicated whether the herd was observed to be among the 25% 'worst' herds when ranking herds according to their welfare score. For each welfare measure system, predictions of the expected welfare status were made by fitting nine different logistic regression models for each outcome level (median, P25/P75, P10/P90). That is: One model for each of the three ABMs including only one ABM at a time; three models with all possible two-way combinations of the three ABMs; one model including all three ABMs in combination; one model including all additive effects of the three ABMs combined with as many of the factors of variation that can be identified in the data; and one model also including significant two-way interactions. Models with multiple explanatory variables were reduced using stepwise backward elimination sequentially removing explanatory variables with a *P*-value > 0.05 from the model (Dohoo et al., 2003). Furthermore, for each outcome level three additional models were constructed by fitting conditional inference tree models (using R packages 'party' and 'partykit'). Coefficients from the models were inspected (data not shown) in order to confirm that associations between an ABM and the overall welfare measure was as would be biologically expectable. That is: high levels of mortality, SCC and lameness would be associated with poorer welfare. Regarding the factors of variation, the coefficients were inspected (data not shown) in order to establish the nature of any detected associations.

In order to compare model results and thereby evaluate the use of ABMs for benchmarking welfare and the value of adding the factors of variation different model evaluation techniques were applied. The overall fit of the models with the same outcome was compared by Bayes Information Criteria (BIC). Also, all fitted models (except the tree models) were cross validated and the prediction errors were calculated using Leave-One-Out-Cross-Validation (LOOCV, R package 'Boot'). The ability of the models to correctly predict the welfare status of the herds was assessed by performing a ROC analyses (R package 'pROC'). The area under the curve (AUC) and the belonging 95% confidence interval was cross validated using Leave-One-Out-Cross-Validation (R package 'cvAUC'). The effect of combining the different explanatory variables was then compared by comparing the cross validated AUC under the different ROC curves using the χ^2 test developed by DeLong et al. (1988) (roc.test function, R package 'pROC').

The value of BIC decreases when the unexplained variation in the dependent variables and/or the number of explanatory variables decreases. Hence, lower BIC implies either fewer explanatory variables, better fit, or both. For the conditional inference tree model, the BIC is not relevant. The MSPE is the expected value of the squared difference between the fitted values and the function. Here, the MSPE is calculated based on the LOOCV cross validation procedure and is the mean MSPE from the series of model fits run in the cross validation. The cross validation is used to evaluate how the results of a statistical analysis will generalize to an independent dataset and the resulting MSPE thus is an indicator of the predictive value of the given model; the lower the prediction error the fewer false predictions. The Area Under the Curve (AUC) in the ROC curves is another indicator of the predictive value. When used to compare models, the AUC depicts the probability that a model will rank a randomly chosen positive case higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative'). Thus, the closer the AUC is to 1 the more likely the model is to produce true positives.

RESULTS

7. Objective 1

7.1. Identification of the WAEs

The median severity and prevalence ratings by the consortium experts were calculated for each adverse effect listed. There were no additional adverse effects described by the experts. Table 17: shows the median severity and prevalence for each adverse effect.

Table 17: Median ratings by the survey respondents for severity and prevalence for each adverse effect

Severity	Prevalence	Adverse Effect
4	0	Mortality (unassisted)
		Mortality (euthanised)
3	3	Foot disorders & associated effects e.g. lameness (cubicle and tie-stall)
		Leg injuries & associated effects (cubicle and tie-stall)
		Behavioural disruption -flooring conditions/space (including fear, pain) (cubicle and tie-stall)
3	2	Foot disorders & associated effects e.g. lameness (straw-yard)
		Exhaustion associated with prolonged high metabolic demand
		Leg injuries & associated effects (straw-yard)
		Behavioural disruption-feeding (including social stress, pain, hunger, exhaustion, fear, frustration) (cubicle)
		Behavioural disruption-rest (incl too little rest, pain, fear) (cubicle and tie-stall)
		Behavioural disruption -flooring conditions/space (including fear, pain) (straw-yard)
3	1	Clinical Mastitis & associated conditions (all systems)
		Reproductive disorders e.g. reduced fertility, dystocia, infections
		Behavioural disruption-feeding (including social stress, pain, hunger, exhaustion, fear, frustration) (tie stall, straw-yard)
		Pain or fear (due to handling, milking, dehorning/other 'surgery', downer cows) or frustration (due to management factors) (cubicle, straw-yard)
3	0	Pain or fear (due to handling, milking, dehorning/other 'surgery', downer cows) or frustration (due to management factors) (tie stall)
		Metabolic disorders e.g. SARA (sub-acute ruminal acidosis), lipomobilisation syndrome, abomasal displacement (all systems)
2	2	Behavioural disruption-rest (incl too little rest, pain, fear) (straw-yard)
	1	Thermal discomfort (including cold and heat) (all systems)
		Mortality (slaughtered, including prior to end of expected productive life)
		Respiratory distress, pain, discomfort - air quality (all systems)

By professional judgement within the consortium, adverse effects with a severity of 4, or a severity of 3 and prevalence score 2 (10-24%) or 3 (25-49%) were considered as the worse adverse effects. Thus, based on the results shown in Table 17: , the worse adverse effects were found to be:

- Mortality – unassisted ;
- Mortality – euthanised;
- Leg injuries;

- Behavioural disruption – flooring;
- Foot disorders;
- Behavioural disruption – Rest;
- Exhaustion;
- Behavioural disruption – Feeding (cubicles only).

7.2. Identification of routinely collected ABMs

7.2.1. ABMs collected routinely on most farms by qualified personnel

The survey about routinely collected ABMs resulted in a total of 21 responses. The respondents (and response rates in brackets) were: 10 researchers (67%); 6 producer organisation representatives (50%); 1 retailer (14%); 1 practising veterinarian (25%); 1 CA (25%); 2 NGOs (100%). The response rates for each country were: Italy 100% (4/4); Belgium and France 67% (2/3 each); Netherlands and Sweden 50% (2/4 each); Denmark 40% (2/5); UK, Germany, and Poland 33% (5/15, 1/3, 1/3, respectively). Proportionately, a large number of researchers were surveyed, reflecting the composition of the consortium (also researchers) and their recommendations for respondents. The unusually large number of surveys sent to UK recipients was due to the high number of recommendations received. One reason for this could be the intensity of activity by researchers, industry, NGOs and retailers in the development of ABMs for the UK dairy industry. There was a low response rate by the (UK) retailers, practising veterinarians and competent authorities. The reason for the low response by the veterinarians was (informally) reported as due to conflicts with other work commitments. The reasons for the low response rates by the remaining groups are not clear.

The total number of ABMs being recorded (over all 21 respondents) was 58. Four new ABMs were added by the respondents. These were: calving ease; signals for welfare of the Swedish Dairy Association; slaughter house data (post mortem inspection data); and medical treatment data. There was only one response for each of these categories. Figure 1: shows the frequencies of recording of each ABM by the respondents (data based on all responses).

Only three ABMs were recorded relatively frequently. These were ‘Evidence of mastitis’, ‘Number of dead animals (unassisted)’, and ‘Number of dead animals (euthanised)’ with 67%, 52% and 48%, respectively, of respondents indicating that these ABMs were being recorded routinely. Less than 25% of the respondents reported that the remaining ABMs were collected routinely.

‘Evidence of mastitis’ was reported to be collected in seven countries (not Poland and Italy), ‘number of dead animals (unassisted)’ was reported to be collected in six countries (not Poland, Netherlands and Belgium), and ‘number of dead animals (euthanised)’ was reported to be collected in five countries (Denmark, France, Italy, Sweden and the UK).

The low and variable numbers of respondents from the different professional groupings (see Methods section) precludes further sensible analysis based on area of expertise.

The descriptions of the ABMs recorded (as provided by the respondents) are shown in Appendix D. These descriptions were provided for 42 of the 54 responses. It is clear from these data that there is little consistency between respondents in definitions of the ABMs, and a lack of precision in the definitions. For example, 12 respondents provided a description for ‘evidence of mastitis’, and only two were identical. A similar pattern was evident for all the other ABMs with multiple responses. Thus, the descriptions of the ABMs, as written, are too inconsistent to be useful in identifying WAEs.

The description of the person recording the ABM was provided in 38 (70%) of cases. For these cases, the recordings were undertaken by farmers (39%), competent authority (32%) or other (29%).

Of the 58 responses where an ABM was recorded, there were 17 instances only where both the herd prevalence and percentage of ABMs in the database were recorded (two each for 'Evidence of mastitis', and 'Number of dead animals (unassisted)', two each for 'Measures of lameness' and 'Number of dead animals (unassisted)', and one or zero each for the remaining ABMs). These response rates were too low to allow further meaningful analysis of the reliability of the recording of ABMs in the databases.

7.2.2. ABMs routinely collected on a sample of herds reasonably representative of the population

The total number of ABMs being recorded (over all 21 respondents) was 180. Nine new ABMs were added by the respondents. These were: calving ease; avoidance distance; number of broken tails, response of cattle to stockperson; cows needing further care; lameness management; verifying self-assessment; avoidance test; water trough space (although it is noted that this is not an ABM). There was one response for each of these categories.

Figure 2: shows the frequencies of recording of each ABM (data based on all responses). Six ABMs were recorded relatively frequently. These were 'Evidence of mastitis', 'Measures of nutritional status', 'Measures of lameness', 'Cleanliness score', 'Numbers of foot lesions or infections' and 'Numbers of hock, knee, skin lesions and swellings' with 71%, 67%, 62%, 57%, 57%, and 52%, respectively, of respondents indicating that these ABMs were being recorded.

'Evidence of mastitis' and 'Numbers of foot lesions or infections' were reported to be collected in eight countries (not Poland), 'Measures of lameness' was reported to be collected in seven countries (not France and Poland), and 'Cleanliness score' and 'Numbers of hock, knee, skin lesions and swellings' were reported to be collected in five countries (not Belgium, France, Germany and Poland).

Forty three percent of respondents noted that 'Number of dead animals (unassisted)', and 'Number of dead animals (euthanised) were being recorded. The remaining ABMs were recorded by less than 33% of respondents.

The low and variable numbers of respondents from the different professional groupings (see Methods section) precluded further sensible analysis based area of expertise.

The descriptions of the ABMs recorded (as provided by the respondents) are shown in Appendix D. These descriptions were provided for 127 of the 180 responses. The descriptions were provided most frequently for the following ABMs: 'Evidence of mastitis', 'Measures of nutritional status', 'Measures of lameness', 'Cleanliness score', 'Numbers of foot lesions or infections' and 'Numbers of hock, knee, skin lesions and swellings'. It is clear from these data that there is little consistency between respondents in definitions for the ABMs, and a lack of precision in the definitions. For example, 10 respondents provided a description for 'Cleanliness score', and none were identical. A similar pattern was evident for all the other ABMs with multiple responses.

The description of the person recording the ABM was provided in 173 (96%) of cases. For these cases, the recordings were undertaken by researchers (40%), assurance assessors and veterinarians on-farm (11% each), farmers (9%), competent authority (9%), other veterinarians (2%) or other (18%).

Of the 180 responses indicating recording of an ABM, there were two instances only where both the herd prevalence and percentage of ABMs in the database were recorded (one each for ‘evidence of mastitis’, and ‘numbers of foot lesions or infections’). These response rates were too low to allow further meaningful analysis of the reliability of the recording of ABMs in the databases.

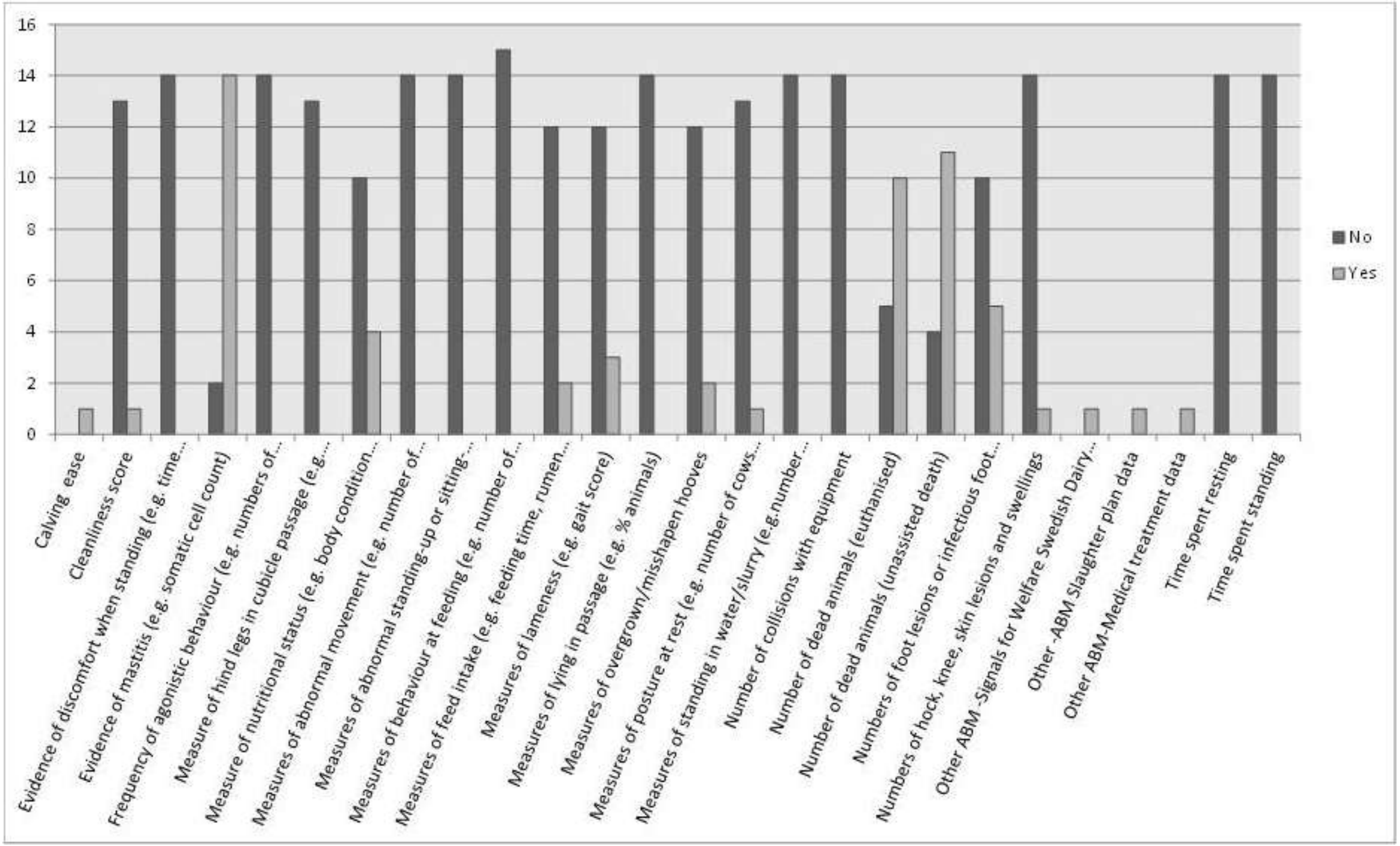


Figure 1: For ABMs collected routinely on most farms by qualified personnel, the frequency of recording animal based measures (ABMs) by the survey respondents.

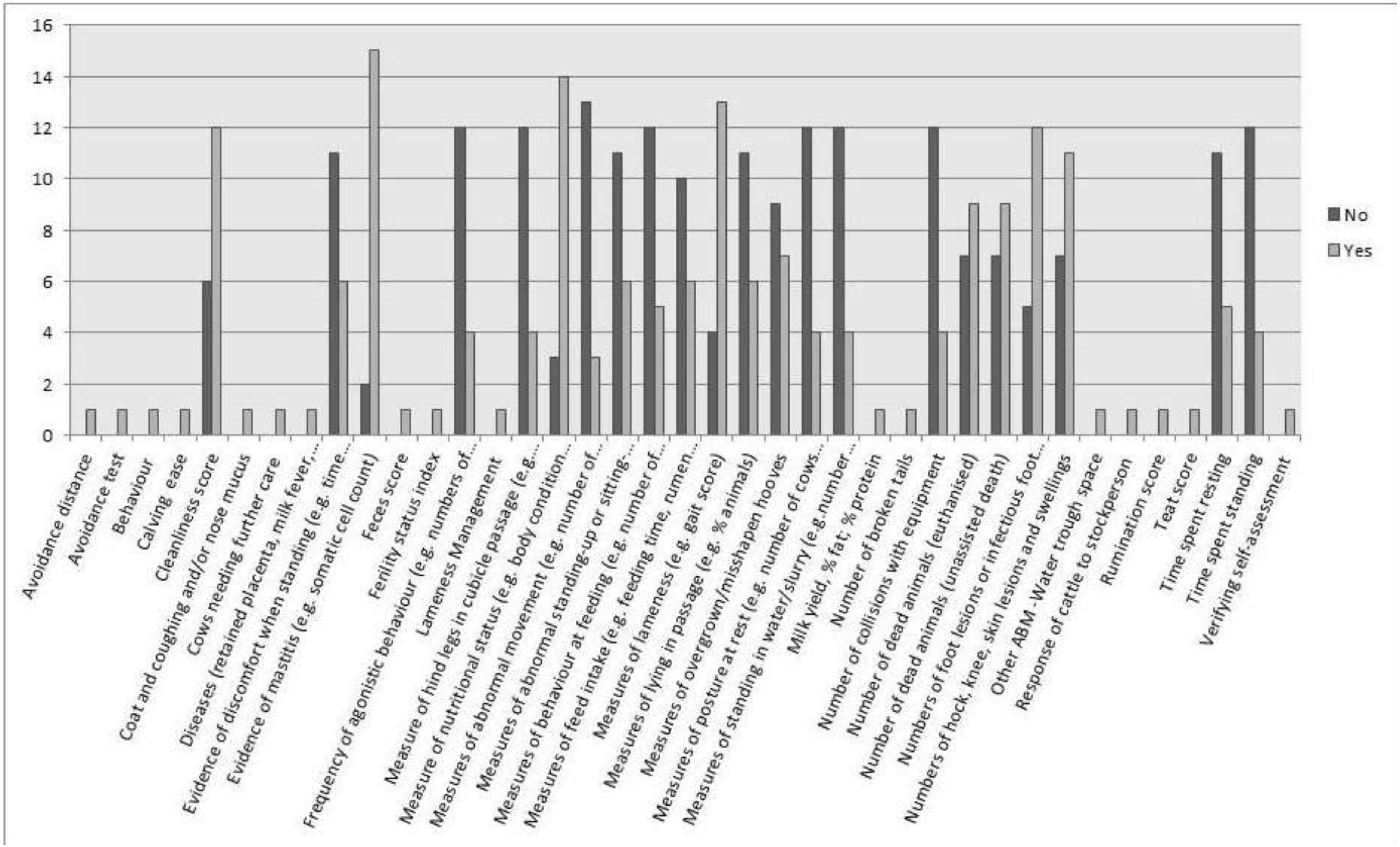


Figure 2: For ABMs routinely-collected on herds reasonably representative of the population, the frequency of recording animal based measures (ABMs) by the survey respondents.

7.2.3. Evaluation of ABMs already collected in the field and relevant for detecting the WAEs

From the survey of ABMs collected routinely on most farms by qualified personnel and routinely collected on herds reasonably representative of the population a total of eight unique ABMs were recorded sufficiently frequently to warrant consideration for detection of the WAEs (and to be considered as candidates for further evaluation in objective 2). The process described in Section 2.2 was used to determine the recommended ABMs for detecting WAEs.

The diagrams resulting from the expert elicitation procedure showing the linkages between ABMs and WAEs, and amongst WAEs are shown in Appendix E. For the various WAE, associations between ABMs and WAE shown in Part a) of Appendix E. were:

a) 'Mortality – unassisted' was reported as being associated with just one ABM ('No. of deaths – unassisted')

b) 'Mortality – euthanised' was reported as being associated with two ABMs ('No. of deaths – euthanised' and 'No. of foot lesions')

c) 'Foot disorders' was reported as being associated with four ABMs ('No. of deaths – euthanised', 'No. of foot lesions', 'Measures of lameness' and 'No. of leg lesions/swellings')

d) 'Behavioural disruption: flooring' was reported as being associated with four ABMs ('No. of foot lesions', 'Measures of lameness', 'No. of leg lesions/swellings', 'Cleanliness score')

e) 'Leg disorders' was reported as being associated with four ABMs ('Evidence of mastitis', 'No. of foot lesions', 'Measures of lameness', 'No. of leg lesions/swellings')

f) 'Behavioural disruption: rest' was reported as being associated with five ABMs ('Evidence of mastitis', 'No. of foot lesions', 'Measures of lameness', 'No. of leg lesions/swellings', 'Cleanliness score')

g) 'Exhaustion' was reported as being associated with three ABMs ('No. of deaths – euthanised', 'No. of leg lesions/swellings', 'Measures of nutritional status')

h) 'Behavioural disruption: feeding' was reported as being associated with four ABMs ('Evidence of mastitis', 'No. of foot lesions', 'Measures of lameness', 'Measures of nutritional status').

The associations amongst WAEs, shown in Part (b) of Appendix E. were:

a) 'Mortality – euthanised' was reported as being associated with 'Foot disorders', 'Exhaustion'

b) 'Mortality – unassisted' was reported as being associated with 'Foot disorders', 'Leg disorders' and 'Exhaustion'

c) 'Behavioural disruption: rest' was reported as being associated with 'Foot disorders', 'Exhaustion', 'Leg disorders', 'Behavioural disruption: flooring', and 'Behavioural disruption: feeding'

d) 'Exhaustion' was reported as being associated with 'Mortality – euthanised' , 'Mortality – unassisted', 'Foot disorders', 'Behavioural disruption: rest' and 'Behavioural disruption: feeding'

e) 'Foot disorders' was reported as being associated with 'Mortality – euthanised', 'Mortality – unassisted', 'Leg disorders', 'Behavioural disruption: rest', 'Behavioural disruption: feeding', 'Exhaustion', 'Behavioural disruption: flooring'

f) ‘Behavioural disruption: flooring’ was reported as being associated with ‘Foot disorders’, ‘Leg disorders’, ‘Behavioural disruption: rest’ and ‘Behavioural disruption: feeding’

g) ‘Leg disorders’ was reported as being associated with ‘Mortality – euthanised’, ‘Foot disorders’, ‘Behavioural disruption: rest’ and ‘Behavioural disruption: flooring’

h) ‘Behavioural disruption: feeding’ was reported as being associated with ‘Foot disorders’, ‘Behavioural disruption: rest’, ‘Behavioural disruption: flooring’ and ‘Exhaustion’

8. Objective 2

8.1. Literature review

This section provides a review on the recent literature about the sensitivity (Se) and specificity (Sp) of the ABMs themselves/for detecting the worse adverse effects and about the intra-observer repeatability (IOR) and inter-observer agreement (IOA) of the selected ABMs.

The selection of which ABMs are relevant for each WAE is based on the recommendations from Objective 1. The review will go through the ABMs one by one. However, the ABMs ‘Unassisted deaths’ and ‘Euthanised cows’ will be presented in one section, as most literature on mortality do not distinguish between these.

8.1.1. **Unassisted deaths and euthanised cows**

Mortality can be measured either as mortality risk (a proportion of a population) or mortality rate (the speed of dying in a population at risk, Toft et al., 2004):

Mortality risk = # animals that die during the study period / # animals in the population at the start of the study

Mortality rate = # animals that die during the study period / time at risk of dying.

Thus the measures are analogue to the measures incidence risk and incidence rate when measuring diseases. Often the time span used is 1 year and the measures will then be “annual mortality risk” or “annual mortality rate”. Sometimes also “lactational mortality risk” using number of lactations as denominator has been used.

In addition, case fatality (or lethality) can be used for expressing mortality due to a certain disease:

Case fatality = # death due to disease in time period / # new cases in time period

Thus the case fatality is the probability that a diseased animal eventually dies from the disease. Note that “case fatality” is sometimes referred to as “case fatality rate”, although it is a proportion, and not a rate.

8.1.1.1. Validity

Despite the huge importance of mortality in relation to animal health economics as well as animal welfare, the literature on mortality is relatively limited. Thus, a review in 2006 identified only 19 relevant studies on the occurrence of mortality in countries with intensive dairy production (Thomsen and Houe, 2006).

When using the search string: (‘dairy cow*’ or ‘cow*’ or ‘cattle’) and (‘mortality’ or ‘dead’ or ‘death’ or ‘deaths’ or ‘survival’ or euthan*’) and (‘sensitivity’ or ’specificity’ or ’reliability’ or ‘repeatability’ or ‘reproducibility’ or ‘agreement’), approximately 500 publications were retrieved (Appendix F.). However, when studied for relevance, very few have information on using mortality as a measure for welfare measures. Still, some publications have information on the magnitude of mortality as well as important causes, which will be outlined first.

In most studies, the mortality ranged between 1 and 5% (Thomsen and Houe, 2006) per year or per lactation. But also important are the causes of death that may indicate whether the death has been painful. Important causes of death include accidents, calving disorders, digestive disorders, locomotor disorders, metabolic disorders and udder/teat disorders (Burnside et al., 1971; Esslemont and Kossaibati, 1997; Faye and Perochon, 1995; Gardner et al., 1990; Harris, 1989; Menzies et al., 1995; Milian-Suazo et al., 1988; Stevenson and Lean, 1998; Thomsen et al., 2004; Thomsen et al., 2012; White and Nichols 1965). Furthermore, a number of causes for sudden or unexpected deaths has been listed (Radostits et al., 2007), but these are considered very seldom and not relevant in this context.

As stated, the publications available are mostly dealing with the occurrence and not the validity of the recordings of death. However when it comes to correct classification of dead animals one can obviously anticipate that the classification is 100% correct, i.e. 100% sensitivity and 100% specificity.

An important circumstance in relation to mortality is whether the cows died unassisted or were euthanised. However, very few studies have distinguished between these categories (Alvåsen et al., 2012; Thomsen et al., 2004; Thomsen et al., 2012). Still, it seems fair to conclude that welfare is most likely compromised regardless of whether the cow died unassisted or were euthanised. Thus, a Danish study on the relevance of using register data on mortality concludes that whether a cow is euthanised or is dying unassisted, there has often been a history of a trauma or a disease, which has caused some level of suffering. Still, the amount of suffering depends on how quickly the cow is euthanised (Houe et al., 2012; Houe and Jensen, 2012;) and it must be anticipated that cows dying unassisted suffer the most.

Instead of mortality as such, it has also been suggested to look at the total life span of the cows including recordings of production and diseases (Houe et al., 2011). However, this methodology has the draw back that welfare problems can only be obtained retrospectively and not give a measure of the welfare here and now.

Concerning the association between mortality and other adverse effects (Table 18:) a Swedish study in 55 herds showed that mortality together with two fertility measures, still births, mastitis and feed related diseases had a sensitivity of 96% and a specificity of 56% of correctly classifying herds with welfare problems (Nyman et al., 2011). A Danish study in 40 dairy herds evaluating the performance of register data as predictors for dairy herds with high lameness prevalence identified mortality as a significant predictor for lameness (Otten et al., submitted). Thus at a predefined cut-off of an annual mean mortality rate per 100 cow years =5.7, a sensitivity of 40% (CI=17-69%) and a specificity of 67% (CI=49-81%) was calculated. At a data driven cut-off of 3.6, a sensitivity of 100% (CI=72-100%) and a specificity of 53% (CI=36-70%) was calculated.

Table 18: Associations between mortality and animal welfare problems

Reference	Location and year	Design	N herds	Predictors	Response	Measures
Nymann et al., 2011	Sweden 2005	Random (stratified on herd size)	55	Mortality, fertility, still birth, mastitis, and feed related diseases	The herd did not score among the 10 % worst (among 9 welfare indicators)	Se = 96% Sp= 56 %
Otten et al., submitted	Denmark 2004	Random among large herds (>100 cows)	40	Mortality rate per 100 cows>5.7	Lameness prevalence >16%	Se = 40 % Sp = 67%
Otten et al., submitted	Denmark 2004	Random among large herds (>100 cows)	40	Mortality rate per 100 cows > 3.6	Lameness prevalence >16%	Se = 100 % Sp = 53 %

8.1.1.2. Robustness

No literature was identified on this issue. However, similar to the conclusion on high validity on this ABM, there is no reason to believe that there should be any intra- or inter-observer variation for this ABM, because mors (death) is an irreversible event.

Mortality (unassisted and euthanised) is defined as a WAE. Mortality as an ABM is a direct measure of the WAE mortality and it is thus considered a useful measure for animal welfare. Furthermore, it is relevant also to study mortality as an ABM for other adverse effects.

8.1.2. **Evidence of mastitis (SCC)**

As SCC is in essence used as a surrogate measure for mastitis, this section will initially highlight the main findings on the recordings of mastitis in databases and in particular the relation between somatic cell count (SCC) and mastitis.

8.1.2.1. Validity

Mastitis is characterised by physical, chemical as well as microbial changes in the milk. Furthermore, there are pathological changes in the glandular tissue. Visual inspection of foremilk has been the traditional way of detecting mastitis. A Se of 80% and Sp of 100% has been given for detecting cows with mastitis during foremilking (Hillerton, 2000).

In a Swedish study among 177 farmers, the completeness of mastitis recordings to the database was 78% (Mörk et al., 2009). The criteria for mastitis included clinical signs such as flakes, clots or swellings. Completeness was defined as:

(“number of disease events in the farmers’ data also found in the database”) /

(“number of disease events in the farmers’ data reported as veterinary treated”)

Thus, the completeness can be said to reflect the sensitivity of the database to record veterinary treated mastitis cases. In a later study including 580 farmers in 4 Nordic countries, the completeness (proportion of clinical mastitis cases found in the national database) was 94% (Denmark), 56% (Finland), 82% (Norway) and 78% (Sweden) (Wolff et al., 2012). Finland has a relatively low completeness which may possibly be explained by the fact that many of the treatments are given after a bacteriological result, only hereafter the cow is treated and a recording made.

In conclusion, clinical mastitis is fairly easy to diagnose correctly. The quality of databases in the Nordic countries seems relatively good and most cases are recorded in the database. However, this might not comply in all countries but information about other (national) databases is not available.

The somatic cells in the milk consist of epithelial cells and white blood cells. The distribution of cells in a healthy gland has been given as neutrophils (<11%), macrophages (66-88%), lymphocytes (10-27%), and epithelial cells (0-7%) (Lee et al., 1980 cf. Radostits et al., 2007). The white blood cells will increase in numbers in response to intramammary infections. In a healthy cow, SCC values of 20-30,000/ml have been given (Østerås, 2003) or at least it is less than 100,000 in healthy animals (Sharma et al., 2011). From about 50,000/ml, there will gradually be changes in the milk. The cell count in a diseased cow can vary from less than 100,000 up to many millions per ml. Generally SCC above 200,000/ml is considered abnormal and would indicate that there is an inflammation in the udder (Harmon, 2001). The SCC is mostly influenced by the infection status (Dohoo and Meek, 1982; Schepers et al., 1997). For example, an Italian retrospective study found that the mean bulk milk somatic cell count among dairy farms in North Italy (925,069 milk samples) decreased from 321,000 cells/ml in 2005 to 284,000 in 2011 while, in the same period, the isolation of contagious pathogens (Streptococcus agalactiae and Staphylococcus aureus) decreased from 55.44% in 2005 to 26.96% in 2011 (736,797 composite or quarter milk samples) (Bertocchi et al., 2013). In addition, the SCC may show variation according to lactation stage, age, breed and other factors as reviewed by Sharma et al. (2011). However, it seems that normal secretion from uninfected quarters could not be significantly influenced (i.e. exceeding 200,000 per ml) by parity, stage of lactation, or heat stress (Harmon, 1994). Very high SCC is practically only influenced by an udder infection.

Several different cut-off values between normal and mastitic milk has been given, i.e. ranging from 100,000 to 500,000 cells/ml (Sharma et al., 2011). Using SCC as a screening test for intramammary infections at a threshold of 200,000 per ml, a Se = 72.5% and Sp = 85.5% was obtained (Dohoo and Leslie, 1991). At a threshold of 200,000 cells per ml Schepers et al. (1997) found similar Se and Sp for intra mammary infections namely 74.5% and 89.6%, respectively. They also made the calculations for thresholds of 100,000 and 400,000 cells per ml (see Table 19:). A study by Sargeant et al. (2001) showed that in very early lactation the specificity is lower (i.e. more false positives) when using SCC to predict infection. Thus, it seems obvious to conclude that associations between occurrence of mastitis/udder health problems and other welfare problems will also be reflected in associations between SCC and other welfare problems.

The use of SCC for predicting intramammary infections at the animal level can also be extended to use SCC of bulk milk to predict mastitis problems at herd level. At bulk milk SCC (BMSCC) levels of 200,000, 500,000, 1,000,000 and 1,500,000, the corresponding percentage of infected quarters has been estimated to: 6, 16, 32 and 48% (Radostits et al., 2007). At bulk milk levels from less than 100,000 up to 600,000 an increase in % infected cows from approximately 5% up to 80% were seen (Philpot and Nickerson, 1991). In addition, it was found by Barkema et al. (1998) that there are correlations between management practises and BMSCC: in herds with "low" BMSCC (<150,000) significantly more attention was paid to general hygiene ($p < 0.05$) than the higher BMSCC herds (divided in "mid" BMSCC (150-250,000) and "high" BMSCC (250-400,000) herds).

Table 19: Studies on sensitivity and specificity of SCC as test for mastitis

Author	SCC threshold	Mastitis definition	Animals	Se and Sp	Remarks
Dohoo and Leslie 1991	200,000	Major and minor pathogens. The same organism had to be recovered from the quarter on any two of the three weekly samples and at least one of those samples had to have had a cell count > 300 000 cells/ml	25 cows 1565 milk samples	Se =72.5 % Sp = 85.5%	
Schepers et al., 1997	100,000	>100 cfu/ml of the same infectious agent	544 cows; 7 herds; 22,467 quarter milk samples	Se=83.2% Sp=80.5%	
Schepers et al., 1997	200,000	Do		Se=74.5% Sp=89.6%	
Schepers et al., 1997	400,000	Do		Se=60.8% Sp=95.0%	
Sargeant et al., 2001.	100,000	'Presence of one or two bacterial species in one or both quarter milk samples taken within 12 h of calving and at d 3 postcalving' Only 'any infection' shown	131 cows; 3 herds; 520 quarter milk samples	Se=96.7% Sp=4.4%	DIM[(a)] = 1
Sargeant et al., 2001.	250,000			Se=83.5 % Sp=21.3%	DIM[(a)] = 1
Sargeant et al., 2001.	500,000			Se=69.8% Sp=68.8%	DIM[(a)] = 1

(a) DIM = Days in milk.

Concerning the association between SCC and any of the adverse effects (Table 20:), a Danish study of 6,839 herds on the risk of mortality during the first 100 days of lactation showed that an increase in mean SCC of 100,000 cells/ml had an OR of 1.16 (95% CI=1.14-1.19) for an increase in mortality (Thomsen et al., 2006). An Italian study compared bulk milk somatic cell count (BMSCC) with the outcomes of an Italian dairy cow welfare assessment protocol, carried out in 265 farms. The study showed a better animal welfare score with a lower BMSCC (Bertocchi et al., 2012). In particular, it was found that in farms, with the worst welfare score, BMSCC value was higher than 345,000 cells/ml (Bertocchi et al., 2012). A Danish study in 40 dairy herds evaluating the performance of register data as predictors for dairy herds with high lameness prevalence identified BMSCC as a significant predictor for lameness (Otten et al., submitted). Thus at a predefined cut-off of BMSCC=245,000 cells/ml, a sensitivity of 40% (CI=17-69%) and a specificity of 80% (CI=63-90%) was calculated. At a data driven cut-off at BMSCC=213,491 cells/ml, a sensitivity of 80% (CI=49-94%) and a specificity of 67% (CI=49-81%) was calculated. This could indicate an association between SCC and foot disorders that might be relevant to explore further.

Some studies have identified a negative association between SCC and lameness in some farms (Archer et al., 2011). Thus, in some herds, cows with lameness had lower SCC than cows without lameness. They suggested that an explanation could be that lame cows are standing more and therefore less exposed to udder pathogens.

Concerning the association between mastitis and any of the adverse effects, the previously mentioned study from Sweden also included mastitis which together with four other measures had a sensitivity of 96% and a specificity of 56 % of correctly classifying herds with welfare problems (Nyman et al., 2011). Sogstad et al. (2006) found an association between clinical mastitis and wounds and swellings at the tarsus.

Table 20: Associations between SCC and animal welfare problems (adverse effects)

Reference	Location	Design	N herds	Predictors	Response	Measures
Thomsen et al., 2006	Denmark 2000-2001	Cross-sectional, all Danish dairy herds (>10 cows)	6839	Increase in mean SCC of 100,000 cells/ml	Mean mortality risk during first 100 days of lactation	OR=1.16
Bertocchi et al., 2012	Italy	Correlation study among dairy herds in Northern Italy	265	BMSCC, 345,000 cells/ml	Animal welfare score	ρ = - 0.33
Otten et al., submitted	Denmark 2004	Random among large herds (>100 cows)	40	BMSCC >245,000	Lameness prevalence >16%	Se = 40 % Sp = 80%
Otten et al., submitted	Denmark 2004	Random among large herds (>100 cows)	40	BMSCC >213,000	Lameness prevalence >16%	Se = 80 % Sp = 67 %

8.1.2.2. Robustness

The analytical variation due to different processing in the laboratory (cooling, stirring etc.) only have minor influence on the result although there are larger uncertainty at very high cell counts (Rasmussen and Frimer, 1993). The analytical uncertainty is given on the technical specifications for the analytical apparatus and in general the robustness is high for this measure. For frozen samples there can be a reduction of 10-20% (ISO 13366-2:2006 (IDF 148-2: 2006). The agreement between different cell counters is very high. Thus, the correlation between the Fossomatic method and the standard 'Breed' method has been reported as 0.98, the correlation between Breed method versus Delaval method as 0.97 and the correlation between the Delaval method and Fossomatic method as 0.96 (Kawai et al., 2013). Also many on-farm tests for prediction of SCC show high agreement with laboratory instruments (Brandt et al., 2010).

From the literature review, SCC can be considered a robust and valid measure, as it has associations with mortality, overall welfare and lameness.

8.1.3. **Measures of lameness**

8.1.3.1. Validity

Lameness scoring has been used in different contexts for several years. Research in dairy cow gait has primarily been driven by an interest in detecting lame cows, defined as cows having an abnormal gait resulting from pain or discomfort from hoof or leg injuries and diseases. Lameness has been associated with substantial production losses (recently reviewed by Bicalho and Oikonomou (2013)) and in the context of animal welfare, lameness has been used as a proxy of cows suffering from pain and thus cows having a compromised welfare according to some of the most widely recognised animal welfare definitions (Welfare Quality, (2009), Farm Animal Welfare Advisory Committee, (1979)).

Most lameness scoring systems assess deviations in locomotion relative to 'normal gait'. However, these changes can be difficult to identify. Studies have shown that producers underestimate the prevalence of lameness in their herd (Whay et al., 2003, Leach et al., 2010) and data from farmers' recordings of lameness should therefore be used with some precaution.

Given the definition of lameness, the validity of a lameness scoring system may thus be considered as its ability to detect animals suffering from pain in the feet or legs. An indication of the implication of pain in lameness is the reduction in lameness score obtained by the administration of analgesia to lame cows as demonstrated by Rushen et al. (2007). Measuring pain in animals is not as straightforward as in humans and most indicators of pain in animals are based on the animals' behavioural or physiological responses toward noxious stimuli (Molony and Kent, 1997, Prunier et al., 2013). In a recent study, Tadich et al. (2013) found positive associations between different pain indicators and a 5-point lameness score. For instance, hyperalgesia increased with increased severity of lameness, thereby confirming the findings in previous studies (Ley et al., 1996, Whay et al., 1998). Dyer et al. (2007) measured pain using a claw compression test. They found that the pain increased with increased severity of lameness. However, they also found painful lesions in some non-lame animals indicating that some cows may suffer from (subclinical) pain that cannot be detected by a visual lameness scoring.

To our knowledge, no studies have specifically evaluated the sensitivity or specificity of visual lameness scoring as a test of pain in feet or legs of dairy cows.

The association of increased gait scores with the presence of different hoof lesions has been evaluated in many studies. Recently, in a study of 1340 dairy cows from 42 Danish herds, Thomsen et al. (2012) found increasing odds of hoof lesions with increasing gait scores regardless of the type of lesion. Generally, increased gait scores have most consistently been associated with the presence of sole ulcers, white line disease or interdigital purulent inflammation (phlegmon) whereas other hoof lesions may show more subtle symptoms that are not easily detected by a visual gait scoring (Flower and Weary, 2006, Frankena et al., 2009, Tadich et al., 2010, Sogstad et al., 2012, Thomsen et al., 2012). This confirms the findings of Dyer et al. (2007) as previously mentioned.

Studies specifically evaluating the validity in terms of Se and Sp of gait scoring for the detection of hoof lesions are scarce. Chapinal et al. (2009) specifically evaluated the Se and Sp of detecting cows with sole ulcers. A visual, 5-point numerical gait score was used, where the observers could allocate a half-integer score if a cow exceeded the requirements of a particular score. Se and Sp were calculated for three different clinical thresholds. With the threshold at score 3, the Se/Sp was 0.85/0.38; with threshold at score 3.5 they were 0.54/0.70; and at score 4 they were 0.38/0.95. Thus the best accuracy (= likelihood that a cow was correctly classified at the threshold) was obtained at threshold at score 4 (0.81).

Resting behaviour can change in dairy cows suffering from lameness. Yet, studies of the behaviour of lame cows are equivocal. Lameness has been associated with longer resting times (Chapinal et al., 2010, Ito et al., 2010, Calderon and Cook, 2011), shorter resting times (Gomez and Cook, 2010, Pavlenko et al., 2011), longer mean duration of lying bouts (Chapinal et al., 2009, Ito et al., 2010, Thomsen et al., 2012), and longer time spent standing in the stall (Cook et al., 2007, Dippel et al., 2011). Thus, the effect of lameness on resting behaviour is complex and depended on climate, housing type, type of resting area, time available for rest and lactation stage, among other things (Olmos et al., 2009, Gomez and Cook, 2010, Calderon and Cook, 2011).

We have not been able to localise any studies specifically evaluating the validity of lameness for the detection of resting behaviour disruptions.

8.1.3.2. Robustness

Table 21: gives an overview of studies of the IOA and in Table 22: there is an overview of studies of the IOR of some commonly used lameness scoring systems. This shows that the IOA as well as the IOR varies substantially between the studies. The robustness of different gait scoring systems can be affected by the observers themselves (e.g. errors of omission, observer expectations, and experience) as well as characteristics of the scoring system (e.g. level of detail in the description and the use of general terms) (reviewed by Flower and Weary (2009)).

Many lameness scoring systems are based on 3 to 5-point ordinal scales with higher number describing more severe lameness. Alternatively, a visual analogue scale (VAS) can be used. Here, the observers mark on a horizontal line the point that represent their perception of the magnitude/extent of the given lameness criteria (Tuyttens et al., 2009). Flower and Weary (2006) compared a 5-point ordinal scale with a 100-point VAS and demonstrated satisfying levels of IOR and IOA with both scales, whereas Tuyttens et al. (2009) found that the IOA was significantly better when using a VAS compared to a 3-point ordinal scale.

In summary, lameness seems to have moderate validity in predicting adverse effects and also moderate robustness. Lameness scorings may have relevance towards some foot disorders. Some problems with validity (e.g. specificity) may be solved by repeating measures. The moderate validity and robustness may be a minor problem if used on the herd level.

Table 21: Studies of inter-observer agreement (IOA) of lameness scoring.

Reference	Type of scale tested	Measure of IOA	Value	N	Notes
Tuyttens et al. (2009)	Ordinal	Estimated from a linear model: Ratio of the variance component of random video effect over the total variance.	0.35 (SE = 0.025)	40 cases. 53 observers	Estimated IOAs higher for VAS than ordinal scale. High correlation btw. mean ordinal and mean VAS score. No effect of order of videos – suggests no fatigue or learning within observers.
Tuyttens et al. (2009)	Continuous VAS	Estimated from a linear model: Ratio of the variance component of random video effect over the total variance.	0.44 (SE = 0.025)	40 cases. 53 observers	Estimated IOAs higher for VAS than ordinal scale. High correlation btw. mean ordinal and mean VAS score. No effect of order of videos – suggests no fatigue or learning within observers
Winckler and Willen (2001)	Ordinal	No formal statistical test.	68% full agreement 30% differed by 1 unit 2% differed by 2 units	147 cases. 3 observers	Most deviation (62%) occurred within lameness score 1 and 2
March et al. (2007)	Ordinal	PABAK[a]	0.32 – 0.94	21 – 68 cases. 2 observers	Testing was done at 9 occasions. The IOA improved with training
		Weighted Kappa	0.41 – 0.86		
		Spearman rank correlation coefficient	0.55 – 0.89		
		Proportion agreement	0.46 – 0.95		
Brenninkmeyer et al. (2007)	Ordinal	PABAK of gait scoring with different categorisation:			
		With all five categories	0.37 – 0.53	21 – 144 cows per session.	Categorisation done retrospectively. Mean IOAs for 4 observers (6 pairs) calculated for each of 4 sessions.
		With only four categories (score 1 and 2 merged)	0.39 – 0.60	4 observers (6 pairs)	
		With dichotomised scale (lame > score 2)	0.59 – 0.70		

(a)Prevalence and Bias Adjusted Kappa

Reference	Type of scale tested	Measure of IOA	Value	N	Notes
Flower and Weary (2006)	Ordinal	Coefficient of correlation (R^1)	0.69	46 cases. 2 observers	Video recordings once daily for 7 consecutive days (n = 309). Each sequences was watched 14 times in total – the 13th time for the overall gait score on the VAS scale and the 14th time for the overall gait score on the ordinal scale. For IOA, a second trained observer scored and rescored this same sample of recordings
Flower and Weary (2006)	Continuous VAS	Coefficient of correlation (R^1)	0.76 – 0.82	46 cases. 2 observers	
von Keyserlingk et al. (2012)	Ordinal	PABAK	0.84	228 cows.	
Borderas et al. (2008)	Ordinal	Pearson–product moment correlation	0.92	65 cows	Video recordings.
Bicalho et al. (2007)	Ordinal	Weighted Kappa	0.45 – 0.48	402 – 459 cows per pair of observers. 3 observers	
		ROC analysis (area under curve)	0.74 – 0.77		
Thomsen et al. (2008)	Ordinal	Kappa without training - mean, all observers	0.32	50 cows per session. 5 observers	
		Weighted kappa without training - mean, all observers	0.48		
		Kappa after training - mean, all observers	0.38		
		Weighted kappa after training - mean, all observers	0.52		
		Equidistance – Polychoric correlation coefficient (per observer)	0.76 – 0.96		Suggests that all observers had approximately the same thresholds to categorise lameness. However, the 5 lameness categories were not equidistant

Table 22: Studies of intra-observer reliability (IOR) of lameness scoring.

Reference	Type of scale tested	Measure of IOR	Value	N	Notes
De Rosa et al. (2003)	Ordinal	Repeatability - Kendall coefficient of concordance (W)	0.43 – 0.66	30-40 cows in four herds. 7 observers	Three occasions of observations, 15-20 day intervals. Low level of significance in 2 (out of 3) herds: One with only two recording sessions and one with very low prevalence of lameness.
Flower and Weary (2006)	Ordinal	Coefficient of correlation (R^2)	0.76 – 0-85	46 cases. 2 observers	Video recordings once daily for 7 consecutive days (n = 309). Each sequences was watched 14 times in total – the 13th time for the overall gait score on the VAS scale and the 14th time for the overall gait score on the ordinal scale. For the IOR, observer 1 rescored recordings from 1 d selected at random, at least 7 d after the first recording.
Flower and Weary (2006)	Continuous VAS	Coefficient of correlation (R^2)	0.73	46 cases. 2 observers	
Borderas et al. (2008)	Ordinal	Pearson–product moment correlation	0.92	63 cows.	Video recordings. For IOR same cow observed twice (n=65)
Thomsen et al. (2008)	Ordinal	Kappa without training - mean, all observers	0.46	50 cows per session.	
		Weighted kappa without training - mean, all observers	0.60	5 observers	
		Kappa after training - mean, all observers	0.43		
		Weighted kappa after training - mean, all observers	0.53		

8.1.4. **Number of foot lesions**

This section displays the main literature findings about the validity and robustness of the recording of foot lesions. Here, foot lesions are considered to be lesions or disorders in the feet or hooves. The term 'Foot lesion' will here primarily comprise of claw lesions typically diagnosed during hoof trimming (e.g. claw horn lesions or sole ulcers) or medically treated diseases (e.g. interdigital phlegmon).

8.1.4.1. Validity

Data on numbers of foot lesions can be obtained from different sources: farmers, veterinarians and hoof trimmers can keep records of diagnoses and treatments and these might eventually be collected in larger databases that in some countries can be national databases. Recently, the validity of using register data from central national databases for evaluating the occurrence of leg and foot disorders has been investigated in the Nordic countries. Here, Lind et al. (2012) compared data obtained by farmers recording foot and leg disorders during a two-month periods with data from the national database of disease records to estimate and compare completeness of the national databases in the four Nordic countries. The conclusion was that the ability to estimate the true hoof disease occurrence was generally low and varied between the countries. Furthermore, Lind et al. (2012) found that some disease cases occurred in the national databases but not in the farmers' recording thereby indicating that the farmer had simply failed to record the case. Other studies of the national Nordic databases have found that completeness of disease data varies with different diagnoses and that the incidence of disease in the databases are at best a conservative estimate of the true disease incidence (Mörk et al., 2009, Espetvedt et al., 2012, Rintakoski et al., 2012, Wolff et al., 2012). In the Nordic countries, it is mandatory to record the use of medication at least at farm level and most of these data are also collected in the central databases. In this way, the databases contain information about most veterinary treatments but do not necessarily depict the true morbidity of disease (Mörk et al., 2009). Many hoof lesions are observed and eventually non-medically treated by the farmer or a hoof trimmer and this is likely to partly explain the low completeness of locomotor disorders in the national databases (Lind et al., 2012).

To our knowledge, there are no recent studies of validity of national databases outside the Nordic countries and the problems described above indicates that studies of database validity are essential and even then, the use of central register data should be used with some caution as they are likely to represent an underestimation of the true disease incidence.

8.1.4.2. Robustness

The information in the literature about the robustness of hoof lesion scoring is scarce. Capion et al. (2008) calculated the IOA (by weighted Kappa values) between a trained researcher and four professional hoof trimmers. Five diagnoses were considered (the range of the weighted Kappa values displayed in brackets): heel horn erosion (0.03-0.48), sole haemorrhage (0.76-0.88), white line lesions (0.40-0.75), interdigital dermatitis (0.27-0.63) and digital dermatitis (0.42-0.78). This illustrates that the robustness of hoof lesion scoring very much depends on which diagnoses are included.

We have not been able to locate any studies considering the IOR of hoof lesion scoring.

As foot lesions have severe problems with completeness (i.e. being recorded), and also on some occasions have low robustness, they are not considered as a suitable ABM.

8.1.5. Number of leg lesions

In the current project, this ABM is defined as carpal and tarsal lesions and swellings; other minor integument alterations (e.g. hairless patches) have not been included.

8.1.5.1. Validity

Overall, the problem of validity of the database records for leg lesions is very similar to the problems mentioned for number of foot lesions (Section 0); in particular low completeness (recording).

Leg lesions are associated with a number of other welfare problems. For the current project, the most relevant associations are the ones to “loser cows” (Thomsen et al., 2007). Loser cows have been defined as a cow that has a number of clinical signs deviating from the normal (healthy) condition, among them severe hock lesions. Loser cows have higher risk of unassisted death or euthanasia, and a higher risk of needing medical treatment.

8.1.5.2. Robustness

We have found very few studies of inter observer reliability for either carpal or tarsal lesions or swellings. The only article that directly addresses the question of the reliability of carpal lesions is one by Gibbons et al. (2012) (indeed this article also states that no other such studies exist). The situation is however somewhat better for tarsus lesions and swellings.

Some of the studies that report on the intra and inter observer agreement on tarsal lesions report high values, around 80% agreement (Zurbrigg et al 2005, Rutherford et al 2008). These values are however only reached after training, and should not be taken as representative of the level usually attained by veterinarians, technicians or farmers.

Thomsen and Baadsgaard (2006) report the agreement among 5 veterinarians for hock lesions, before any explicit training. The prevalence adjusted bias adjusted kappa (PABAK) values ranged from 0.55 to 0.83 depending on the cut-off. For this study, the veterinarians were specifically asked to score a small number of different conditions and the values obtained should therefore probably be seen as the maximum agreement between observers that have not obtained any specific training.

Gibbons et al. (2012) looked specifically at the effect of training, both on carpal lesions as well as tarsal lesions. The PABAK for the first day of training was 0.73 for tarsal lesions, i.e. similar to the values reported by Thomsen and Baadsgaard (2006). For carpal lesions however the values were much lower (0.51). The effect of training was not dramatic; 8 days of training increased the agreement to 0.75 for the tarsal lesions, and to 0.58 for the carpal lesions.

As leg lesions have severe problems with completeness (i.e. being recorded), and also on some occasions have low robustness they are not considered as a suitable ABM.

8.2. Descriptions of the ABMs

A detailed description and the definition of cut-offs of these three selected ABM (‘Number of dead cows’, ‘Somatic cell count’ and ‘Measures of lameness’) deemed as relevant for this project based on the results from objective 1 and in the literature review, were initially approached by expert discussions within the consortium. These discussions were done via email, on ScienceNet and during a teleconference. The results of the discussion among consortium members are presented in Table 23: .

Table 23: Definitions of the ABMs as discussed by consortium members.

ABM	Description	Positive outcome	Suggested cut-offs
Number of dead cows	On-farm mortality is a herd-level ABM detected with **annual mortality rates (AMR)**. **Denominator:** The average number of cows in a herd during the last year is calculated by (1) averaging the total number of cows for each day of the last year (365 d) or (2) calculate the cow-years at risk (cow year =(sum of days a cow has been in the herd during a year)/365). **Nominator:** Cows that died on the dairy farm. Cows are lactating or non-lactating animals that have calved at least once. It includes 'uncontrolled' death of animals as well as cases of euthanasia. Emergency slaughter is or is not included in dead animals, and therefore needs to be specified in the calculation formula. Therefore, there are 2 outcomes: AMR(including emergency slaughter) = **AMR_ES** AMR(not including emergency slaughter) = **AMR_NES**	Lower no. of dead animals	1) 5% (used in IZSLER/CReNBA protocol, Italy) 2) 4.5% (as defined in WQ® protocol) 3) 8% (based on recent studies from EU and US)
Evidence of mastitis - Somatic Cell Counts	Evidence of mastitis is a herd-level ABM detected with **measures of somatic cell counts** at individual level (SCC) or in the bulk tank (BMSCC). With individual measures - Percentage of cows with High Somatic Cell Count (HSCC): (=BMSCC) - Percentage of tests with high somatic cell count:	Lower risk of mastitis	For SCC at cow level: 1) 400000 (most likely threshold to retrieve from data in the field) 2) 300000 (used as indicator of low SCC in many publications) For BMSCC 1) 400000 (most likely threshold to retrieve from data in the field) 2) 300000 (used as indicator of low SCC in many publications)

ABM	Description	Positive outcome	Suggested cut-offs
Measures of lameness	Lameness is a herd-level ABM measured with **locomotion scores**. At herd level, the outcome is percentage of lame cows during 365 days: **The nominator:** At individual level, a cow is categorized as lame if it for example: • Scores 3, 4 and 5 on the Sprecher scoring system, or • Scores 1 and 2 on WQ® scoring system, or • Score 3 on scoring system from Amory et al., 2006 (3-point), or • Score 3 and 4 on Nordlund et al., 2004 (4-point), or • Score 3 and 4 on Cook et al., 2003 (4-point), or • Score 3, 4, and 5 on scoring system from Flower & Weary, 2006 (5-point) This measure applies to cows kept in loose housed systems as well as tie-stall systems. Preferably, tied animals need to be loosened and walked for locomotion scoring. Alternatively, use WQ® protocol for tied animals.	Lower prevalence of lameness	1) 8% (used in used in IZSLER/CReNBA protocol, Italy) 2) 10 or 20% (based on recent studies from EU and US)

8.3. Identification of ABMs and WAEs in the available data

Three datasets were available at ScienceNet. One was from ILVO, Belgium; one was from IZSLER/CReNBA, Italy and the third was derived from INRA, France. The database protocol and details of the datasets can be found in Appendix J. Appendix L. In the project application two additional datasets were mentioned. However, in Denmark as well as Sweden troubles with data ownership prevented the consortium from getting access to these data.

The available data were consulted and the different WAE and ABM measures were identified by systematically going through the matrix shown in Table 24: . Not all WAE/ABMs could be characterised by the available data. The WAE 'Mortality' could be identified in all three datasets as an overall mortality not specifying whether cows died unassisted or were euthanised. For the WAE 'Foot disorders', the only available data were recordings of overgrown claws found in the Belgian data. This was deemed insufficient as a proxy of foot disorders in general. Regarding the WAE 'Behavioural disruptions' found in the Belgian data, it was not specified in what areas of the barn the observations were done and it was therefore not possible to distinguish behavioural disruptions between flooring and feeding areas. However, occurrence of agonistic behaviour (head butts, displacement, fighting, chasing) were used as a representation of an overall measure of behavioural disruptions in the flooring and feeding areas. To describe the WAE 'Behavioural disruption, resting' information about 'Time to lie down', 'Collisions with equipment' and 'Number of cows lying outside lying area' were used. The WAE 'Exhaustion' was assessed by body condition scores which were found in the Belgian data. Also the Italian data had information about BCS, but these observations were aggregated at herd level in a way where fat cows could not be distinguished from thin (= exhausted) cows.

The ABM 'Number of unassisted deaths' and 'Number of euthanised cows' were not distinguishable in the available data. Instead, 'Number of deaths, overall' was used. 'Evidence of mastitis, SCC' was found at cow level in the French and the Belgian datasets and aggregated at herd level (= bulk milk SCC, BMSCC) in the Italian dataset (IZSLER/CReNBA protocol). Likewise, 'Measures of lameness' was aggregated at herd level in the Italian data, whereas cow level samples were available in the French and Belgian data. In the French and the Belgian datasets the 3-point lameness scoring system from the WQ® protocol was used whereas the Italian data used a 4-point scoring system (see data description in the Appendix J. - Appendix K.) where scores were dichotomised in lame (score 1+2) and not lame (score 3+4). As mentioned above, the only measure of 'Number of foot disorders' was registration of overgrown claws in the Belgian data. 'Leg lesions' were defined as occurrence of lesions or swellings on the tarsus.

Table 24: Matrix describing the potential combinations of the selected WAE and ABMs

	Worst adverse effects (WAE)							
	Mortality			Leg lesions	Foot disorders(a)	Behavioural disruptions		Exhaustion (d)
ABM selected in objective 1	Unassisted	Euthanised	Total			Overall (b)	Resting (c)	
# of deaths, unassisted # of deaths, euthanised	NA	NA	NA	NA	NA	NA	NA	NA
# of deaths, overall			DM	B	NA	B	B	B
Evidence of mastitis (SCC)	NA	NA	I/B/F	B	NA	B	B	B
Measures of lameness			I/B/F	B	NA	B	B	B
# foot disorders			NA	NA	NA	NA	NA	NA
# leg lesions			B	DM	NA	B	B	B

DM = Direct measure
NA = Not available
I = Data from IZSLER, Italy
B = Data from ILVO, Belgium
F = Data from INRA, France

(a) = Overgrown claws is the only foot disorders available (ILVO data)
(b) = Overall behavioural disruption assessed by number of agonistic behaviours (head butts, displacement, fighting, chasing) - not possible to separate feeding area and flooring area in the available data
(c) = Behavioural disruption in resting assessed by time to lie down OR number of collisions with equipment OR Number of cows lying outside lying area
(d) = Exhaustion assessed by BCS

8.4. Descriptions of cut-offs for ABMs and WAEs

It has afterwards been necessary to choose one cut-off for each parameter, or for some parameters one cut-off on animal level and one on herd level (Table 25: Table 26:). Multiple cut-offs multiply the number of estimates. Including all suggestions presented by the consortium during the process could potentially have resulted in approximately 4,000 pairs of sensitivities and specificities solely based on the available data.

As argued above, the cut-offs will depend on the purpose of the calculation of the Se and Sp. Since this was not defined in the project call or application, we have in general chosen the cut-offs based on a discussion with the project members, available prevalence data and practical considerations. In the following, some specific arguments for deciding the specific cut-offs are summarised. For the WAE 'Mortality' and the ABM 'Number of dead cows', the cut-off found in the data from IZSLER/CReNBA was chosen. This level was also a compromise between the levels suggested in the WQ® protocol and the mortality levels found in recent studies (Alvåsen et al. 2012; review by Thomsen and Houe, 2006). Regarding the behavioural disruptions, we used thresholds given in the WQ® protocol. If more than 10% of the animals had a 'Time to lie down' above the 6.3 seconds (= threshold used in WQ® protocol) it was considered as a behavioural disruption. Using BCS as an indication of exhaustion the presence of cows with a BCS less than 2 were deemed as undesirable. For the ABM 'Evidence of mastitis' two cut-offs were defined: 1) A mean bulk milk somatic cell count (BMSCC) above 400000 and 2) Proportion of cows with SCC above 400000 larger than 10%. The cut-off for BMSCC at 400000 cells/ml was chosen out of feasibility in that multiple consortium members stated that this is a cut-off used in field data. However, in the Italian data (IZSLER/CReNBA) set the distribution of the data at a cut-off at 400000 cells/ml was very skewed (Table 27:) and it was therefore decided also to include a threshold at 300000 cells/ml (a threshold also found in the IZSLER/CReNBA dataset). For cow level SCC, the threshold at 400000 was chosen for the same reasons as stated above. For the ABM 'Measures of lameness', cut-off found in the data from IZSLER/CReNBA was chosen. The highest cut-off from the IZSLER/CReNBA data was chosen in order to reach some compromise between the level used in the dataset and the levels suggested by consortium members based on recent studies of dairy cow lameness.

Table 25: Definitions of cut-offs for the WAE

WAE	Definition	Cut-off
Mortality, overall	Annual mortality rate	5%
Leg lesions	Proportion of cow with at least one lesion or swelling on the tarsus	8%
Behavioural disruptions		
Overall	Mean number of agonistic behaviours per animal per hour [(a)]	5
Resting	Proportion of cows where 'Time to lie down' > 6.3 seconds > 10%[(b)]	1[(c)]
	Proportion of cows lying outside lying area >5%[(b)]	
	Proportion of cows with collisions with equipment > 30%[(b)]	
Exhaustion	Proportion of cows with BCS<2	0%

(a): Definition from WQ® protocol. Cut-off based on WQ® stating that a maximum expected frequency of agonistic behaviour is 500 per hour in a group of 100 cows
(b): WQ® limit for serious problem.
(c): At least one of the defined behavioural disruptions present

Table 26: Definitions of cut-offs for the ABMs

ABM	Definition	Cut-off
# deaths, overall	Annual mortality rate	5%
Evidence of mastitis (SCC)		
BMSCC	Geometric mean during last three months	400000
BMSCC	Geometric mean during last three months	300000
Cow level SCC	Proportion of cows with SCC > 400,000	10%
Measures of lameness	Proportion of moderately/severely lame cows	8%
# leg lesions	Proportion of cow with at least one lesion or swelling on the tarsus	8%

8.5. Descriptive analysis of available data

Having described and defined the ABMs and the WAE and their respective cut-offs, it was investigated how the data were distributed between the two levels of each variable (Table 27: and Table 28:).

Table 27: The distribution of data below and above defined threshold for each of the ABMs

ABM	Measure	cut-off	Data source	N herds	Distribution of herds (%)	
					< cut-off	≥ cut-off
Number of deaths	Annual mortality rate	5%	F	131	81	19
			B	210	65	35
			I	442	92	8
Measures of lameness	Proportion of lame cows in herd	8%	F	131	38	62
			B	145	6	94
			I	442	89	11
Evidence of mastitis (SCC)	Proportion of cows with SCC > 400000	10%	F	128	66	34
			B	63	11	89
	Geometric mean of BMSCC last three months	400000	I	442	98	2
		300000	I	442	80	20
Number of leg lesions	Proportion of cow with at least one lesion or swelling on the tarsus	8%	B	145	30	70

F = Data from INRA, France, B = Data from ILVO, Belgium, I = Data from IZSLER/CReNBA, Italy

Table 28: The distribution of data below and above defined threshold for each of the WAE

WAE	Measure	cut-off	Data source	N herds	Distribution (%)	
					< cut-off	≥ cut-off
Mortality, overall	Annual mortality rate	5%	F	131	81	19
			B	210	65	35
			I	442	92	8
Leg lesions	Proportion of cow with at least one lesion or swelling on the tarsus	8%	B	145	30	70
Behavioural disruption, overall	Mean number of agonistic behaviours per animal per hour1	5%	B	92	41	59
Behavioural disruption, resting	Proportion of cows where 'Time to lie down' > 6.3 secs > 10%, OR Proportion of cows lying outside lying area >5%, OR Proportion of cows with collisions with equipment > 30%	1	B	92	92	8
Exhaustion	Proportion of cows with BCS < 2	5%	B	145	71	29

F = Data from INRA, France, B = Data from ILVO, Belgium, I = Data from IZSLER/CReNBA, Italy

8.6. Relative sensitivity and relative specificity estimates

Estimates of the relative sensitivity and the relative specificity of the selected ABMs as predictors of the WAE are presented in Table 29: . As indicated by the P-value, 'Measures of lameness' and 'Evidence of mastitis' were the only ABMs which were informative about the WAE and therefore, the only ABMs where interpretation of the estimates of Se and Sp is meaningful.

In the Italian data (ISZLER/CReNBA protocol, 'Measures of lameness' showed some potential in predicting 'Mortality, overall'. This test had a reasonably high Sp (0.90), whereas the Se was low (0.26). In the French and the Belgian data, 'Measures of lameness' did not correlate with 'Mortality, overall'. 'Measures of lameness' also correlated with the WAE 'Leg lesions'. Here, a high Se was found (0.97) whereas the Sp was low (0.11). This indicated that herds with a high proportion of lame cows were also likely to have a high proportion of cows with leg lesions. However, precaution should be taken in the interpretation as 94 out of 100 herds with lameness observations in the Belgian dataset had a proportion of lame cows > 8%. In the Belgian data, 'Evidence of mastitis' measured as the proportion of cows in a herd with SCC > 400000 correlated to 'Mortality, overall' and had a very high Se (1.00) but a low Sp (0.13). Also, this ABM correlated to 'Behavioural disruption, overall' and likewise, a high Se (1.00) and a low Sp was estimated. Again, data distribution has to be considered in the interpretation of these test results. Based on information from the available data, the ABMs 'Number of deaths' and 'Number of leg lesions' were not suitable as predictors of any of the WAE.

Table 29: Estimates of relative sensitivity (Se) and relative specificity (Sp) of selected ABMs as predictors of the WAE

ABM	WAE	Data source	Se	Se 95% CI	Sp	Sp 95% CI	P-value (χ^2)
Number of deaths	Leg lesions	B	0.33	0.23 - 0.43	0.60	0.44 - 0.76	0.43
	Behavioural disruption, overall	B	0.40	0.29 - 0.51	0.72	0.60 - 0.84	0.19
	Behavioural disruptions, resting	B	0.50	0.15 - 0.85	0.66	0.57 - 0.75	0.35
	Exhaustion	B	0.40	0.24 - 0.56	0.67	0.57 - 0.77	0.43
Measures of lameness	Mortality, overall	B	0.98	0.94 - 1.00	0.02	0 - 0.05	0.96
		F	0.72	0.54 - 0.9	0.41	0.32 - 0.50	0.26
		I	0.26	0.11 - 0.41	0.90	0.87 - 0.93	**0.01**
	Leg lesions	B	0.97	0.94 - 1.00	0.11	0.02 - 0.20	**0.04**
	Behavioural disruption, overall	B	0.95	0.90 - 1.00	0.07	0.01 - 0.13	0.72
	Behavioural disruptions, resting	B	0.88	0.65 - 1.00	0.05	0.01 - 0.09	0.37
	Exhaustion	B	0.98	0.94 - 1.00	0.07	0.02 - 0.12	0.29
Evidence of mastitis (SCC)	Mortality, overall	F	0.44	0.25 - 0.63	0.68	0.59 - 0.77	0.24
		B	1.00	1.00 1.00	0.13	0.05 - 0.21	**0.03**
	Leg lesions	B	0.89	0.82 - 0.96	0	0	0.13
	Behavioural disruption, overall	B	0.84	0.74 - 0.94	0	0	**0.01**
	Behavioural disruptions, resting	B	1.00	1.00 - 1.00	0.09	0.03 - 0.15	0.40
	Exhaustion	B	0.95	0.85 - 1.00	0.10	0.03 - 0.17	0.29
Evidence of mastitis (BMSCC > 400000)	Mortality, overall	I	0	0	0.98	0.97 - 0.99	0.41
Evidence of mastitis (BMSCC > 300000)	Mortality, overall	I	0.23	0.09 - 0.37	0.80	0.76 - 0.84	0.65
Number of leg lesions	Mortality, overall	B	0.68	0.54 - 0.82	0.25	0.16 - 0.34	0.43
	Behavioural disruption, overall	B	0.67	0.57 - 0.77	0.27	0.16 - 0.38	0.47
	Behavioural disruptions, resting	B	0.75	0.45 - 1.00	0.31	0.23 - 0.39	0.74
	Exhaustion	B	0.60	0.45 - 0.75	0.26	0.18 - 0.34	0.09

F = Data from INRA, France, B = Data from ILVO, Belgium, I = Data from IZSLER/CReNBA, Italy

9. Objective 3

9.1. Literature search for relevant factors of variation

After using the search strings listed in the methods and investigating references from EFSA reports, papers of interest as well as review papers, a total of 471 potentially interesting articles were obtained. There were 113 articles dealing with mortality, 238 articles dealing with mastitis/high SCC and 120 articles dealing with lameness. Each abstract was checked and when the article actually dealt with the subject of interest, i.e. a relationship between one or several of the above categories with one or several specific studied parameter(s), the article was scrutinised more thoroughly, and included in this work when the relationship was verified. The parameters were then listed in the manner that they were used in each article (allowing sometimes synonyms), with corresponding citing articles (see Appendix P.). A total of 199 parameters, classified under five entity classes: animal, housing, management, sampling and general parameters, were listed.

Most of these parameters were quite specific and could be regrouped under a higher-level parameter. Some were redundant and were thus merged. A few parameters were anecdotal, e.g. 'inhabitant density', or irrelevant, e.g. 'tail-docking', and were eliminated. This finally resulted in a list of 63 parameters, classified under five entity classes and 16 parameter types (Table 30:).

Table 30: Refined list of parameters obtained from literature search.

Entity class	Parameter type		Factor name
Animal class	Genetic	1	Breed
		2	Milk production/yield (individual, herd)
		3	Udder conformation (level of the udder from the hock…)
		4	Weight
	Physiology	5	Age
		6	Body condition score (skinny/normal/fat…)
		7	Days in milk/after calving
		8	Lactation and reproductive status (open/dry/pregnant)
		9	Parity (primiparous/multiparous, number of parities)
	Health	10	Reproductive problems (high levels of abortion, retained placentas, dystocia, metritis...)
		11	Injuries, accidents, pain (wounds, swellings…)
		12	Udder status (teat end quality, cleanliness, hyperkeratosis, contamination by pathogens)
		13	Hoof status (overgrown, trimming)
Housing class	Housing system	14	Housing system (tie-stalls, free-stalls with deep litter, free-stalls with cubicle)
		15	Milking system (Automatic MS, milking parlour, traditional milking)
	Space organisation	16	Density (overstocking: cubicles per animal; feeding space per animal, drinkers per animal…)
		17	Outside/exercise area (availability, size…)
		18	Access to pasture (duration, period, distance, presence of shed…)
		19	Calving pen (multiple animals/single animal…)
		20	Beef cow unit
		21	Contact with other species (poultry, cat)
	Material	22	Ergonomics (cubicle: brisket board, neck rail; tie*stall: electric cow trainers; steps in front of the manger…)
		23	Bedding material (cubicle, calving pen…)
	Floor	24	Floor type (slatted, concrete, mattress)
		25	Floor quality (broken ground, slippery ground, stairs, slope…)

Entity class	Parameter type		Factor name
	Ambience	26	Indoor environment (ventilation, Temperature Humidity Index)
Management class	Stockmanship	27	Behaviour towards the animals (at milking, herding…)
		28	Empathy, threshold to recognise disease and pain, call veterinarian
		29	Ability to detect/report diseases and welfare issues (detection of risk injury, of lameness…)
		30	Organic dairy production
		31	Available workforce at milking
		32	Relation with the outside (member of milk control program, contract with foot trimmer, classes attendance)
		33	Rules for mixing animals (introduction of transition cows in the herd, heifers…)
		34	Calving interval
	Hygiene/ Maintenance	35	Cleanliness of alleys and lying area
		36	Cleanliness of water trough and manger
		37	Manure used for bedding
	Herd health	38	Animal care/hygiene (cleanliness, hoof care, footbath, hair clipping)
		39	Biosecurity (rules for outside persons, quarantine duration…)
		40	Veterinary treatments (vaccines, antibiotics, foot care…)
	Milking	41	Order of milking (high producing cows, high SCC cows)
		42	Procedure (frequency, waiting time, individual care)
		43	Milking machine cleanliness (controls, dipping or backflushing units…)
		44	Hygiene procedures during milking (teat dipping, gloves, teat preparation…)
	Dry cows	45	Management of dry cows (dry cow therapy, specific surveillance, pre-calving hoof care)
		46	Housing of dry cows (comfort, cleanliness…)
	Feeding	47	Water quality and availability
		48	Concentrate consumption (proportion, amount)
		49	Forage or fibre availability / % in the ration
		50	Supplementation with minerals and vitamins (selenium, vitamin E/A…)
		51	Cows blocked at feeding (duration, prolonged standing…)
		52	Feeding frequency
		53	TMR feeding (Total Mixed Ration)
		54	Negative Energy Balance (confirmed by tests e.g. MUN Milk Urea Nitrogen test)
		55	Nutrition of calves (quality of milk: mastitic milk, antibiotics…)
		56	Specific nutrition according to lactation or physiological status (feeding transitions)
	Culling	57	Culling rules (lesions, udder conformation, DIM…)
		58	Replacement rate and origin (heifers bought outside…)
		59	Economics (price of milk, replacement heifers, culled cows…)
General factors class	General factors	60	Herd size
		61	Outdoor conditions (Temperature Humidity Index…)
		62	Geographical region
		63	Milk quota

9.2. Selection of the final parameters based on consortium partners' opinion

Consortium members were provided with a spreadsheet (Appendix N.). Eleven members answered (Appendix Q.). The answers are shown in Appendix Q. (relevance scores), Appendix R. feasibility to collect) and Appendix S. (characterisation of population).

The relevance of parameters ranged from (mean ± sd) 1.3 ± 0.5 (‘contact with other species’) to 4.4 ± 0.7 (‘density’) on the 1 to 5 scale. The median score ranged from 1 to 4. Twenty-four parameters were attributed a mean relevance score ≥ 3.5 (Appendix Q.). For these factors, the median score ranged from 3 to 4.

As regard to feasibility of collection, 54 parameters were selected by at least 1 member, and 14 parameters selected by at least 8 members (Appendix S.).

Seven parameters complied with the thresholds of inclusion for the criteria “relevance” and “feasibility”. These parameters are, in order of relevance, ‘parity’, ‘housing system’, ‘age’, ‘floor type’, ‘days in milk/after calving’, ‘access to pasture’ and ‘milk production/yield’. The median score was 4 for all of these parameters, which confirms that they are very important for the majority of the consortium experts. It was decided by the members of the consortium that concerning ‘parity’ and ‘age’, only ‘parity’ should be kept. In fact, ‘parity’ and ‘age’ are highly correlated and keeping them both could be redundant. Moreover, since reproductive cycles (i.e. time between parturitions) can be variable between individuals, animals of the same age can have experienced varying physiological pressure due to reproduction and lactation. Therefore, it was decided that ‘parity’ should be kept as it reflects better similar levels of physiological strain on the individual. Thus, six parameters were included regarding the criteria “relevance” and “feasibility”: ‘parity’, ‘housing system’, ‘floor type’, ‘days in milk / after calving’, ‘access to pasture’ and ‘milk production / yield’. Regarding the relevance of these parameters, only one (‘access to pasture’) was scored 2 or less by two experts and three (‘housing system’, ‘days in milk’, ‘milk production’) were scored 2 or less by only one expert. The last two parameters (‘parity’ and ‘floor type’) were always scored 3 or higher by all experts.

In the selection of parameters characterising the the population, 43 factors of variation were selected by at least 1 member and 7 parameters were selected by at least 8 members (Appendix S.). These factors were, from the most often chosen to the least often chosen, ‘geographical region’, ‘breed’, housing system’, ‘herd size’, access to pasture’, milk production/yield’ and ‘organic dairy production’. Of these seven parameters, three were already included in those selected for relevance/feasibility. The parameters to add for their ability to characterise a given population were therefore: ‘geographical region’, ‘breed’, ‘herd size’ and ‘organic dairy production’.

In summary, ten epidemiological parameters were identified by partners from the consortium as being the most important to collect: ‘parity’, ‘housing system’, ‘floor type’, ‘days in milk / after calving’, ‘access to pasture’, ‘milk production / yield’, ‘geographical region’, ‘breed’, ‘herd size’ and ‘’organic dairy production’ (Table 31:). Following is a description of each of these parameters.

9.2.1. Housing system

Dairy cows can be housed (confined) from a period of only the winter months to an all year-round period. A housing system is required to provide animals and humans with protection from adverse climatic conditions, when keeping animals outside for grazing is not the most efficient or cost-effective use of the land, or simply to allow an easier control and management of the herd (CIGR, 1994; Haskell et al., 2006).

Housing systems can be of several types: free stall (or “loose housing”), tie stall, and mixed systems including stall and outdoors access (International Committee for Animal Recording (ICAR), 2012).

A loose housing system is a housing system in which the cows can move around freely. This housing is typically divided according to function into a resting, feeding, and milking area. There are two main loose housing systems: cubicle and deep litter. In cubicle system, the resting area is divided into cubicles that function as resting places for each individual cow. Cubicles must be a delineated, comfortable and sheltered resting places and in sufficient number for all cows. In deep litter, the resting area is undivided and the bedding consists of a deep litter mat.

In a tied housing system, the cow's movement is restricted: cows are tied up by tether or the like. The boxes serve both as resting and feeding area. Milking is carried out in the stalls or in a milking parlour (Anonymous, 2001).

9.2.2. Floor type

Types of floors can be concrete, cement with wood shavings, slats, sand, rubber, straw, pasture (International Committee for Animal Recording (ICAR), 2012). Good slip resistance, ease of cleaning, encouraging locomotion, and promotion of claw health are characteristics of an appropriate flooring system, which could be described by a general term, floor ergonomics (Telezhenko and Bergsten, 2005).

9.2.3. Organic dairy production

Organic farming is a value-based method of agricultural production. The values are held in four principles that were recently articulated by the International Federation of Organic Movements (IFOAM). The four principles transferred in a statement are: principles of health, ecology, fairness and care.

The principles are the basis for development of prescriptions and standards that hold for any farming system operation or product that can legally be labelled "organic". EU adopted specific regulations regarding organic production of agricultural products and foodstuffs to include in livestock production (Council regulation no 1804/1999).

9.2.4. Days in milk (DIM)

Days in milk are the number of days since the last calving. Dairy cows are generally milked for 10 months and days in milk are then between 1 and in average 305 days, a 305-d lactation yield being the basis of genetic evaluation.

9.2.5. Milk production/yield

Milk production is the volume of milk produced, usually quoted for a year or a lactation, sometimes quoted as kg of butterfat or of milk solids produced. It is used as the benchmark of productivity of dairy cows.

Milk production data are records of volume and components of milk produced by individual cows or the whole herd (total quantities recorded or estimations from periodic samplings).The 305-day lactation yields of dairy cows are usually calculated based on monthly test-day yields of milk-recorded cows. The method is to interpolate the mean yield over the interval between monthly tests and to accumulate the resulting value after each test. Intervals between two tests should be close to 30 days (Schaeffer et al., 1977). Two milkings per recording days is the reference method and the interpolation method is the reference method for calculating lactations. Other methods are available.

9.2.6. Parity

The number of times a female has given birth, counting multiple births (twins or more calves born at the same time) as one and usually including stillbirths (Definition Source: NAL Thesaurus Staff). Cows are often divided in primiparous cows (experiencing their first parity) and multiparous cows.

9.2.7. Access to pasture

Access to pasture is measured as the duration that cows are outdoors and can graze (EFSA, 2009). In Welfare Quality®, it is counted as the number of days with access to pasture per day and the number of hours per day on pasture. Unrestricted pasturing during the summer period is compulsory in organic farming (Council regulation 1999).

Pastures are a source of forages and nutrients for dairy cattle, including heifers, dry cows, and the milking herd.

Pasturing access might be recorded by "availability of pasture plots; length of pasture paths; quality of pasture paths; provision of shade and shelter".

9.2.8. Breed

The term most commonly used to describe livestock populations or varieties is 'breed'. A breed is defined as: "a group of animals that has been selected by man to possess a uniform appearance that is inheritable and distinguishes it from other groups of animals within the same species. It is a product of artificial choice of characters that are not necessarily strategies for survival but are favoured by man for economic, aesthetic, or ritual reasons, or because they increase the social status of the owner of the animals." (Clutton-Brock, 1981). The list of breed codes is maintained by the Interbull Centre.

9.2.9. Geographical region

Geographical region is a zone - any of the regions of the surface of the Earth loosely divided according to latitude or longitude. A geographical region could also be e.g. a country.

9.2.10. Herd size

The herd size is the number of dairy cows (lactating and dry) in the herd. Herd size spans from a few individuals to thousands of heads, which relates to differences in e.g. grouping, feeding, reproductive management, calving and milking.

Table 31: Results from the partners' consultation on the 63 main factors of variation identified in the literature search.

	Mean relevance	Median relevance	Feasibility	Characterisation
Density	4.4	4	6	1
Parity	4.2	4	11	4
Housing system	4.2	4	10	10
Empathy, threshold to recognise disease and pain, call veterinarian	4.2	4	0	0
Ability to detect/report diseases and welfare issues	4.2	4	0	0

	Mean relevance	Median relevance	Feasibility	Characterisation
Hoof status	4.0	4	2	1
Veterinary treatments	3.9	4	6	0
Age	3.8	4	9	2
Animal care/hygiene	3.8	4	1	1
Injuries, accidents, pain	3.8	4	2	1
Floor type	3.7	4	10	2
Body condition score	3.6	4	2	2
Days in milk/after calving	3.6	4	10	1
Reproductive problems	3.6	4	3	2
Floor quality	3.6	4	4	1
Management of dry cows	3.6	4	3	1
Culling rules	3.6	4	1	1
Outdoor conditions	3.6	4	4	4
Access to pasture	3.5	4	10	9
Water quality and availability	3.5	3	1	0
Milk production/yield	3.5	4	11	8
Udder status	3.5	4	0	0
Bedding material	3.5	3	7	2
Negative energy balance	3.5	4	0	0
Hygiene procedures during milking	3.4	3,5	3	1
Cleanliness of alleys and lying area	3.3	3	0	0
Manure used for bedding	3.3	3	0	0
Ergonomics	3.3	3	6	1
Relation with the outside	3.3	3	5	1
Concentrate consumption	3.3	3	1	1
Forage or fibre availability / % in the ration	3.3	3	1	2
Economics	3.2	3	2	1
Calving pen	3.2	3	6	1
Housing of dry cows	3.2	4	4	1
Herd size	3.2	4	11	10
Specific nutrition according to lactation or physiological status	3.1	3	2	0

	Mean relevance	Median relevance	Feasibility	Characterisation
Outside / exercise area	3.1	3	8	2
Indoor environment	3.1	3	2	0
Milking machine cleanliness	3.1	3	4	1
Breed	3.0	3	11	10
Behaviour towards the animals	3.0	3	0	0
Milking procedure	3.0	3	1	0
Lactation and reproductive status	2.9	3	6	3
Milking system	2.9	3	10	7
Biosecurity	2.8	3	1	0
Order of milking	2.8	3	1	1
Replacement rate and origin	2.8	3	6	1
Supplementation with minerals and vitamins	2.7	3	2	0
Geographical region	2.7	3	11	11
Cleanliness of water trough and manger	2.7	2,5	1	1
Nutrition of calves	2.6	3	0	0
Organic dairy production	2.5	3	9	8
Rules for mixing animals	2.5	2	2	0
Calving interval	2.5	3	6	1
Cows blocked at feeding	2.5	3	2	0
Feeding frequency	2.5	3	5	1
Available workforce at milking	2.4	2	5	1
TMR feeding	2.4	2	5	1
Milk quota	2.4	2	8	2
Weight	2.3	2	1	0
Udder conformation	2.1	2	0	0
Beef cow unit	1.4	1	7	1
Contact with other species	1.3	1	1	0

9.3. Availability of factors of variations in the Routinely Collected Databases

Four of the five countries involved in the project responded to the request. Answers are summarised in Appendix T. Five of the ten selected parameters (‘parity’, ‘days in milk’, ‘herd size’, ‘breed’, ‘geographical region’) and two of the ABMs (‘somatic cell count’, ‘mortality’) are already collected in databases of the four countries who responded. Three of the parameters (‘housing system’, ‘floor type’, ‘access to pasture’) and one ABM (lameness') are never collected routinely. The other parameters ('milk production' and 'organic dairy production') are collected in some countries and not in others.

The five parameters and the two ABMs already collected in the databases are all collected at herd level. Data are available for the majority of the farms and collected continuously. Three of the parameters ('parity', 'milk production', 'breed') and the ABMs are also collected at individual cow level. Data are available for the majority of the animals.

10. Objective 4

In this section, we give an overview of the data uploaded into the DCF, what information is present in all databases, in what form, and what information is lacking.

Table 32: Overview of number of rows in the original and in the inserted database

Dataset	Italy	Denmark ISCC	Denmark HSCC	Denmark AMR	Belgium
# rows inserted	384,840	3,920,136	34,533	7,306	74,071
# rows original	231,536	3,920,136	3,920,136	3,670	24,946

The difference in the number of original rows and the number of rows inserted is due to the deletion of meat breeds (Italy), removing empty rows and placing the 'result values' (values of SCC, DIM, Parity, herd size, access to pasture, density, HSCC, BMSCC and Lameness) underneath each other in the column 'resVal' instead of besides each other as separate columns.

Table 33: Overview of elements found in databases

Element	Italy	Denmark	France	Belgium
Region	✓	✓	×	×
AnimalID	×	✓	✓	×
HerdID	✓	✓	✓	✓
HoldingID	✓	×	×	×
Year	✓	✓	✓	✓
Month	✓	✓	✓	✓
Day	✓	✓	✓	✓
Breed	✓	✓	✓	×
Production type	✓	×	×	×
Housing	×	×	×	×
Flooring	×	×	×	×
Bedding	×	×	×	×
Herd size	✓	✓	×	×
Access to pasture	×	×	×	×
Milk yield	×	×	×	×
Parity	✓	✓	×	×
DIM	×	✓	×	×
Density	×	×	×	×
HSSC	×	✓	×	✓
BMSCC	✓	×	×	×
AMR	✓	✓	×	×
HLAME	×	×	×	×

× = not present in database
✓ = present in database

Information that is lacking in all databases includes: housing, flooring, bedding, access to pasture, density and lameness score. Since the animalID's, herdID's and holdingID's were random numbers made by the data providers to ensure anonymity of the farms, no missing values were found here.

For the Italian database (IZSAM + IZSLER databases), information about the production system was not available for many farms. The percentage of missing values was 79% (this is excluding the SCC database which contained no information about production system). Breed was not specified in the SCC database. For the elements that were present in the Danish and Belgian database, there were 0% missing values.

Differences between reported information availability of data elements as stated earlier by the data providers as part of Objective 3, and the actual data we received are listed in Table 34: Contrary to expected, the Danish database that we received did not contain information about parity and the Italian database did not contain information on days-in-milk. Only Denmark and Italy are shown because the French database was not usable and incomplete and the Belgian data provider did not send information about data availability beforehand.

Table 34: Differences between reported availability (Objective 3) and actually received (Objective 4) data

Elements	**Denmark**			**Italy**
	Reported	Received	Reported	Received
Parity	Y	N	Y	Y
Housing system	N	N	N	N
Floor type	N	N	N	N
Days in milk	Y	Y[(a)]	Y	N
Access to pasture	N	N	N	N
Milk production	Y	Y[(a)]	N	N
Herd size	Y	Y	Y	Y
Breed	Y	Y	Y	Y
Geographical region	Y	Y	Y	Y
Organic dairy production	N	N	N	N
Lameness	N	N	N	N
Somatic cell count	Y	Y	Y	Y
Mortality	Y	Y	Y	Y

Y = Yes
N = No
(a): Provided but could not be used

11. Objective 5

11.1. Descriptive analyses

11.1.1. The IZSLER/CReNBA data

The overall welfare measure is given as the percentage of the maximum a given herd obtained. Thus the overall welfare measure (the score percentage) is a continuous variable potentially ranging from 0 to 100. The mean of the overall score percentage was 69.23 (SD = 10.63, min = 30.64, max = 93.02). The frequency distribution of the overall score percentage is shown in Figure 3: .

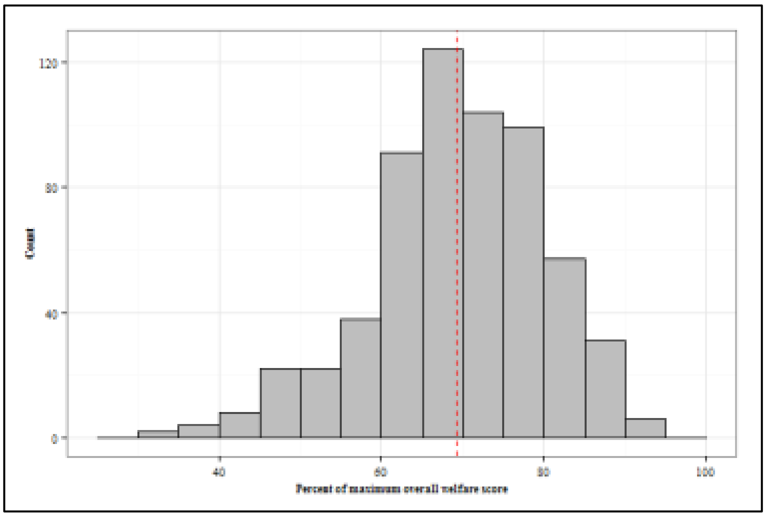


Figure 3: Frequency distribution of the overall score percentage, IZSLER/CReNBA data, N = 608. Dashed line indicates mean overall score percentage

All three ABMs were identified in the data. All three ABMs were categorical variables: mortality (<2% / 2-5% / >5%), SCC (<300,000 / 300,000-400,000 / >400,000) and lameness (<4% / 4-8% / >8%). The following factors of variation were included in the descriptive analysis: housing system (categorical: tethering / loose house / loose house with access to outdoor exercise area), floor (categorical: only good surface / >50% good surface / unsuitable), herd size (number of animals, continuous) and milk yield (daily milk yield, continuous).See Appendix M. for detailed description of the variables. The remaining of the selected factors of variation was not identified. Boxplots of the overall score percentage by levels of the ABMs and the categorical factors of variation are shown in Figure 4: and Figure 5: , respectively. The mean herd size was 261 cows (SD = 198, min = 12, max = 1413). The mean milk yield per cow per day was 28,2 kg (SD = 4.2, min = 11, max = 41). Scatterplots of the overall score percentage versus herd size and milk yield are shown in Figure 6: . The relationships between the ABMs and the factors of variation were illustrated graphically by boxplots (Figure 7: and Figure 8:) and barplots (Figure 9: and Figure 10:).

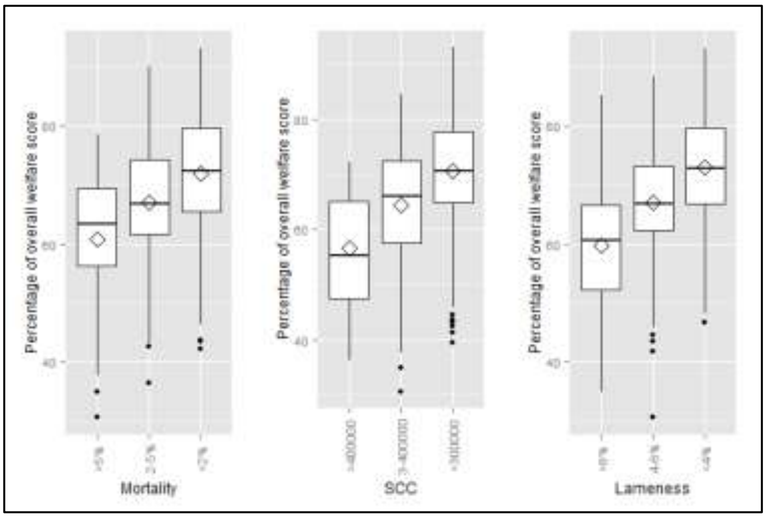


Figure 4: Boxplots of the mean of the overall score percentage by the levels of the ABMs. IZSLER/CReNBA data, N = 608.

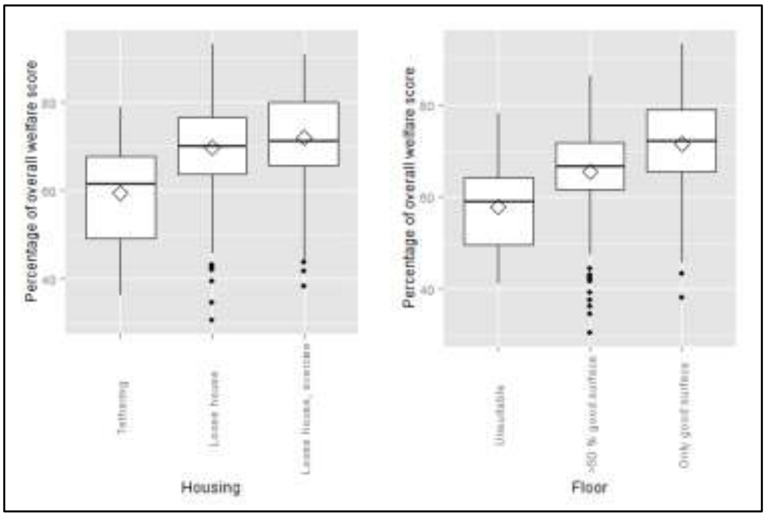


Figure 5: Boxplots of the mean of the overall score percentages by the levels of the categorical factors of variation. IZSLER/CReNBA data, N = 608.

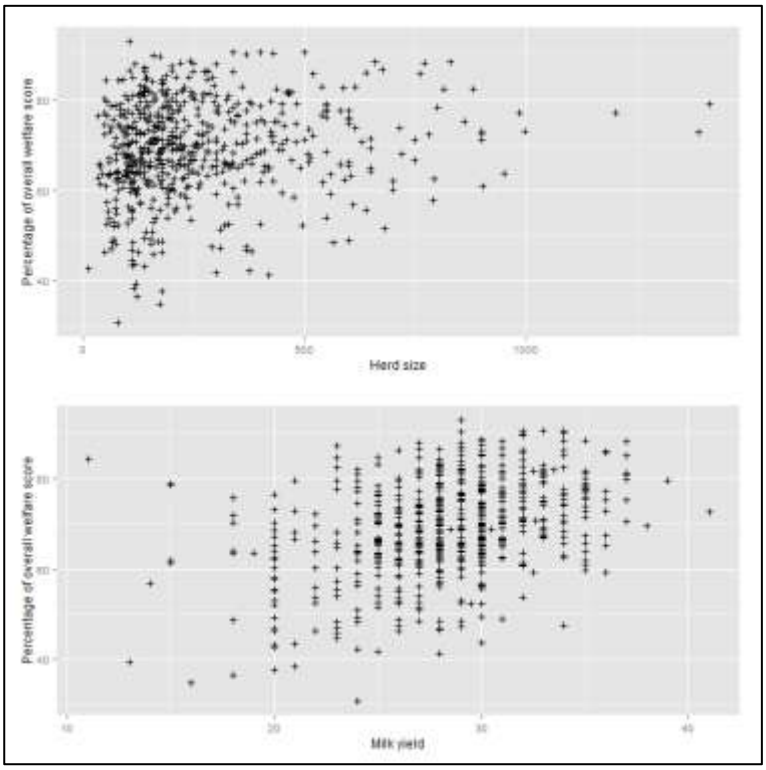


Figure 6: Scatterplots of the overall score percentage versus the herd size and the daily milk yield. IZSLER/CReNBA data, N = 608.

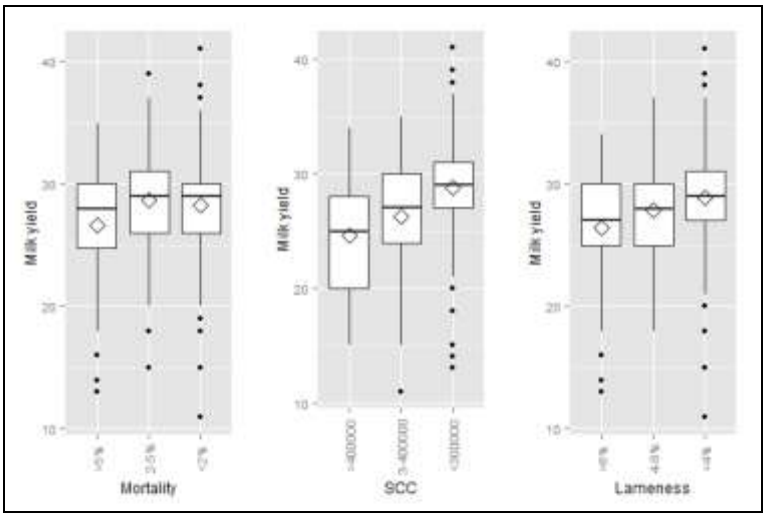


Figure 7: Boxplots of the herd size by levels of the ABMs. IZSLER/CReNBA data, N = 608.

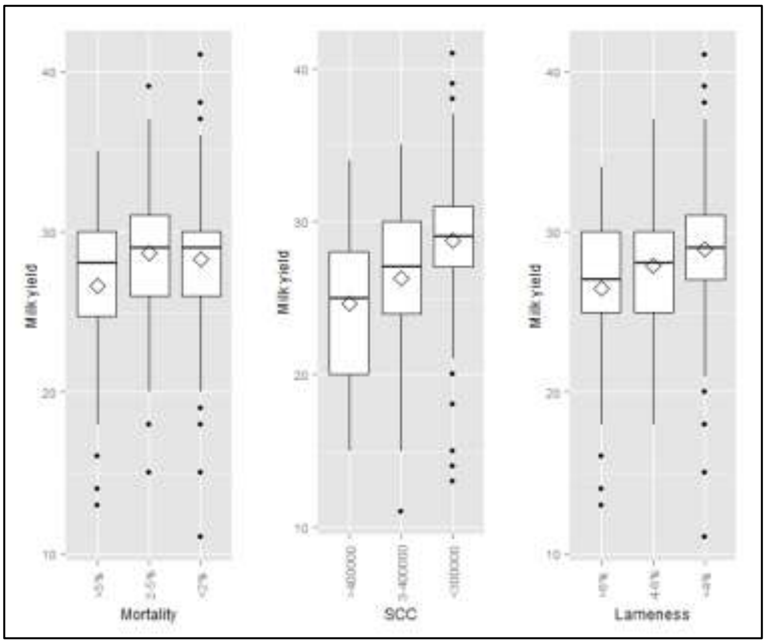


Figure 8: Boxplots of the milk yield by levels of the ABMs. IZSLER/CReNBA data, N = 608.

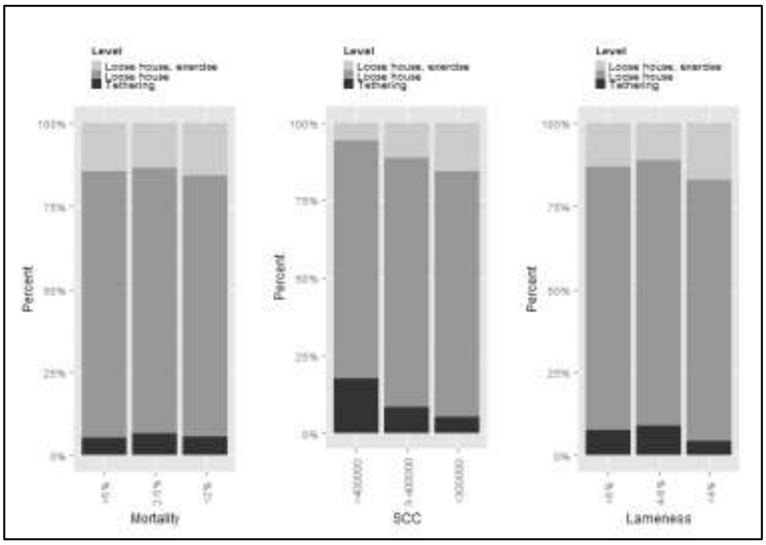


Figure 9: Barplots of the housing system by ABMs. IZSLER/CReNBA data, N = 608.

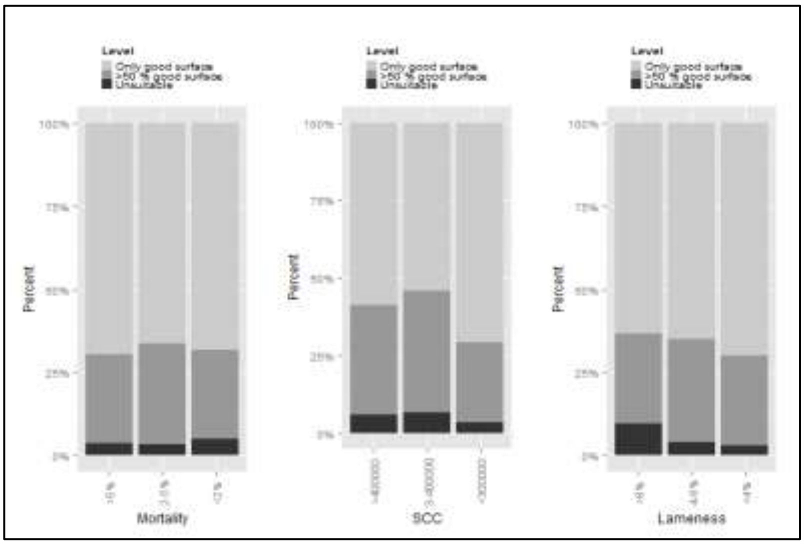


Figure 10: Barplots of the floor type by ABMs levels. IZSLER/CReNBA data, N = 608.

For the statistical analyses, the overall welfare score percentages were dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 35: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 35: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. IZSLER/CReNBA data, N = 608.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	>5%	43	13	1	55	0	56
	2-5%	127	80	32	175	8	199
	<2%	133	212	119	226	53	292
SCC	>400,000	16	1	0	17	0	17
	3-400,000	68	37	16	89	7	98
	<300,000	219	267	136	350	54	432
Lameness	>8%	77	16	6	87	1	92
	4-8%	101	68	24	145	8	161
	<4%	125	221	122	224	52	294
Floor type	Unsuitable	24	2	1	25	0	26
	>50 % good surface	112	59	16	155	4	167
	Only good surface	167	244	135	276	57	354
Housing system	Tethering	30	7	1	36	0	37
	Loose house	234	247	114	367	44	437
	Loose house, exercise	39	51	37	53	17	73
Herd size	Mean	234.4	287.1	296.8	248.9	313.5	255.5
Milk yield	Mean	27.0	29.4	29.8	27.7	29.9	28.1

11.1.2. **The Otten data**

The Otten data contains three different welfare measures: an animal based measure, a register data based measure and a system based measure. These three welfare measures are analysed separately. However, some of the descriptive analyses are equivalent and will only be presented with the first measure; the animal based measure.

11.1.2.1. The Animal Based welfare measure

The Otten animal based welfare measure was a continuous variable potentially ranging from 0 to 3900. The mean of the index was 946.91 (SD = 263.29, min = 449.99, max = 1756.50). The frequency distribution of the animal based index is shown in Figure 11: .

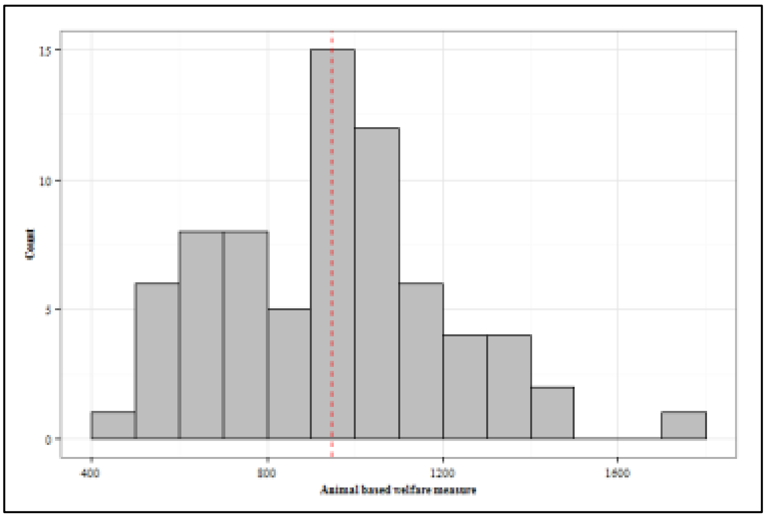


Figure 11: Frequency distribution of animal based welfare measure, Otten data, N=72. Dashed line indicates mean animal based measure.

All three ABMs were identified in the data. All three ABMs are categorical variables: mortality (<5% / >=5%), SCC (<300,000 / >=300,000) and lameness (<28% / >=28%). The following factors of variation were included in the descriptive analysis: production type (binary: conventional / organic), Breed (categorical: Danish Holstein (DH) / Red Danish Dairy (RDD) / Jersey / Crossbreed (CB)), herd size (number of animals, continuous) and milk yield (kg ECM per cow year, continuous). See Appendix M. for detailed description of the variables. The remaining of the selected factors of variation was not identified. Boxplots of the Otten animal based welfare measure by levels of the ABMs and the categorical factors of variation are shown Figure 12: and Figure 13: , respectively. The mean herd size was 181 cows (SD = 80, min = 93, max = 518). The mean milk yield (kg ECM per cow year) was 9126 kg ECM (SD = 1458, min = 5447, max = 11860). Scatterplots of the Otten animal based welfare measure versus herd size and milk yield are shown in Figure 14: . The relationships between the ABMs and the factors of variation were illustrated graphically by boxplots (Figure 15: and Figure 16:) and barplots (Figure 17: and Figure 18:).

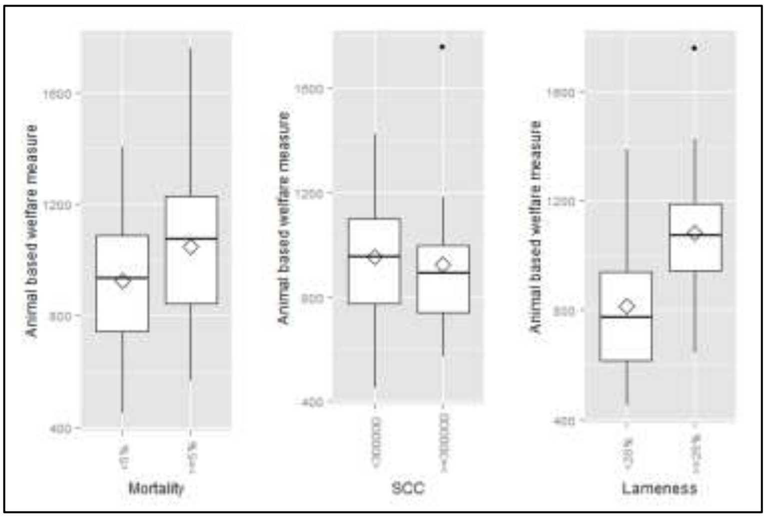
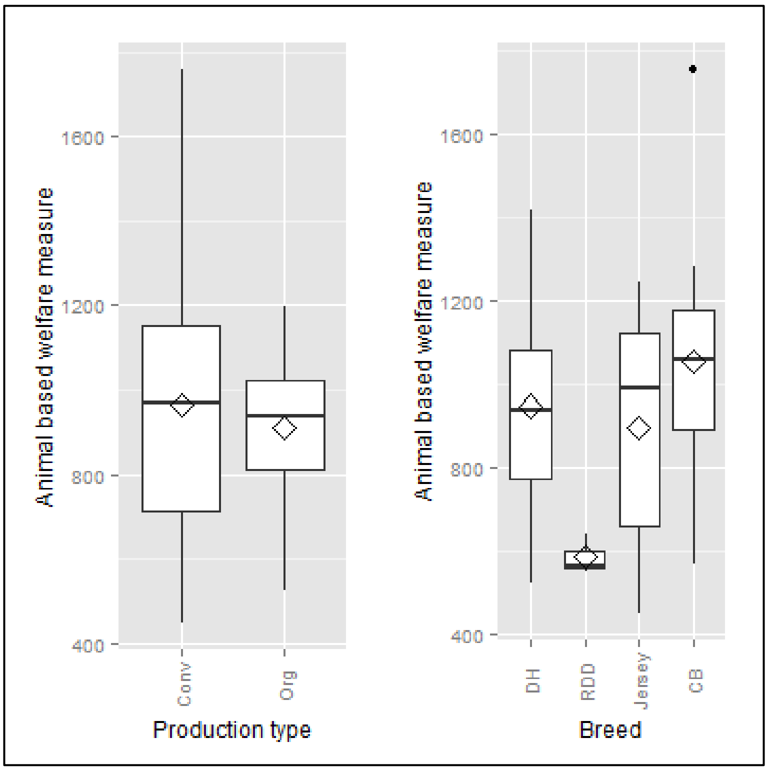


Figure 12: Boxplots of the animal based welfare measure by levels of the ABMs, Otten data, N = 72.



DH: Danish Holstein, RDD: Red Danish Dairy, CB: Crossbreed, Conv: conventional, Org: organic

Figure 13: Boxplots of the animal based welfare measure by levels of categorical factors of variation, Otten data, N = 72.

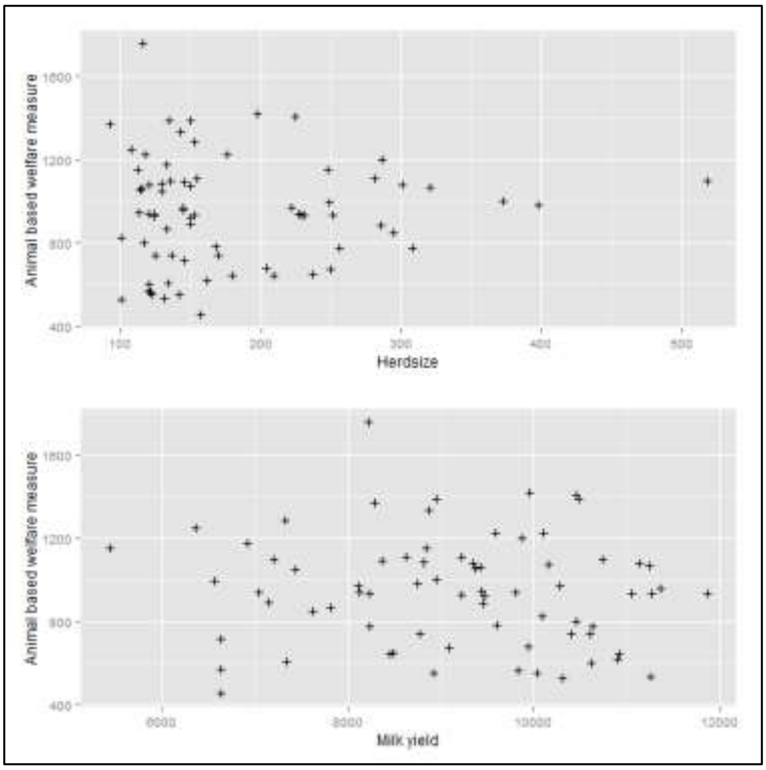


Figure 14: Scatterplots of the herd size and milk yield versus animal based welfare measure, Otten data, N = 72.

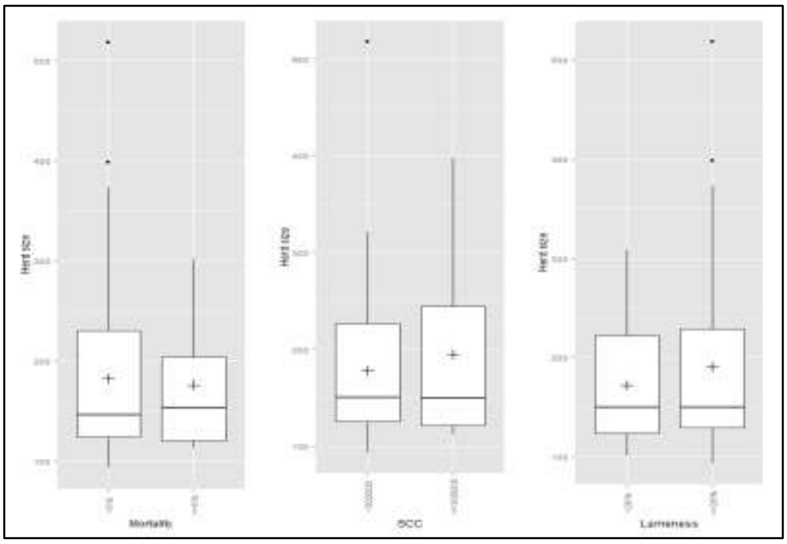


Figure 15: Boxplots of the herd size by levels of the ABMs, Otten data, N = 72.

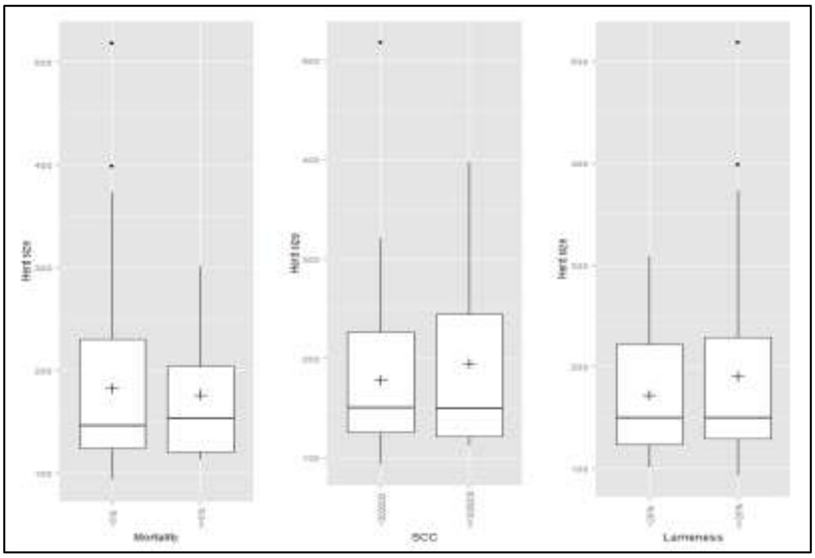
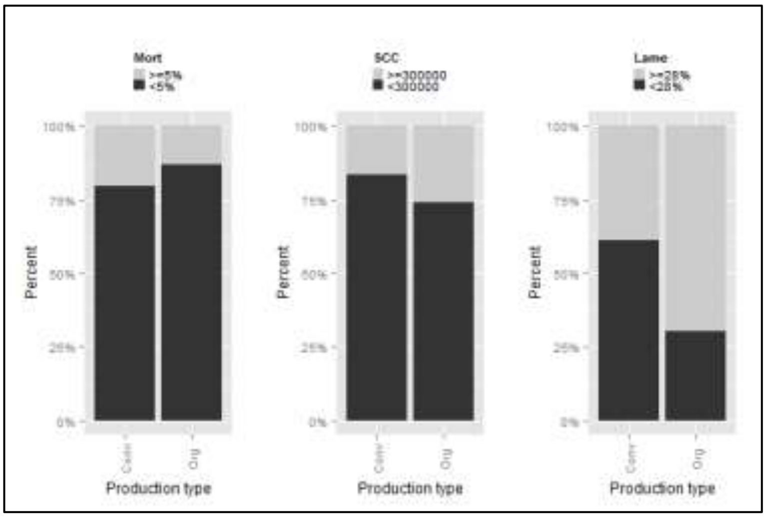
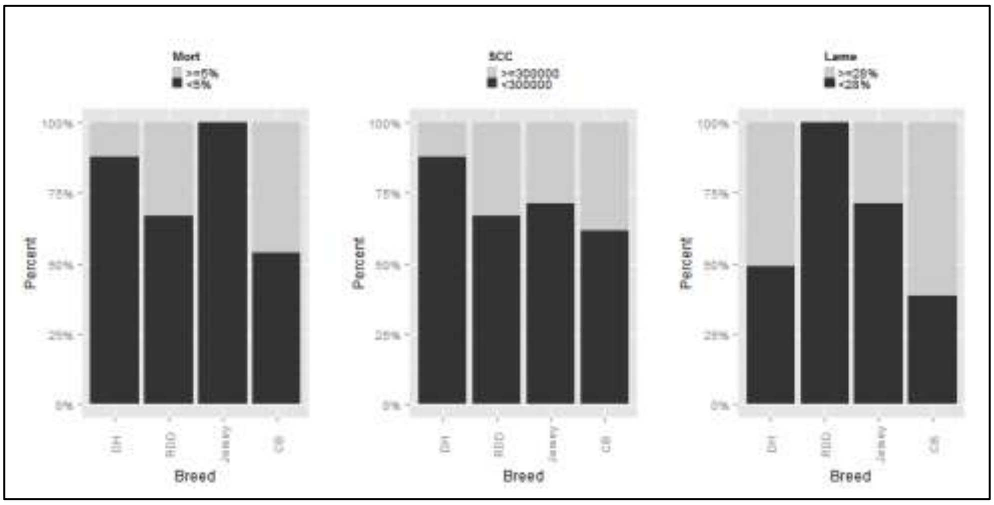


Figure 16: Boxplots of the milk yield by levels of the ABMs, Otten data, N = 72.



Conv: conventional, Org: organic, SCC: somatic cell count, Lame: lameness, Mort: mortality

Figure 17: Barplots of the production type by ABM levels, Otten data, N = 72.



DH: Danish Holstein, RDD: Red Danish Dairy, Jersey: Jersey, CB: Crossbreed. SCC: somatic cell count, Lame: lameness, Mort: mortality

Figure 18: Barplots of the Breed by ABM levels, Otten data, N = 72.

For the statistical analyses, the animal based welfare measure was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 36: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes. Due to the severely skewed distribution in the breed variable, it was decided to omit this variable from the statistical analyses of the Otten data.

Table 36: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Otten data, animal based welfare measure, N = 72.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	28	31	13	46	5	54
	>=5%	8	5	5	8	3	10
SCC	<300,000	31	27	15	43	7	51
	>=300,000	5	9	3	11	1	13
Lameness	<28%	9	28	4	33	2	35
	>=28%	27	8	14	21	6	29
Production type	Conventional	26	23	15	34	8	41
	Organic	10	13	3	20	0	23
Breed	Danish Holstein	24	25	11	38	6	43
	Red Danish Dairy	0	3	0	3	0	3
	Jersey	4	3	3	4	0	7
	Crossbreed	8	5	4	9	2	11
Herd size	Mean	191	171	165	187	152	185
Milk yield	Mean	8953	9299	8595	9303	9064	9134

11.1.2.2. The register- data based welfare measure

The Otten register-based welfare measure was a continuous variable potentially ranging from 0 to 196.5. The mean was 69.94 (SD = 16.79, min = 41.00, max = 125.00). The frequency distribution of Otten register-based welfare measure is shown in Figure 19: .

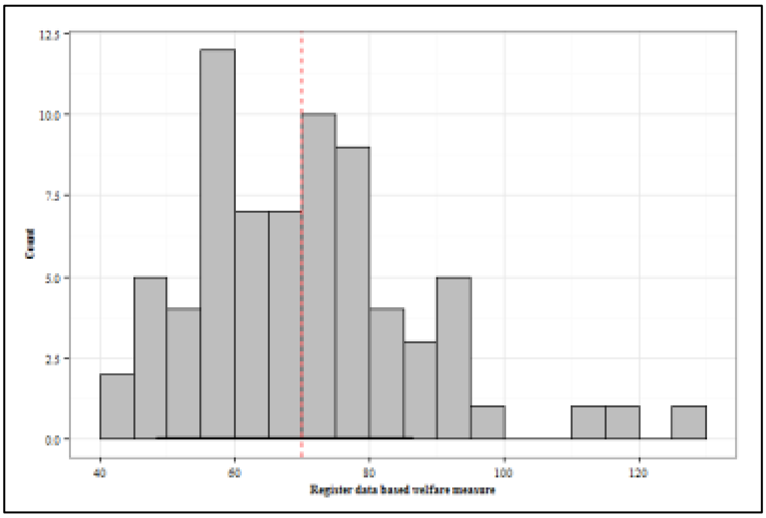


Figure 19: Frequency distribution of the register-based welfare measure, Otten data, N = 72.

The description of the ABMs and the factors of variation in the analysis of the Otten register-based welfare index are equivalent to description in the previous section regarding the Otten animal based welfare index. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them). Boxplots of the Otten register-based welfare measure by levels of the ABMs and the categorical factors of variation are shown in Figure 20: and Figure 21: respectively. Scatterplots of the Otten register-based welfare measure versus the two continuous factors of variation are shown in Figure 22: .

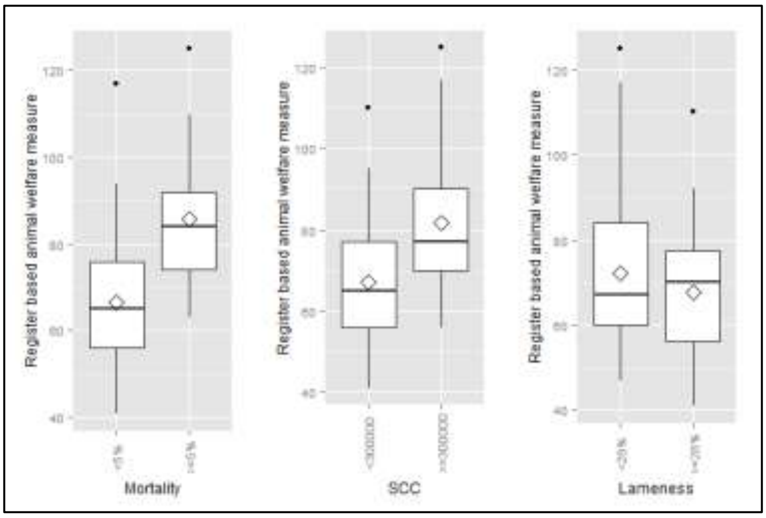
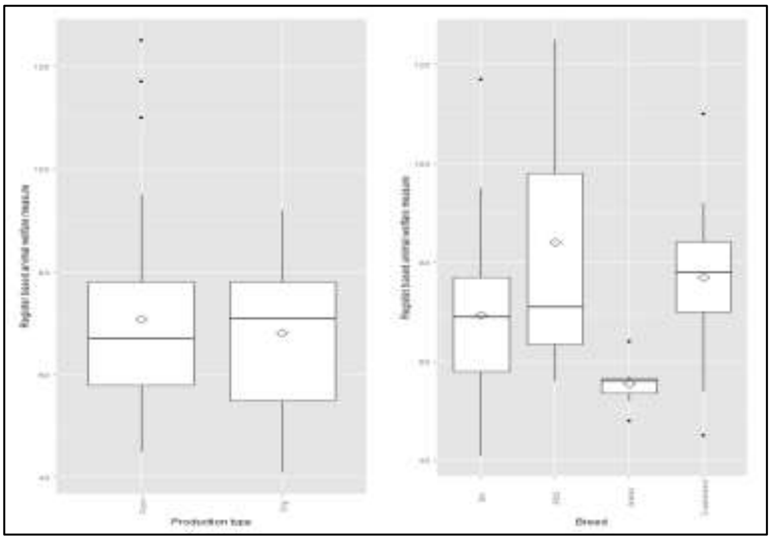


Figure 20: Boxplots of the register- based welfare index by levels of the ABMs, Otten data, N = 72.



Conv: conventional, Org: organic, DH: Danish Holstein, RDD: Red Danish Dairy,

Figure 21: Boxplots of the register- based welfare index by levels of categorical factors of variation, Otten data, N = 72.

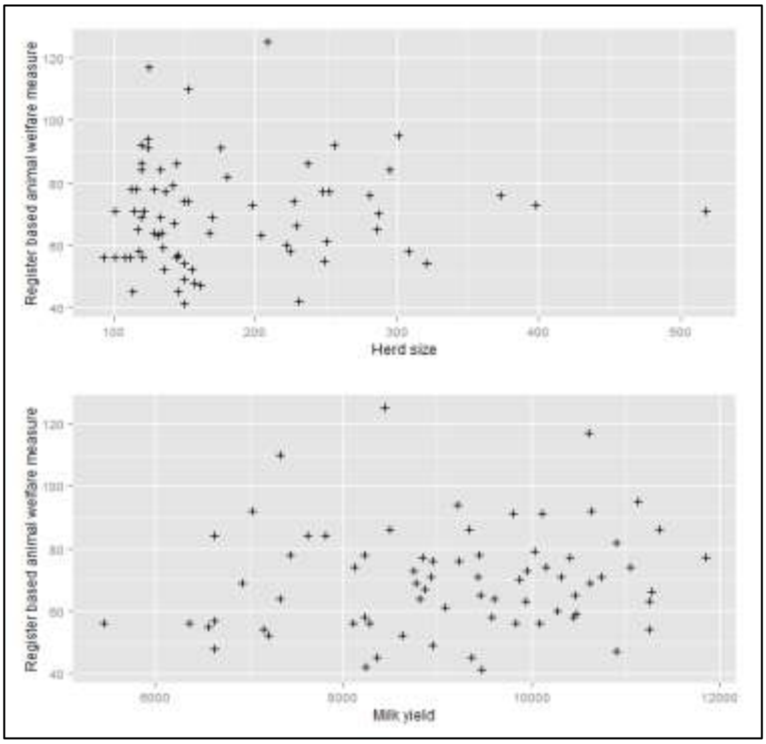


Figure 22: Scatterplots of the herd size and milk yield versus the register-based welfare measure, Otten data, N = 72.

For the statistical analyses, the register based welfare measure was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 37: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 37: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Otten data, animal based welfare measure, N = 72.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	24	35	10	49	3	56
	>=5%	11	2	7	6	4	9
SCC	<300,000	25	33	11	47	3	55
	>=300,000	10	4	6	8	4	10
Lameness	<28%	17	20	11	26	5	32
	>=28%	18	17	6	29	2	33
Production type	Conventional	22	27	12	37	5	44
	Organic	13	10	5	18	2	21
Breed	Danish Holstein	23	26	11	38	4	45
	Red Danish dairy	2	1	1	2	1	2
	Jersey	0	7	0	7	0	7
	Crossbreed	10	3	5	8	2	11
Herd size	Mean	197	167	174	183	184	181
Milk yield	Mean	9377	8889	9206	9102	9199	9118

11.1.2.3. The system based welfare measure

The Otten system based welfare measure is a continuous variable potentially ranging from 0 to 52. The mean was 33.23 (SD = 5.20, min = 18.00, max = 43.50). The frequency distribution of the Otten systems based welfare measure is shown in Figure 23: .

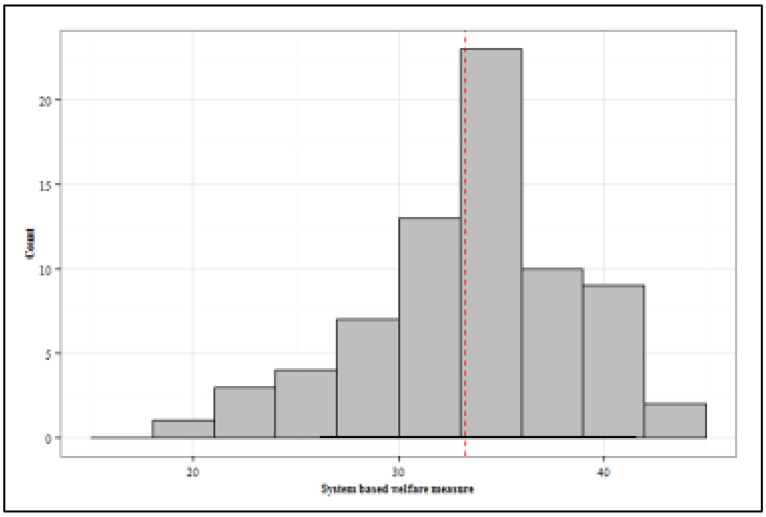


Figure 23: Frequency distribution of the system based welfare measure, Otten data, N =72.

The description of the ABMs and the factors of variation in the analysis of the Otten system-based welfare index is equivalent to description in the previous section regarding the Otten animal based welfare index. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them). Boxplots of the Otten system based welfare measure by levels of the ABMs and the categorical factors of variation are shown in Figure 24: and Figure 25: respectively. Scatterplots of Otten systems based welfare measure versus the two continuous factors of variation are shown in Figure 26: .

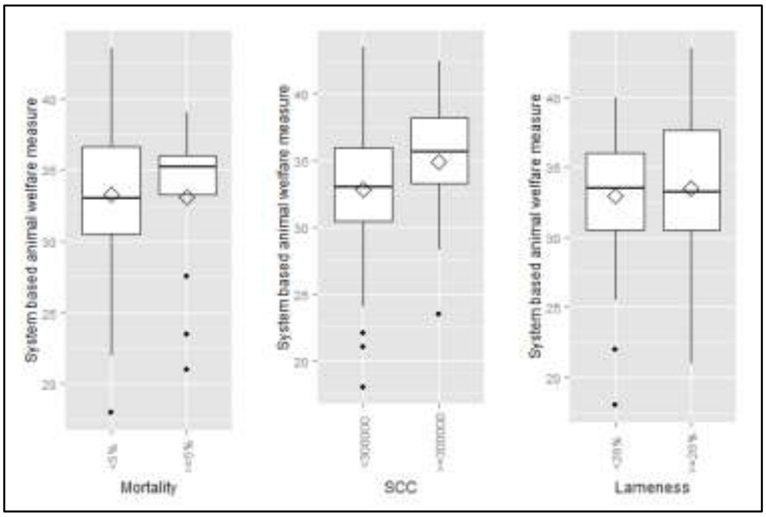
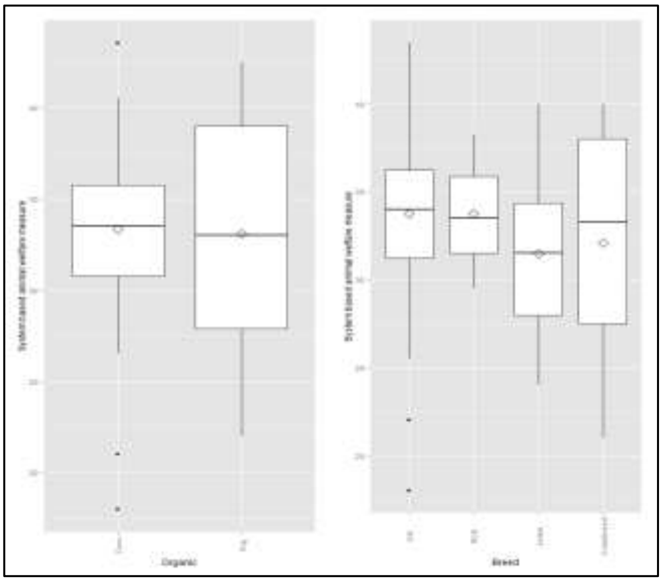


Figure 24: Boxplots of the systems-based welfare score by levels of the ABMs, Otten data, N =72.



DH: Danish Holstein, RDD: Red Danish Dairy, Conv: conventional, Org: organic

Figure 25: Boxplots of the system based welfare measure by levels of categorical factors of variation, Otten data, N =72.

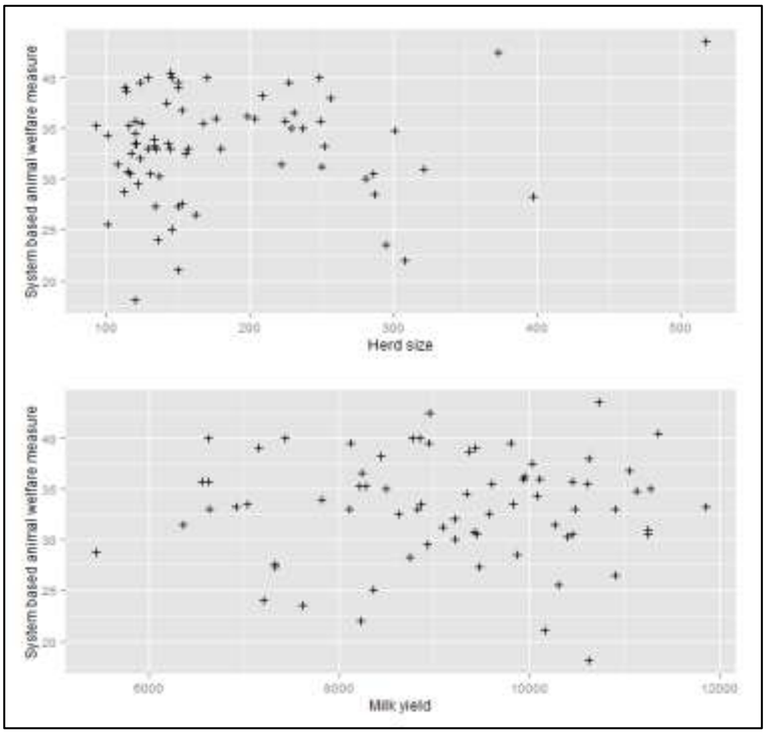


Figure 26: Scatterplots of the herd size and milk yield versus system based welfare measure, Otten data, N = 72.

For the statistical analyses, the system based welfare measure was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 38: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 38: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Otten data, system based welfare measure, N = 72.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	26	33	16	43	7	52
	>=5%	8	5	2	11	0	13
SCC	<300,000	25	33	12	46	4	54
	>=300,000	9	5	6	8	3	11
Lameness	<28%	18	19	8	29	1	36
	>=28%	16	19	10	25	6	29
Production type	Conventional	24	25	8	41	4	45
	Organic	10	13	10	13	3	20
Breed	Danish Holstein	25	24	12	37	5	44
	Red Danish Dairy	1	2	1	2	0	3
	Jersey	2	5	1	6	1	6
	Crossbreed	6	7	4	9	1	12
Herd size	Mean	188	175	200	175	247	174
Milk yield	Mean	9190	9069	9109	9132	8961	9144

11.1.3. **The Burow data**

The overall summer welfare measure is a continuous variable potentially ranging from 0 to 5400. The mean of the overall summer index was 2905.23 (SD = 638.19, min = 1803.59, max = 4727.89). The frequency distribution of the overall summer index is shown in Figure 27: .

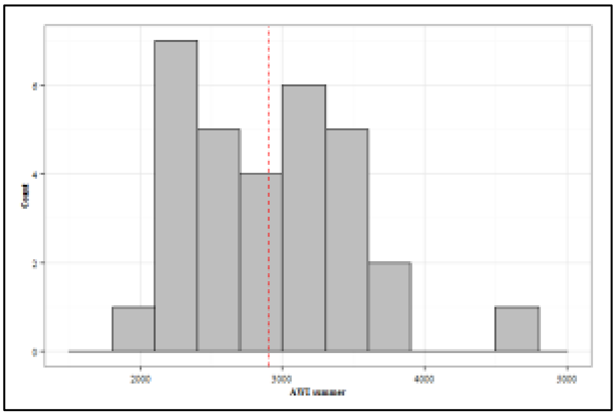


Figure 27: Frequency distribution of AWI, Summer measure, Burow data, N = 31. Dashed line indicates mean.

All three ABMs were identified in the data. All three ABMs are categorical variables: mortality (<3% / >=3%), SCC (<300,000 / >=300,000) and lameness (<46% / >=46%). The following factors of variation were included in the descriptive analysis: production system (categorical: conventional /organic), breed (categorical: DH / Jersey / CB), herd size (number of animals, continuous) and milk yield (kg ECM per cow year, continuous). See Appendix M. for detailed description of the variables. The remaining of the selected factors of variation was not identified. Boxplots of the overall summer index by levels of the ABMs and the categorical factors of variation are shown in Figure 28: and Figure 29: respectively. The mean herd size was 172 cows (SD = 70, min = 93, max = 373). The mean milk yield (kg ECM per cow year) was 8681 kg ECM (SD = 1279, min = 6356, max = 11056). Scatterplots of the overall summer welfare measure versus the two continuous factors of variation are shown in Figure 30: . The relationships between the ABMs and the factors of variation were illustrated graphically by boxplots (Figure 31: and Figure 32:) and barplots (Figure 33: and Figure 34:).

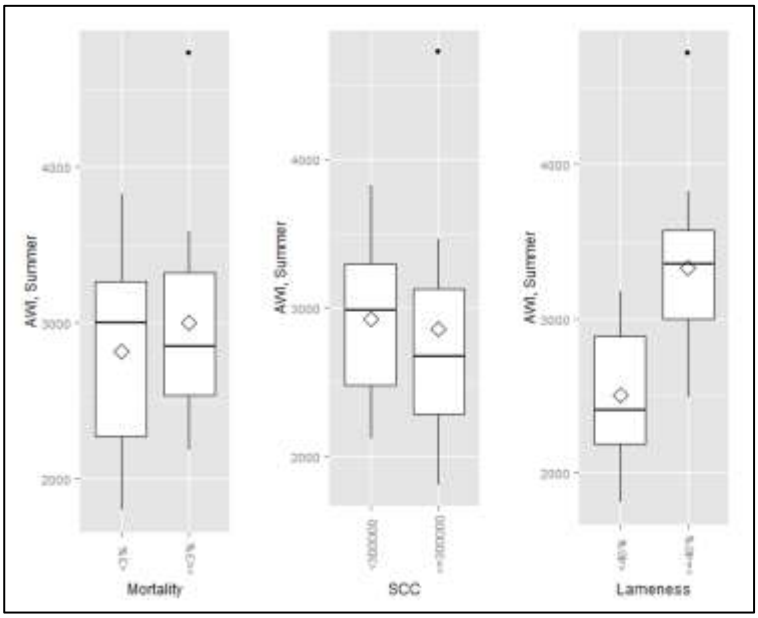
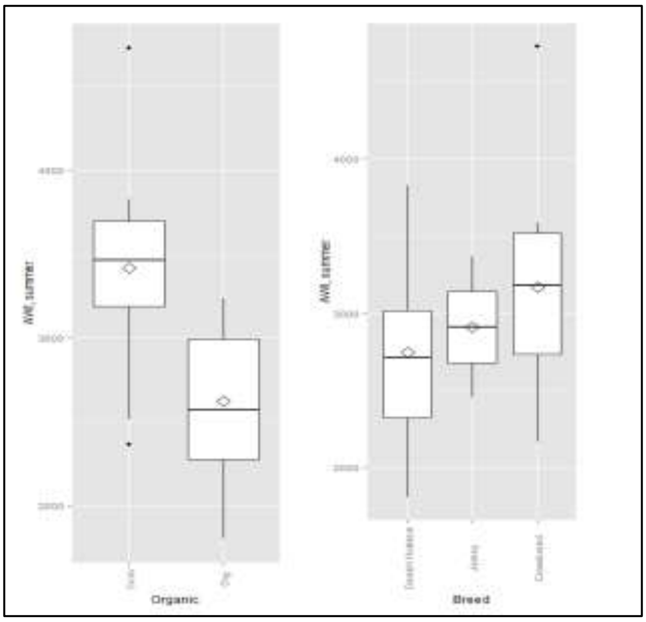


Figure 28: Boxplots of the AWI, Summer welfare measure by levels of the ABMs. Burow data, N = 31.



Conv: conventional, Org: organic

Figure 29: Boxplots of the AWI, Summer welfare measure by levels of categorical factors of variation. Burow data, N = 31.

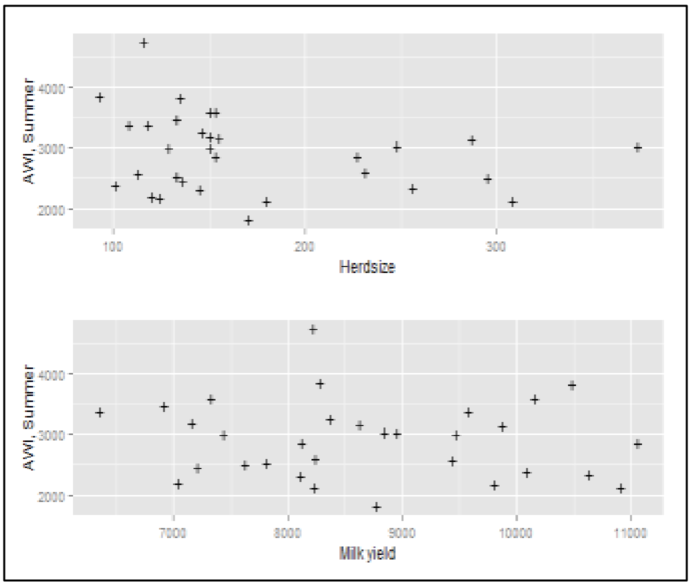


Figure 30: Scatterplots of the herd size and milk yield versus the AWI, Summer welfare measure. Burow data, N =31.

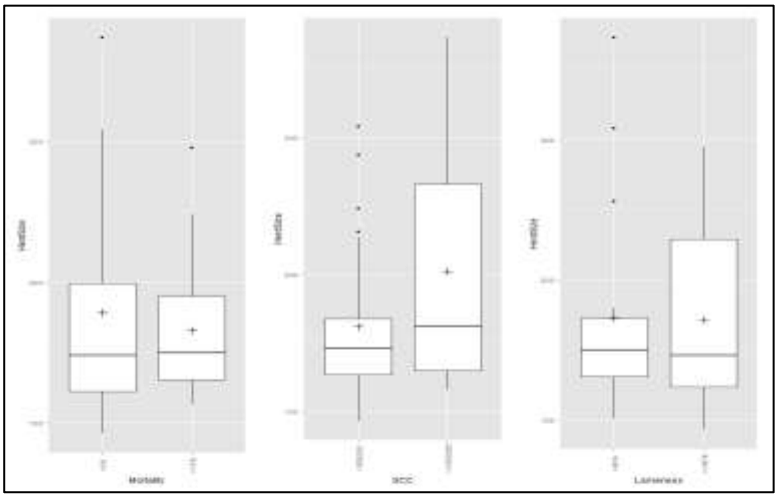


Figure 31: Boxplots of the herd size by levels of the ABMs. Burow data, N = 31.

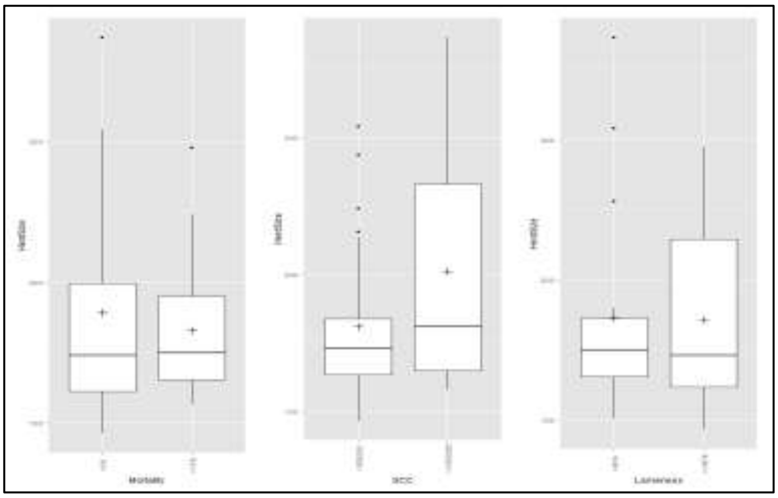
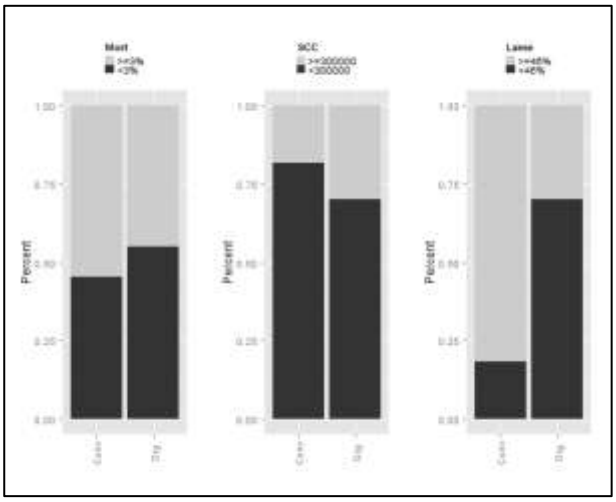
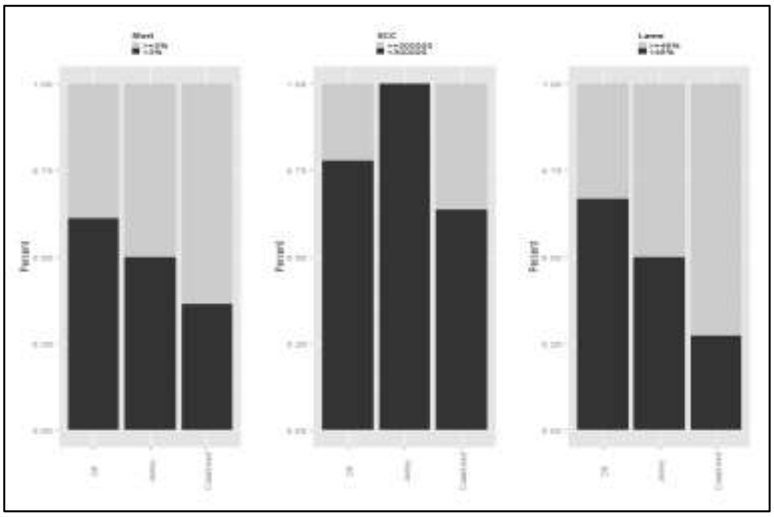


Figure 32: Boxplots of the milk yield by levels of the ABMs. Burow data, N = 31.



Conv: conventional, Org: organic, SCC: somatic cell count, Lame: lameness, Mort: mortality

Figure 33: Barplots of the production type by the levels of the ABMs. Burow data, N = 31.



SCC: somatic cell count, Lame: lameness, Mort: mortality

Figure 34: Barplots of breed by the levels of the ABMs. Burow data, N = 31.

For the statistical analyses, the AWI, Summer welfare measure was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 39: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 39: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Burow data, AWI, Summer welfare measure, N = 31.

		Outcome median		**Outcome P25**		**Outcome P10**	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<3%	8	8	4	12	2	14
	>=3%	7	8	4	11	1	14
SCC	<300,000	12	11	6	17	2	21
	>=300,000	3	5	2	6	1	7
Lameness	<46%	3	13	0	16	0	16
	>=46%	12	3	8	7	3	12
Production type	Conventional	9	2	8	3	3	8
	Organic	6	14	0	20	0	20
Herd size	Mean	166	177	125	188	114	178
Milk yield	Mean	8438	8909	8415	8773	8993	8647

11.1.4. **The French Welfare Quality® data**

The WQ® data contains the criteria scores, the principle scores and the overall classification. The principle scores Health, Feeding, Housing and Behaviour are analysed separately. However, some of the descriptive analyses are equivalent and will only be presented with the first principle.

11.1.4.1. Health principle

The health principle score was a continuous variable. The mean was 30.7 (SD = 7.7, min = 13.2, max =54.4). The frequency distribution of the health principle score is shown in Figure 35: .

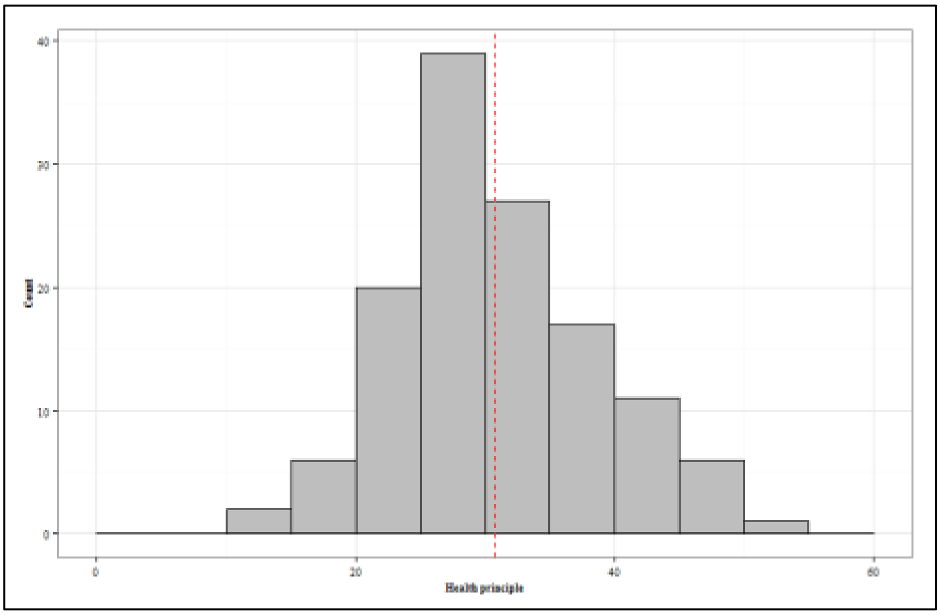


Figure 35: Frequency distribution of WQ® health principle score in the French WQ® data. N = 130. Dashed line indicates mean principle score.

All three ABMs were identified in the data. The three ABMs mortality, SCC and lameness were continuous variables. For the analyses, these were dichotomised according to the thresholds defined in Objective 2: mortality (0 = <5%, 1= >=5%), SCC (0 = <10 % of cows with SCC > 400,000, 1 = >=10 % of cows with SCC > 400,000) and lameness (0 = <8%, 1= >=8%). The following factors of variation were included in the descriptive analysis: pasture (binary, Yes/No), breed (binary, Milking/Double purpose), housing system (binary, Cubicles/Straw) and herd size (number of animals, continuous). See Appendix M. for detailed description of the variables. The remaining of the selected factors of variation was not identified. Boxplots of the health principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 36: and Figure 37: respectively. The mean herd size was 51 cows (SD = 17, min = 21, max = 120). A scatterplot of the WQ® Health principle score versus the herd size is shown in Figure 38: . The relationships between the ABMs and the factors of variation were illustrated graphically by boxplots (Figure 39:) and barplots (Figure 40:).

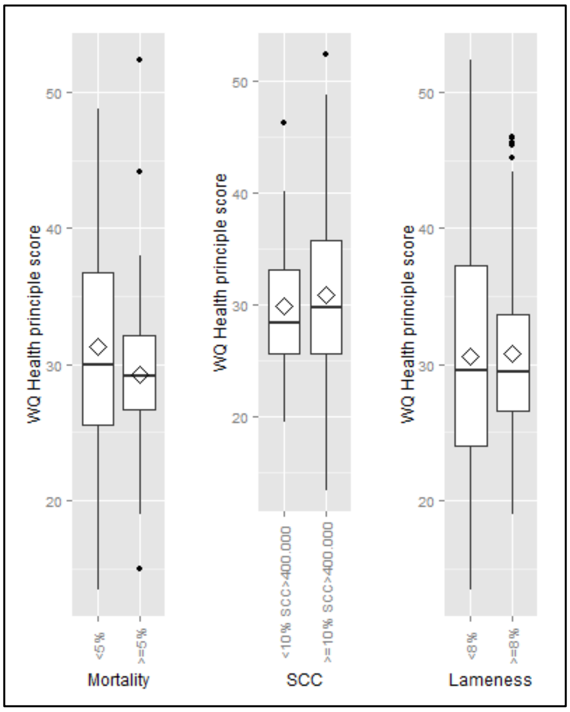
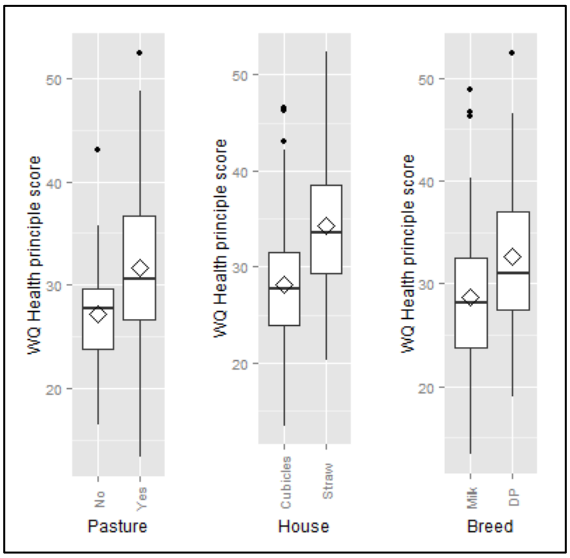


Figure 36: Boxplots of the WQ® health principle score by levels of the ABMs. French WQ® data, N=129.



Milk: milking breed, DP: Double purpose

Figure 37: Boxplots of the WQ® health principle score by levels of categorical factors of variation. French WQ® data, N=129.

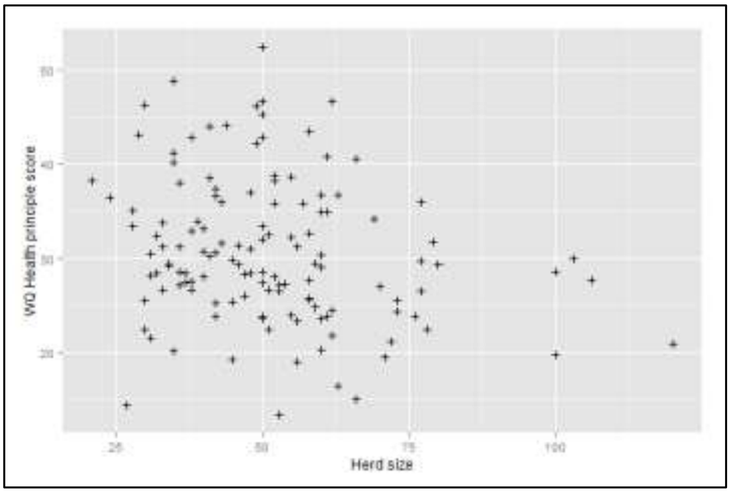


Figure 38: Scatterplot of WQ® health principle score versus herd size. French WQ® data, N=129

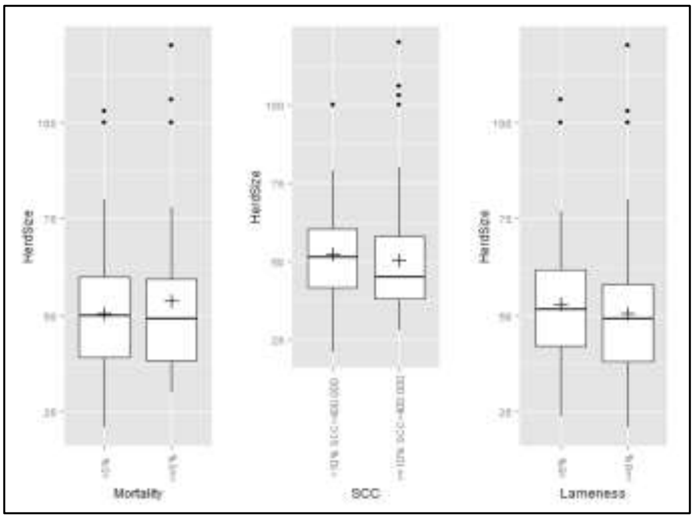
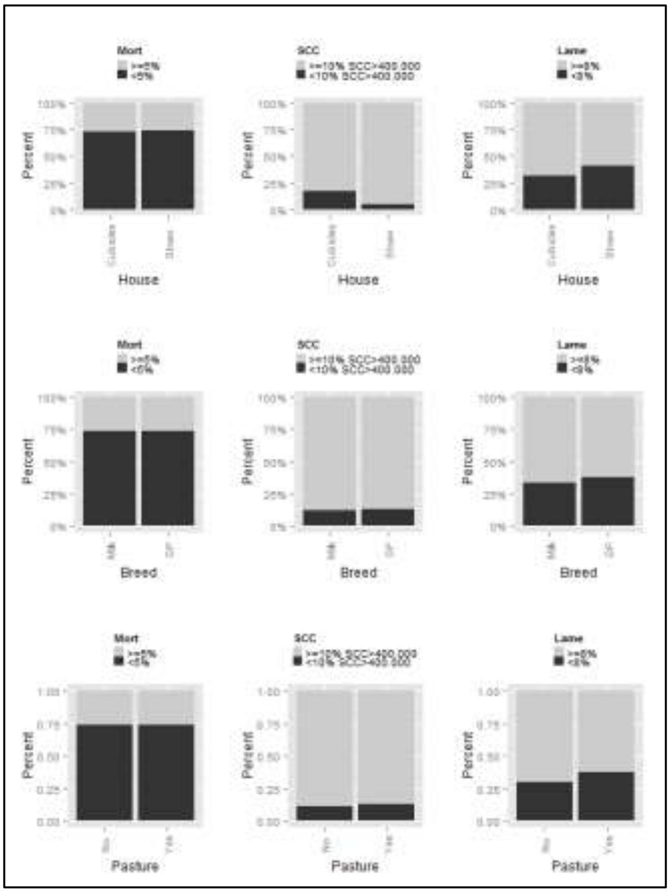


Figure 39: Boxplots of herd size by levels of the ABMs. French WQ® data, N = 129



SCC: somatic cell count, Lame: lameness, Mort: mortality, DP: Double purpose, House: Housing system

Figure 40: Barplots of the categorical factors of variations by the levels of the ABMs. French WQ® data, N=129.

For the statistical analyses, the health principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 40: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 40: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. French WQ® health principle score, N =129.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	45	50	25	70	9	86
	>=5%	19	15	7	27	4	30
SCC	<10% SCC>400,000	33	31	20	44	9	55
	>=10% SCC>400,000	31	34	12	53	4	61
Lameness	<8%	23	23	15	31	6	40
	>=8%	41	42	17	66	7	76
Housing system	Cubicles	49	26	26	49	11	64
	Straw	15	39	6	48	2	52
Breed	Milking Breed	36	24	21	58	11	49
	Double purpose	28	41	11	16	2	67
Pasture	No	20	7	11	16	4	23
	Yes	44	58	21	81	9	93
Herd Size	Mean	55	48	58	49	61	50

11.1.4.2. Housing principle

The housing principle score was a continuous variable. The mean was 58.1 (SD = 10.6, min = 35.2, max =79.8). The frequency distribution of the housing principle score is shown in Figure 41: .

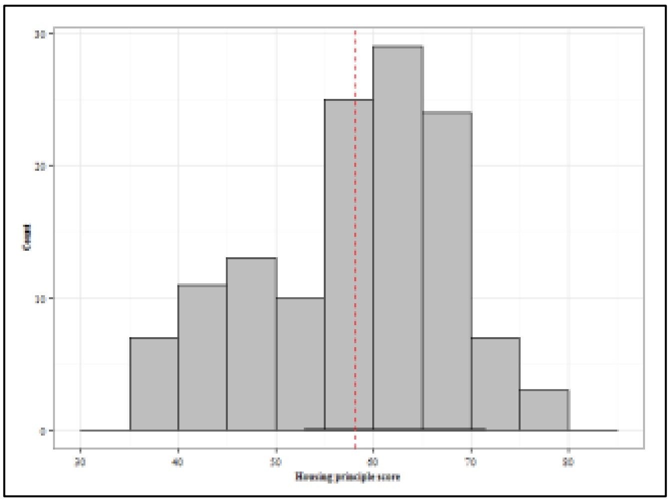


Figure 41: Frequency distribution of WQ® housing principle score in the French WQ® data. N = 129. Dashed line indicates mean principle score.

The description of the ABMs and the factors of variation in the analysis of the WQ® Housing principle score are equivalent to description in the previous section regarding the WQ® Health principle score. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them).

Boxplots of the WQ® Housing principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 42: and Figure 43: respectively. The relationship between the WQ® Housing principle score and the herd size is illustrated in the scatterplot in Figure 44: .

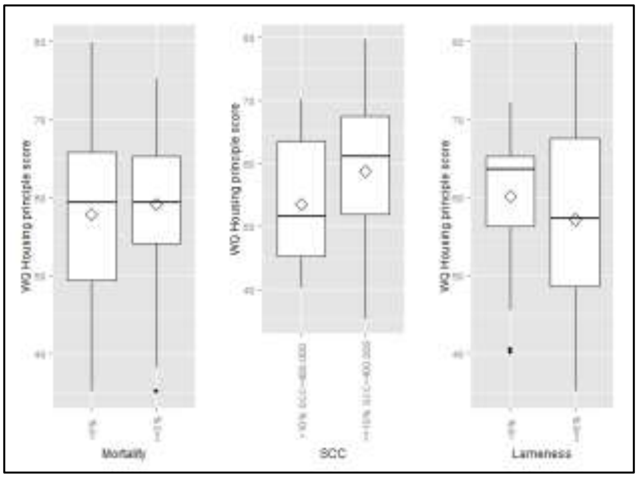
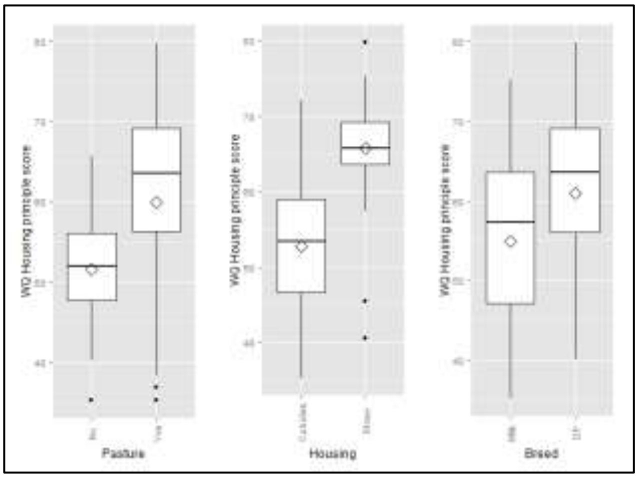


Figure 42: Boxplots of the WQ® housing principle score by the levels of the ABMs. French WQ® data, N = 129.



DP: Double purpose

Figure 43: Boxplots of the WQ® housing principle score by the levels of the categorical factors of variation. French WQ® data, N = 129.

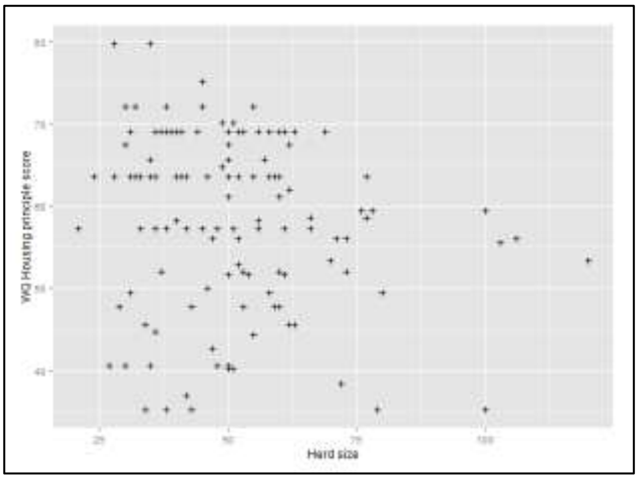


Figure 44: Scatterplot of WQ® housing principle score versus herd size. French WQ® data, N =129.

For the statistical analyses, the housing principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 41: Table 41: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 41: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. French WQ® housing principle score, N =129.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	47	48	25	70	8	87
	>=5%	16	18	6	28	2	32
SCC	<10% SCC>400,000	33	31	15	49	5	59
	>=10% SCC>400,000	30	35	16	49	5	60
Lameness	<8%	18	28	7	39	1	45
	>=8%	45	38	24	59	9	74
Housing system	Cubicles	56	19	29	46	10	65
	Straw	7	47	2	52	0	54
Breed	Milk Breed	34	26	8	37	9	51
	Double purpose	29	40	10	61	1	68
Pasture	No	24	3	10	17	3	24
	Yes	39	63	21	81	7	95
Herd Size	Mean	55	48	50	52	56	51

11.1.4.3. Behaviour principle

The behaviour principle score was a continuous variable. The mean was 36.9 (SD = 13.2, min = 6.5, max =69, N = 128). The frequency distribution of the behaviour principle score is shown in Figure 45: .

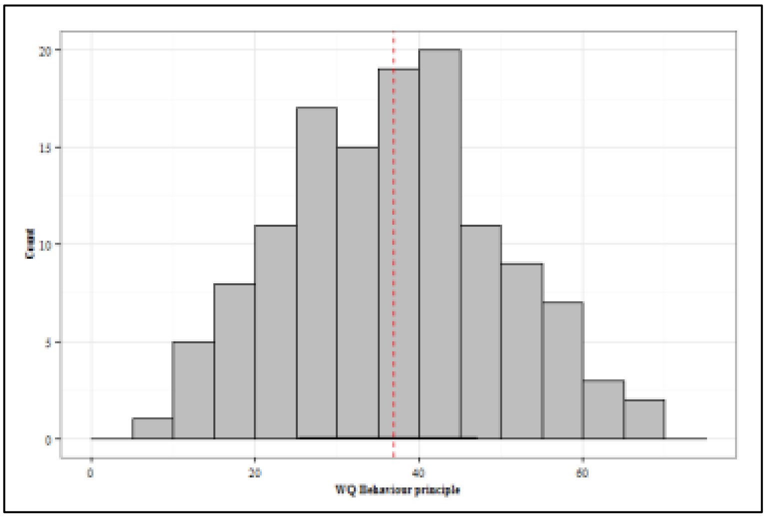


Figure 45: Frequency distribution of WQ® behaviour principle score in the French WQ® data. N = 128. Dashed line indicates mean principle score.

The description of the ABMs and the factors of variation in the analysis of the WQ® Behaviour principle score are equivalent to description in the previous section regarding the WQ® Health principle score. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them).

Boxplots of the WQ® Behaviour principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 46: and Figure 47: respectively. The relationship between the WQ® Housing principle score and the herd size is illustrated in the scatterplot in Figure 48: .

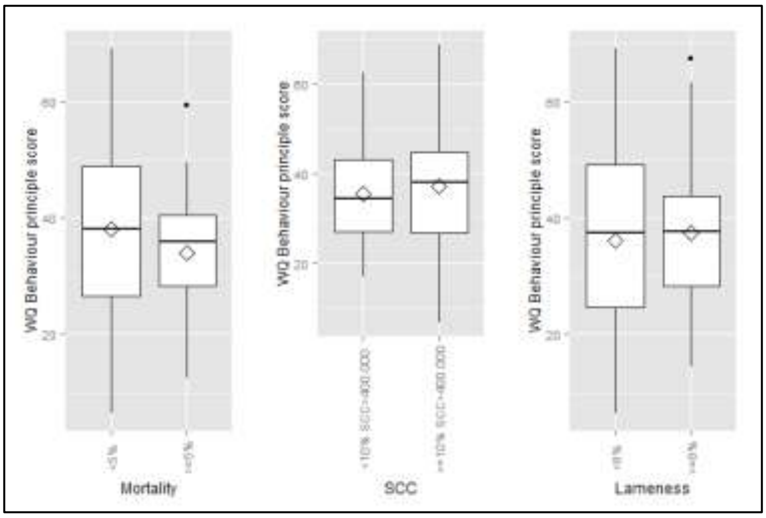
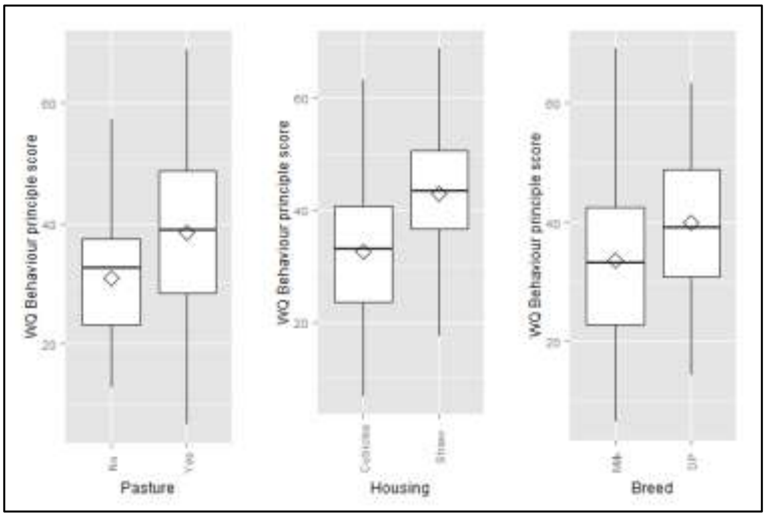


Figure 46: Boxplots of the WQ® behaviour principle score by the levels of the ABMs. French WQ® data, N = 128.



DP: Double purpose

Figure 47: Boxplots of the WQ® behaviour principle score by the levels of the categorical factors of variation. French WQ® data, N = 128.

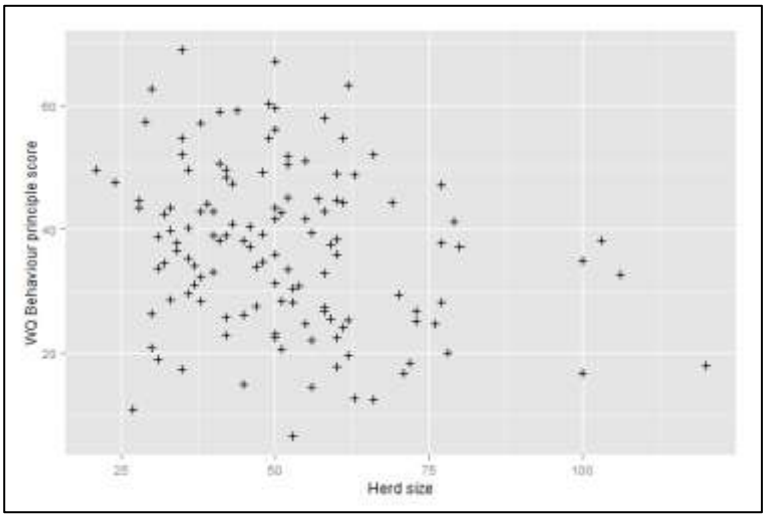


Figure 48: Scatterplot of WQ® behaviour principle score versus herd size. French WQ® data, N =128.

For the statistical analyses, the behaviour principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 42: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 42: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. French WQ® behaviour principle score, N =128.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	45	50	24	71	9	86
	>=5%	19	14	7	26	4	29
SCC	<10% SCC>400,000	33	31	19	45	9	55
	>=10% SCC>400,000	31	33	12	52	4	60
Lameness	<8%	23	22	15	30	6	39
	>=8%	41	42	16	67	7	76
Housing system	Cubicles	49	26	25	50	11	64
	Straw	15	38	6	47	9	51
Breed	Milk	36	24	21	39	11	49
	Double purpose	28	40	10	58	2	66
Pasture	No	20	7	10	17	4	23
	Yes	44	57	21	80	9	92
Herd size	Mean	55	47	57	49	61	50

11.1.4.4. Feeding principle

The feeding principle score was a continuous variable. The mean was 45.2 (SD = 28.1, min = 3.5, max =100, N = 129). The frequency distribution of the housing principle score is shown in Figure 49:

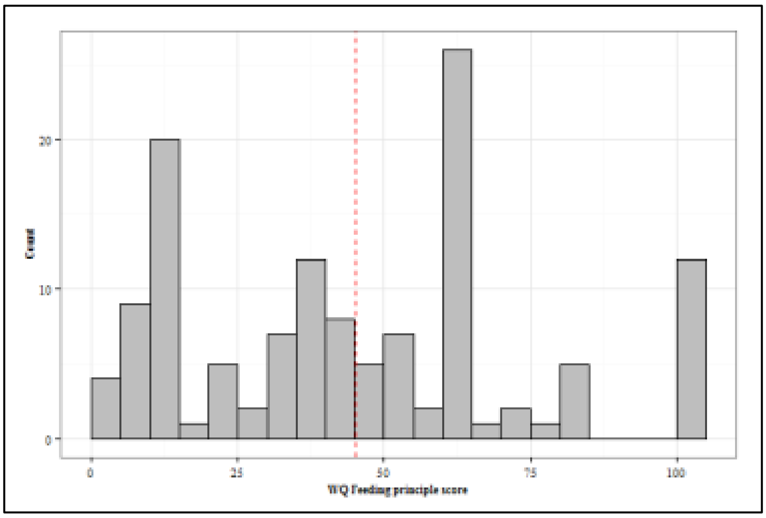


Figure 49: Frequency distribution of WQ® feeding principle score in the French WQ® data. N = 129. Dashed line indicates mean principle score.

The description of the ABMs and the factors of variation in the analysis of the WQ® Feeding principle score are equivalent to description in the previous section regarding the WQ® Health principle score. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them).

Boxplots of the WQ® Feeding principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 50: and Figure 51: respectively. The relationship between the WQ® Feeding principle score and the herd size is illustrated in the scatterplot in Figure 52: .

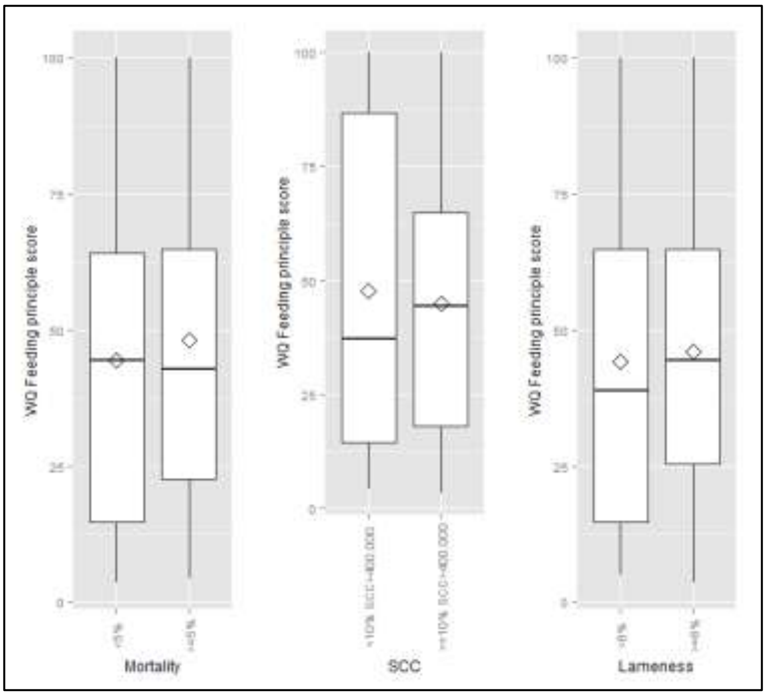
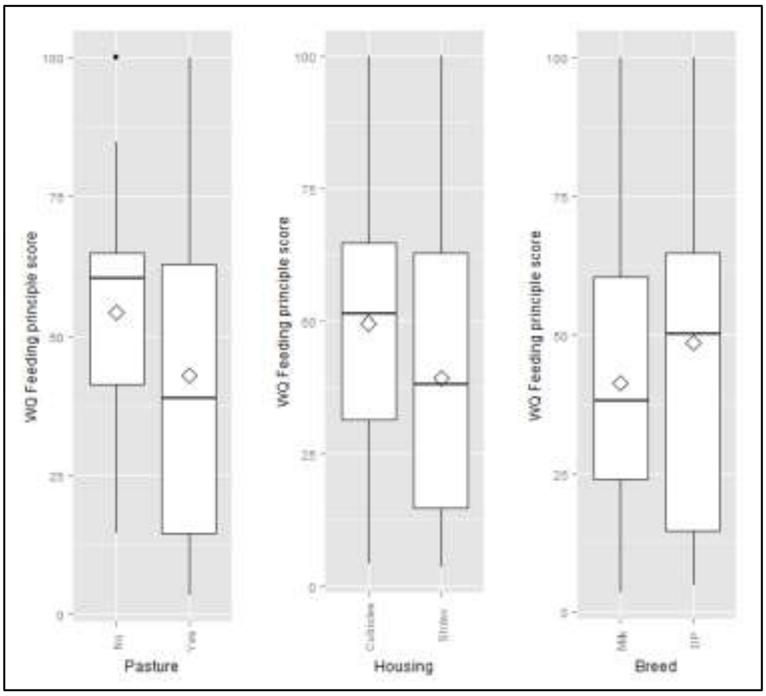


Figure 50: Boxplots of the WQ® feeding principle score by the levels of the ABMs. French WQ® data, N = 129.



DP: Double purpose

Figure 51: Boxplots of the WQ® feeding principle score by the levels of the categorical factors of variation. French WQ® data, N = 129.

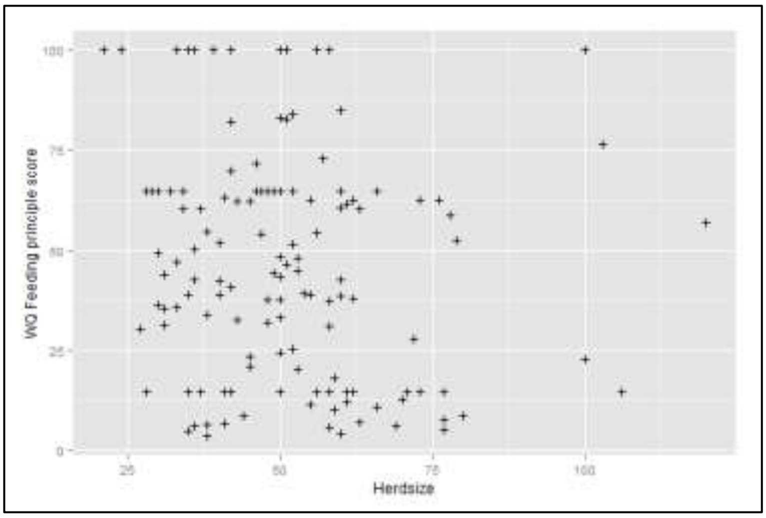


Figure 52: Scatterplot of WQ® feeding principle score versus herd size. French WQ® data, N =129.

For the statistical analyses, the feeding principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 43: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 43: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. French WQ® feeding principle score, N =129.

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<5%	46	49	13	82	10	85
	>=5%	18	16	5	29	3	31
SCC	<10% SCC>400,000	30	34	7	57	6	58
	>=10% SCC>400,000	34	31	11	54	7	58
Lameness	<8%	25	21	8	38	5	41
	>=8%	39	44	10	73	8	75
Housing system	Cubicles	33	42	6	69	4	71
	Straw	31	23	12	42	9	45
Breed	Milking	37	23	7	53	6	54
	Double purpose	27	42	11	58	7	62
Pasture	No	7	20	0	27	0	27
	Yes	57	45	18	84	13	89
Herd Size	Mean	53	50	57	50	55	51

11.1.5. The Belgian Welfare Quality® data

The WQ® data contains the criteria scores, the principle scores and the overall classification. The principle scores Health, Feeding, Housing and Behaviour are analysed separately. However, like the French data, some of the descriptive analyses are equivalent and will only be presented with the first principle.

11.1.5.1. Health principle

The health principle score was a continuous variable. The mean was 32.8 (SD = 7.2, min = 20.8, max =53.7). The frequency distribution of the health principle score is shown in Figure 53: .

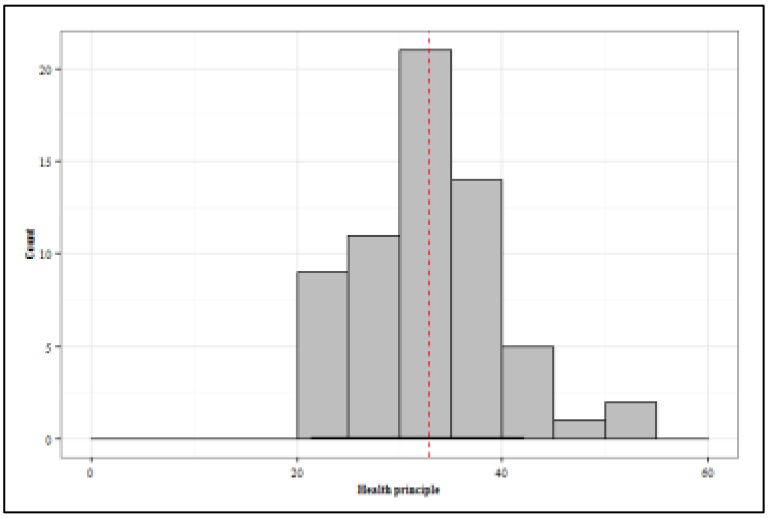


Figure 53: Frequency distribution of WQ® health principle score in the Belgian WQ® data. N = 63. Dashed line indicates mean principle score.

All three ABMs were identified in the data. The three ABMs mortality, SCC and lameness were continuous variables. Using the threshold from objective 2 for the dichotomisation of the ABMs was not biologically meningful. Therefore, thresholds were defined by the median of each ABM: mortality (0 = <3%, 1= >=3%), SCC (0 = <20 % of cows with SCC > 400,000, 1 = >=20 % of cows with SCC > 400,000) and lameness (0 = <22%, 1= >=22%). The following factors of variation were included in the descriptive analysis: pasture (binary, Yes/No), breed (binary, Milking/ Double purpose), housing system (binary, Loose house/ Tiestall) and herd size (number of animals, continuous). See Appendix M. for detailed description of the variables. The remaining of the selected factors of variation was not identified.

Boxplots of the health principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 54: and Figure 55: respectively.

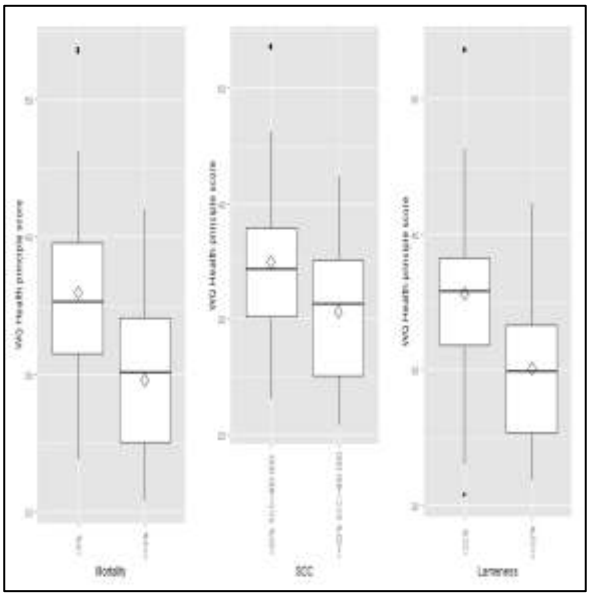
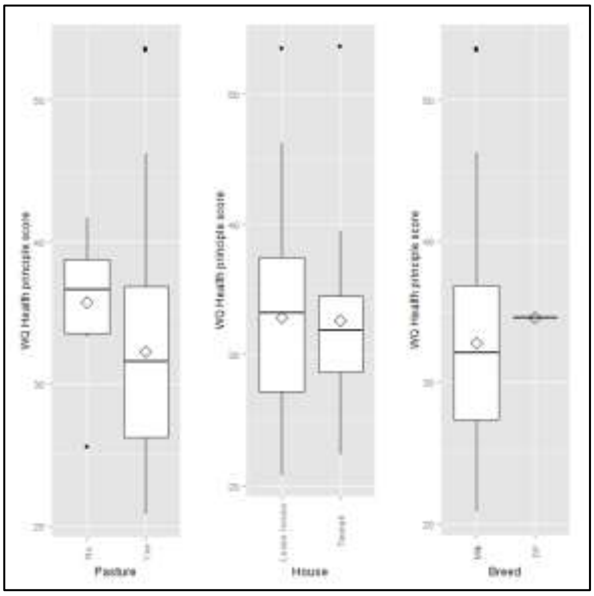


Figure 54: Boxplots of the WQ® health principle score by levels of the ABMs. Belgian WQ® data, N=63.



DP: Double purpose

Figure 55: Boxplots of the WQ® Health principle score by levels of categorical factors of variation. Belgian WQ® data, N=63.

The mean herd size was 61 cows (SD = 29, min = 25, max = 150). A scatterplot of the WQ® Health principle score versus the herd size is shown in Figure 56: .

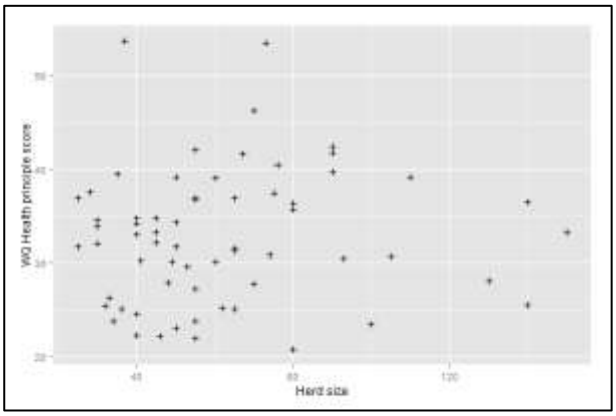


Figure 56: Scatterplot of WQ® health principle score versus herd size. Belgian WQ® data, N=63

The relationships between the ABMs and the factors of variation were illustrated graphically by boxplots (Figure 57:) and barplots (Figure 58:)

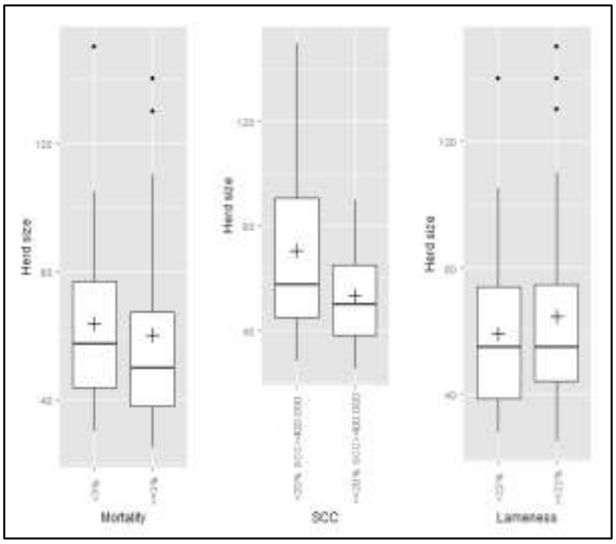
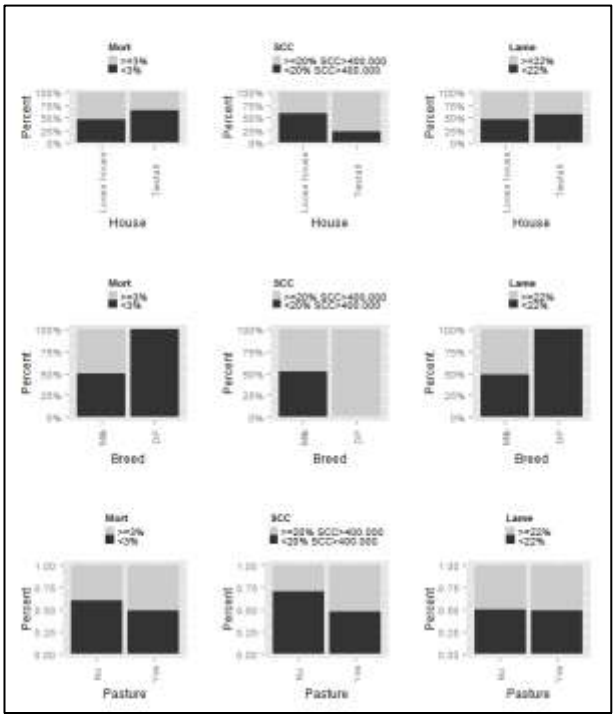


Figure 57: Boxplots of herd size by levels of the ABMs. Belgian WQ® data, N = 63



SCC: somatic cell count, Lame: lameness, Mort: mortality, DP: Double purpose, House: housing system

Figure 58: Barplots of the categorical factors of variations by the levels of the ABMs. Belgian WQ® data, N=63.

For the statistical analyses, the health principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 43: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 44: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Belgian WQ® Health principle score, N =63

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<3%	10	22	12	20	6	26
	>=3%	21	10	4	27	1	30
SCC	<20% SCC>400,000	12	20	10	22	6	26
	>=20% SCC>400,000	19	12	6	25	1	30
Lameness	<22%	9	22	10	21	6	25
	>=22%	22	10	6	26	1	31
Housing system	Loose house	23	26	13	36	6	43
	Tiestall	8	6	3	11	1	13
Breed	Milking	31	31	16	46	7	55
	Double purpose	0	1	0	1	0	1
Pasture	No	1	9	4	6	1	9
	Yes	30	23	12	41	6	47
Herd size	Mean	60.7	63.2	64.4	61	68.9	61.2

11.1.5.2. Housing principle

The housing principle score was a continuous variable. The mean was 46.9 (SD = 15.1, min =18.0, max =93.2). The frequency distribution of the housing principle score is shown in Figure 59: .

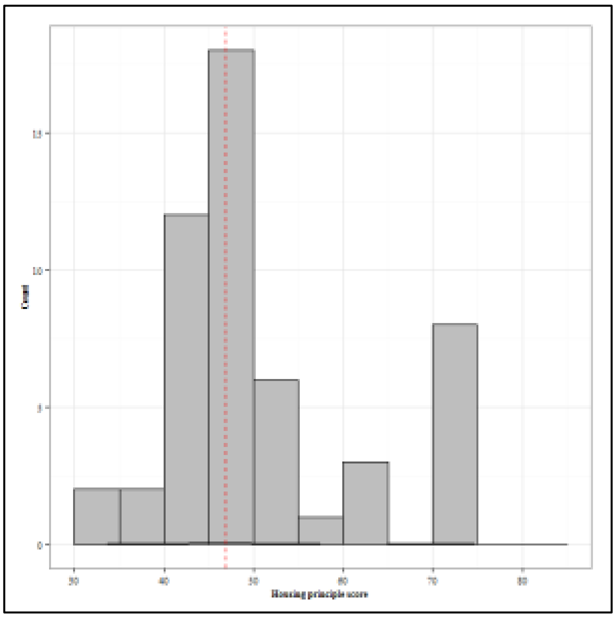


Figure 59: Frequency distribution of WQ® housing principle score in the Belgian WQ® data. N = 63. Dashed line indicates mean principle score.

The description of the ABMs and the factors of variation in the analysis of the Belgian WQ® Housing principle score are equivalent to description in the previous section regarding the Belgian WQ® Health principle score. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them).

Boxplots of the WQ® Housing principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 60: and Figure 61: respectively. The relationship between the WQ® Housing principle score and the herd size is illustrated in the scatterplot in Figure 62: .

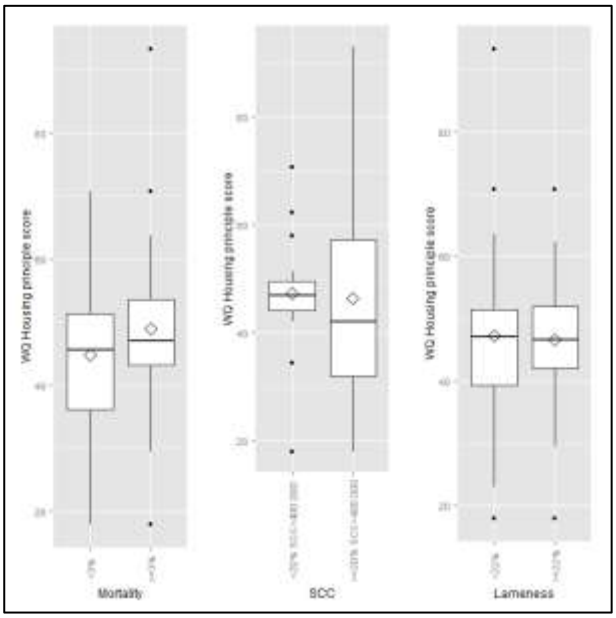
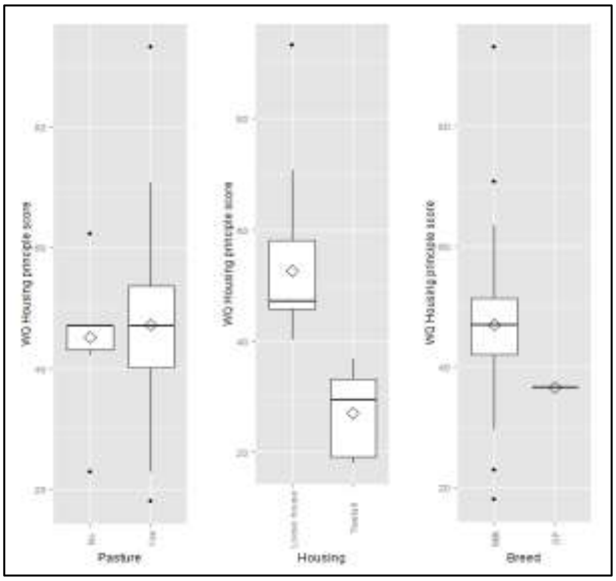


Figure 60: Boxplots of the WQ® housing principle score by the levels of the ABMs. Belgian WQ® data, N = 63.



DP: Double purpose

Figure 61: Boxplots of the WQ® housing principle score by the levels of the categorical factors of variation. Belgian WQ® data, N = 63.

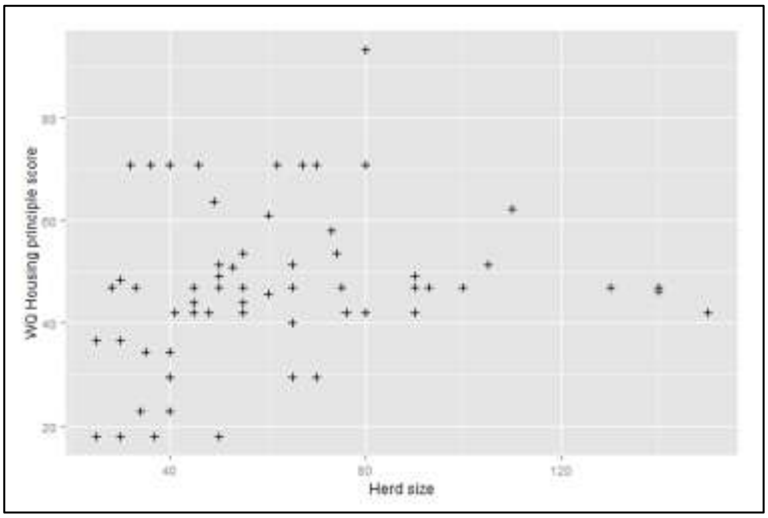


Figure 62: Scatterplot of WQ® housing principle score versus herd size. Belgian WQ® data, N =63.

For the statistical analyses, the housing principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 45: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 45: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Belgian WQ® Housing principle score, N =63

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<3%	16	16	6	26	0	32
	>=3%	12	19	9	22	1	30
SCC	<20% SCC>400,000	12	20	5	27	0	32
	>=20% SCC>400,000	16	15	10	21	1	30
Lameness	<22%	12	19	7	24	1	30
	>=22%	16	16	8	24	0	32
Housing system	Loose house	14	35	15	34	1	48
	Tiestall	14	0	0	14	0	14
Breed	Milking	27	35	15	47	1	61
	Double purpose	1	0	0	1	0	1
Pasture	No	4	6	1	9	0	10
	Yes	24	29	14	39	1	52
Herd size	Mean	55.9	66.7	62.3	61.8	80	61.6

11.1.5.3. Behaviour principle

The behaviour principle score was a continuous variable. The mean was 46.7 (SD = 14.0, min = 18.9, max =72.8, N = 63). The frequency distribution of the behaviour principle score is shown in Figure 63: .

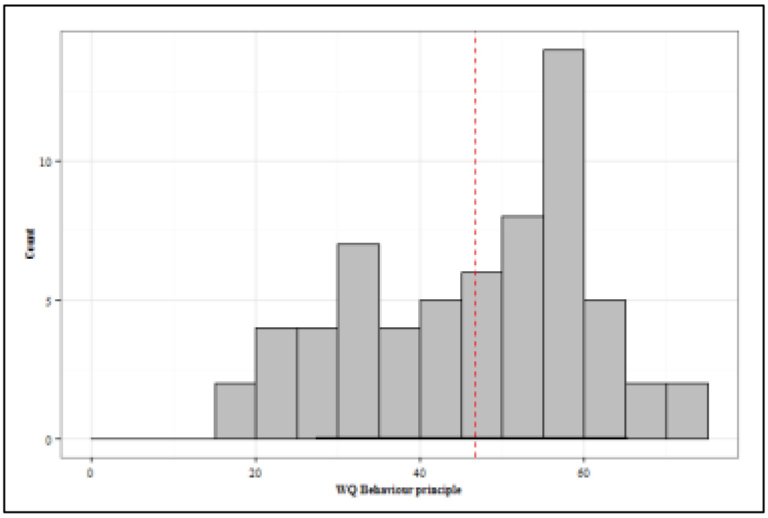


Figure 63: Frequency distribution of WQ® Behaviour principle score in the Belgian WQ® data. N = 63. Dashed line indicates mean principle score.

The description of the ABMs and the factors of variation in the analysis of the Belgian WQ® Behaviour principle score are equivalent to description in the previous section regarding the Belgian WQ® Health principle score. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them).

Boxplots of the WQ® Behaviour principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 64: and Figure 65: respectively. The relationship between the WQ® Behaviour principle score and the herd size is illustrated in the scatterplot in Figure 66: .

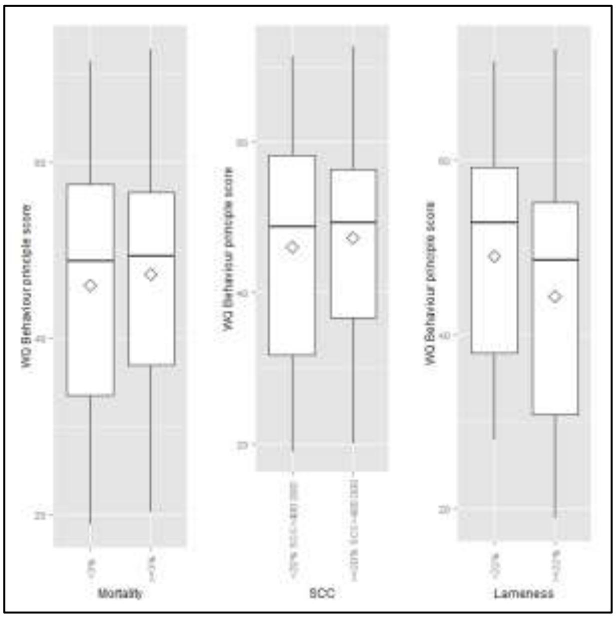
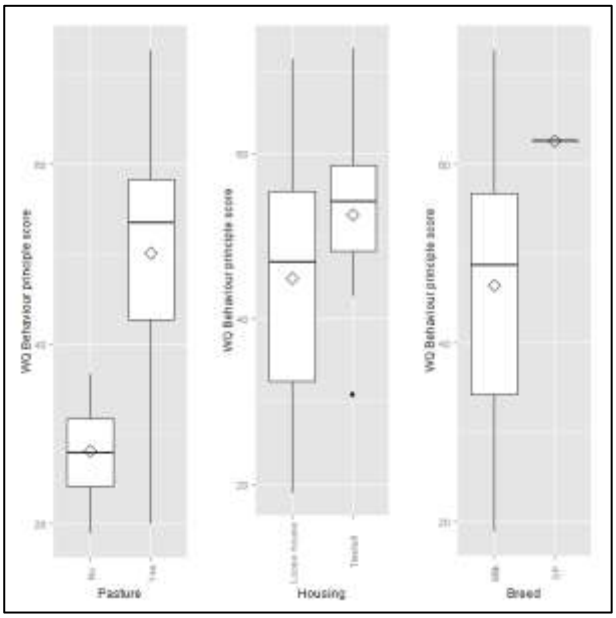


Figure 64: Boxplots of the WQ® behaviour principle score by the levels of the ABMs. Belgian WQ® data, N = 63.



DP: Double purpose

Figure 65: Boxplots of the WQ® behaviour principle score by the levels of the categorical factors of variation. Belgian WQ® data, N = 63.

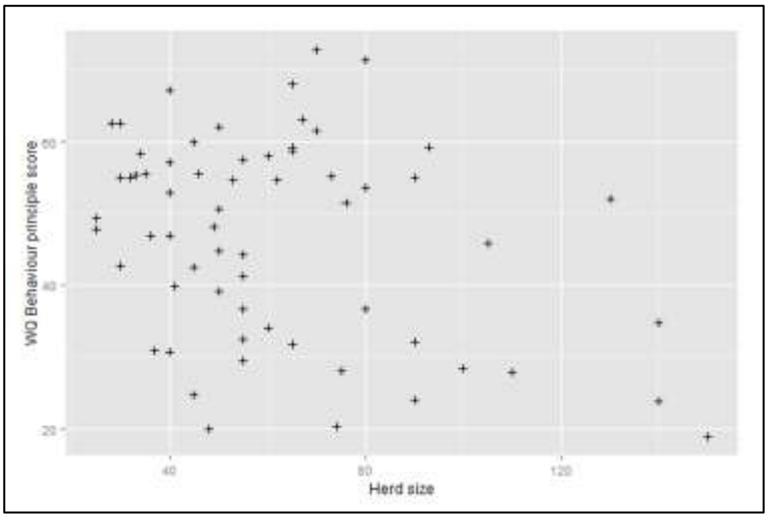


Figure 66: Scatterplot of WQ® behaviour principle score versus herd size. Belgian WQ® data, N =63.

For the statistical analyses, the behaviour principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 46: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 46: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Belgian WQ® Behaviour principle score, N =63

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<3%	16	16	8	24	3	29
	>=3%	15	16	8	23	4	27
SCC	<20% SCC>400,000	16	16	9	23	3	29
	>=20% SCC>400,000	15	16	7	24	4	27
Lameness	<22%	15	16	11	20	4	27
	>=22%	16	16	5	27	3	29
Housing system	Loose house	27	22	11	38	4	45
	Tiestall	4	10	5	9	3	11
Breed	Milking	31	31	15	47	6	56
	Double purpose	0	1	1	0	1	0
Pasture	No	10	0	0	10	0	10
	Yes	21	32	16	37	7	46
Herd size	Mean	67.4	56.6	57.3	63.5	54.3	62.9

11.1.5.4. Feeding principle

The feeding principle score was a continuous variable. The mean was 40.5 (SD = 27.8, min = 5.2, max =100, N = 63). The frequency distribution of the housing principle score is shown in Figure 67: .

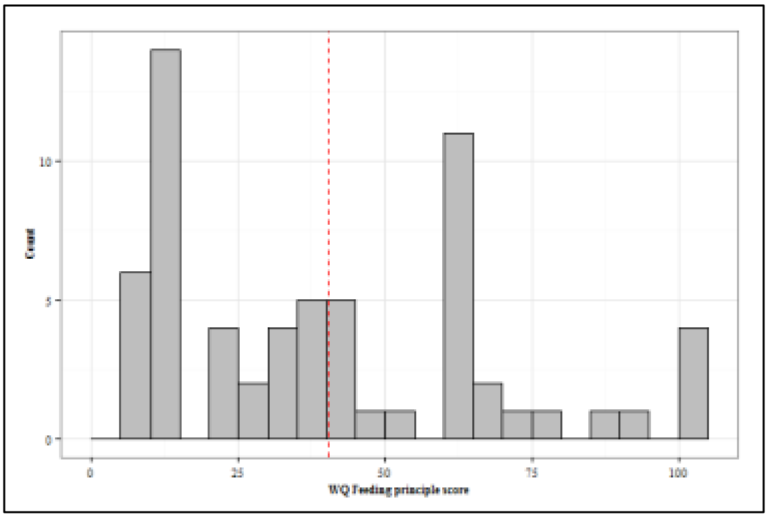


Figure 67: Frequency distribution of WQ® Feeding principle score in the Belgian WQ® data. N = 63. Dashed line indicates mean principle score.

The description of the ABMs and the factors of variation in the analysis of the Belgian WQ® Feeding principle score are equivalent to description in the previous section regarding the Belgian WQ® Health principle score. The reader is therefore referred to the previous section for the description of the ABMs and the factors of variation (the mean and the distribution of the ABMs and the factors of variation and the relationship between them).

Boxplots of the WQ® Feeding principle score by levels of the ABMs and the categorical factors of variation are shown in Figure 68: and Figure 69: respectively. The relationship between the WQ® Feeding principle score and the herd size is illustrated in the scatterplot in Figure 70: .

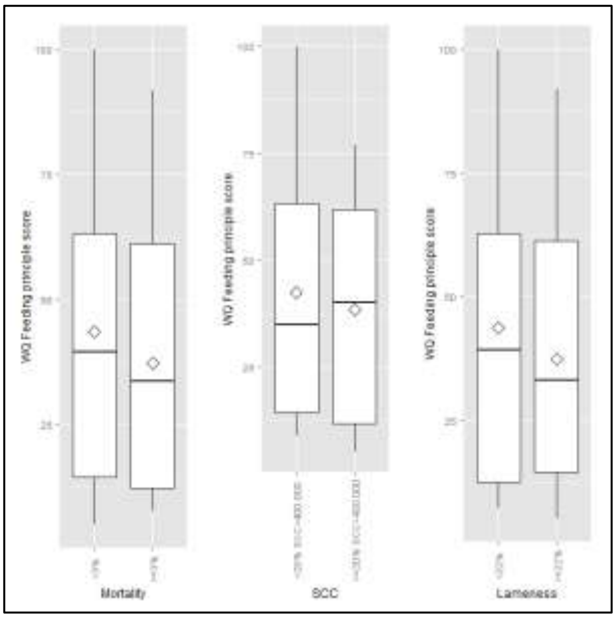
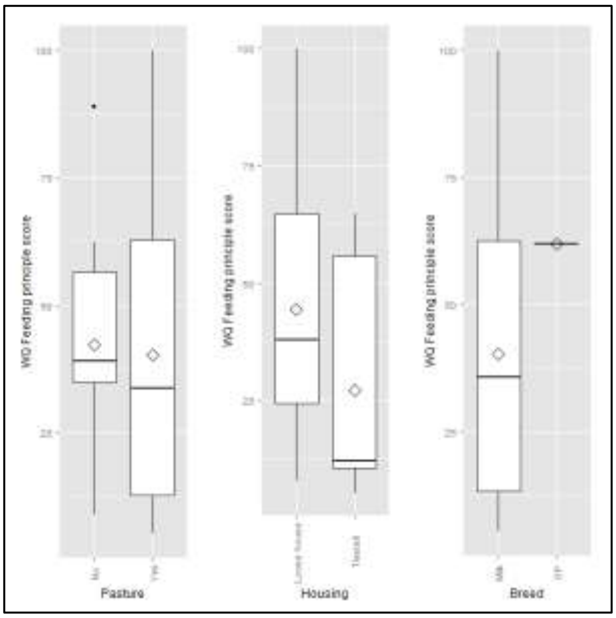


Figure 68: Boxplots of the WQ® feeding principle score by the levels of the ABMs. Belgian WQ® data, N = 63.



DP: Double purpose

Figure 69: Boxplots of the WQ® Feeding principle score by the levels of the categorical factors of variation. Belgian WQ® data, N = 63.

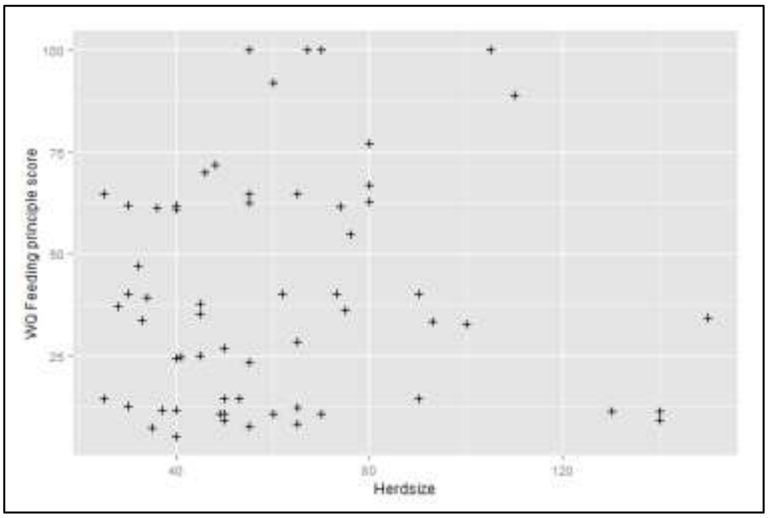


Figure 70: Scatterplot of WQ® feeding principle score versus herd size. Belgian WQ® data, N =63.

For the statistical analyses, the feeding principle score was dichotomised at three different levels: the median, the 25th percentile and the 10th percentile. In Table 47: the number of observations in each category of the ABMs and the categorical factors of variation and the mean herd size and milk yield are given for each level of the dichotomised outcomes.

Table 47: Number of observations (or the mean) in each level of the ABMs and the factors of variation by the levels of the dichotomised outcomes. Belgian WQ® Feeding principle score, N =63

		Outcome median		Outcome P25		Outcome P10	
Parameter	**Level**	**Poorer**	**Better**	**Poorer**	**Better**	**Poorer**	**Better**
Mortality	<3%	14	18	7	25	3	29
	>=3%	17	14	9	22	4	27
SCC	<20% SCC>400,000	17	15	7	25	1	31
	>=20% SCC>400,000	14	17	9	22	6	25
Lameness	<22%	13	18	9	22	3	28
	>=22%	18	14	7	25	4	28
Housing	Loose house	22	27	8	41	3	46
	Tiestall	9	5	8	6	4	10
Breed	Milking	31	31	16	46	7	55
	Double purpose	0	1	0	1	0	1
Pasture	No	3	7	2	8	1	9
	Yes	28	25	14	39	6	47
Herd size	Mean	64.2	59.7	66	60.6	65	61.6

11.1.6. Descriptive analyses for objective 4

Table 48: presents descriptive statistics of the data submitted to DCF in Objective 4. Data is presented at the same level as it was submitted to DCF: Data from Italy (IZSAM+IZSLER databases) and Belgium at the herd level (for the Belgian data in some cases more than one recording/line per herd) and data from Denmark at the herd level for annual mortality rate and at the cow level (individual milk recordings, more than one recording possible for each individual cow) for SCC. The Italian data included information from 85,504 herds, the Danish mortality data included information from 3,652 herds and the Danish SCC data included information from 3,896,014 individual milk recordings. The Belgian data included 24,563 lines (each herd contributed with one or more lines). The distribution of observations is presented in Figures Figure 71: Figure 72: Figure 73: Figure 74: Figure 75: Figure 76: Figure 77: Figure 78: Figure 79: . In order to improve the readability of the figures, in some cases a few extreme observations have not been included in the histograms.

Table 48: Descriptive statistics of the data submitted to DCF in Objective 4

Country	Herd size, cows	SCC	Percent cows with SCC>400,000	Annual mortality rate, %
Italy	Mean: 23.1 Min: 1 Q1: 2 Median: 7 Q3: 22 Max: 2110	Mean: 274,229 Min: 1000 Q1: 186,000 Median: 248,000 Q3: 326,000 Max: 11,013,000	No information available	Mean: 2.25 Min: 0 Q1: 0 Median: 0 Q3: 0 Max: 100
Belgium	Mean: 53.1 Min: 3 Q1: 33 Median: 47 Q3: 65 Max: 399	Mean: 260,715 Min: 29,000 Q1: 160,000 Median: 228,000 Q3: 324,000 Max: 3,216,000	Mean: 22.6 Min: 0 Q1: 12.3 Median: 21.7 Q3: 31.6 Max: 100	No information available
Denmark	Mean: 156 Min: 1 Q1: 77 Median: 136 Q3: 200 Max: 1317	Mean: 245,495 Min: 0 Q1: 37,000 Median: 80,000 Q3: 197,000 Max: 9,999,000	No information available	Mean: 4.4 Min: 0 Q1: 1.8 Median: 3.8 Q3: 6.1 Max: 69

Min: minimum; Q1: 1st quartile; Q3: 3rd quartile; max: maximum

It can be seen that the mean herd size among herds included in these datasets is larger in Denmark compared to Belgium and especially Italy. Somatic cell counts are at the same level for all three countries. No information regarding annual mortality rate was available from Belgium. Comparing Denmark and Italy it can be seen that the mean annual mortality rate was higher in the Danish dataset. This means that a direct comparison between the annual mortality rate in an Italian and a Danish dairy herd is not straightforward and therefore region/country may need to be taken into account when evaluating the potential animal welfare consequences of a certain level of cow mortality.

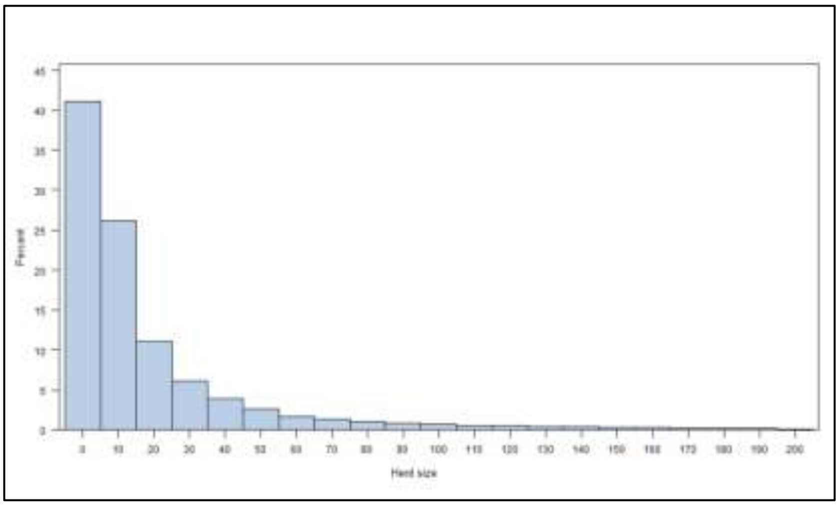


Figure 71: Distribution of herd sizes in 85,504 Italian dairy herds where data was uploaded to DCF during Objective 4 of this project

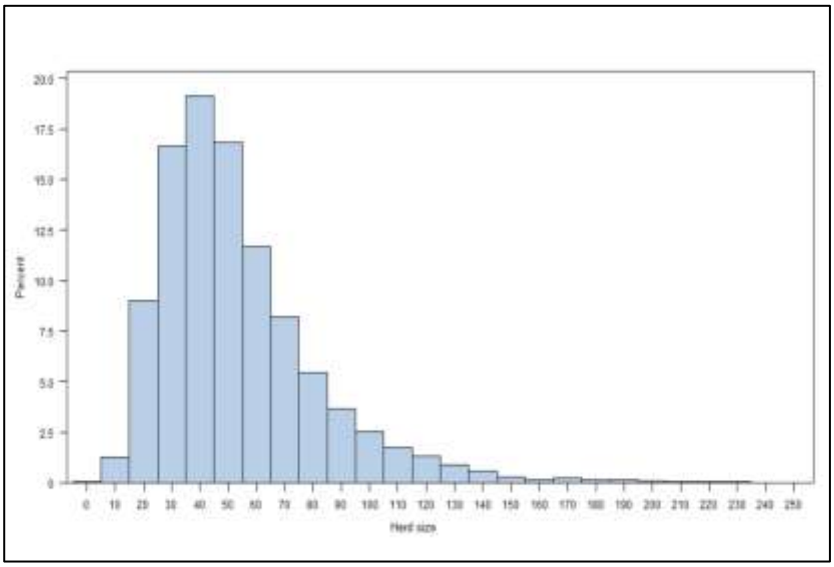


Figure 72: Distribution of herd sizes in 24,504 observations from Belgian dairy herds where data was uploaded to DCF during Objective 4 of this project

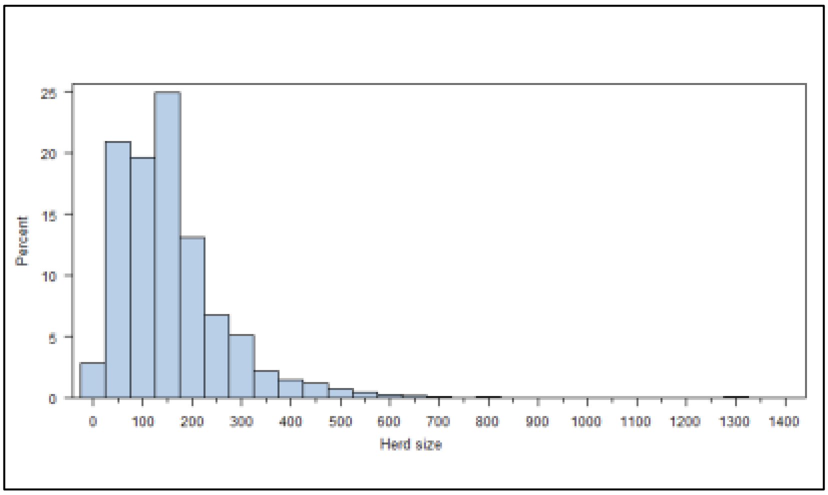


Figure 73: Distribution of herd sizes in 3,652 Danish dairy herds where data was uploaded to DCF during Objective 4 of this project

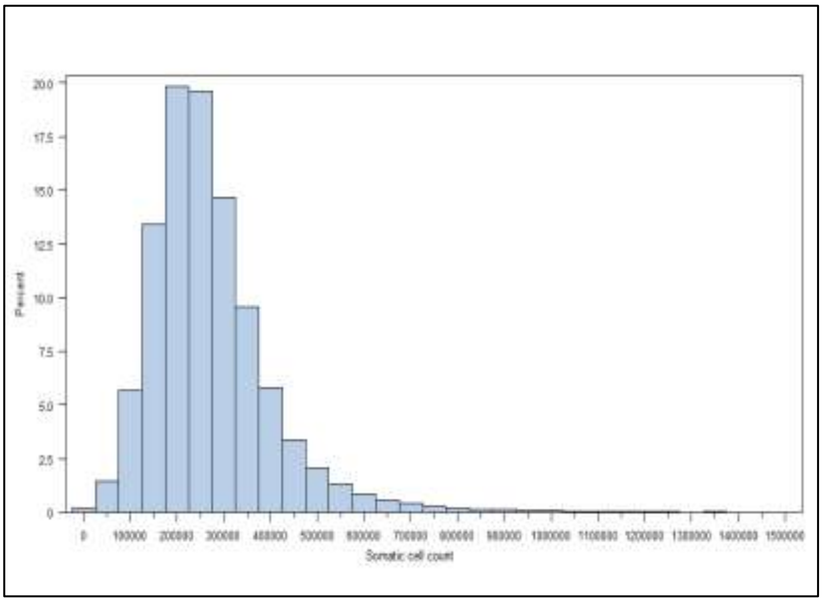


Figure 74: Distribution of somatic cell counts from 85,504 Italian dairy herds where data was uploaded to DCF during Objective 4 of this project

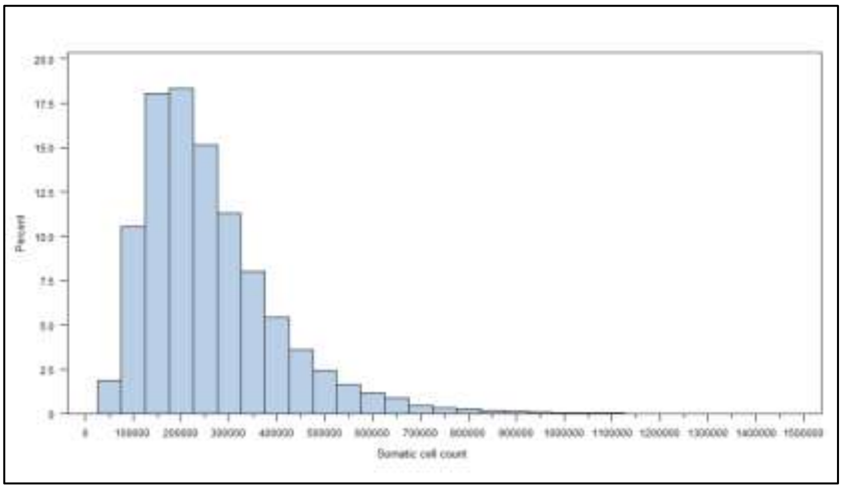


Figure 75: Distribution of somatic cell counts from 24,563 observations from Belgian dairy herds where data was uploaded to DCF during Objective 4 of this project

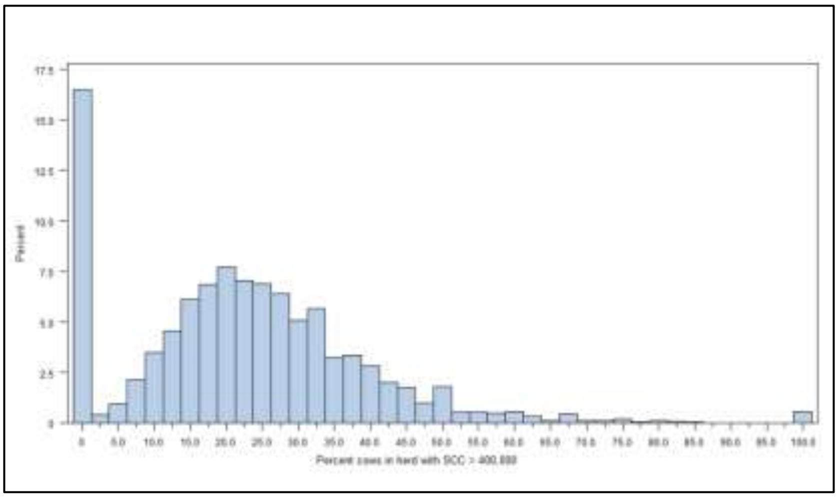


Figure 76: Distribution of the percentage of cows in each herd having a somatic cell count > 400,000 from 24,563 observations from Belgian dairy herds where data was uploaded to DCF during Objective 4 of this project

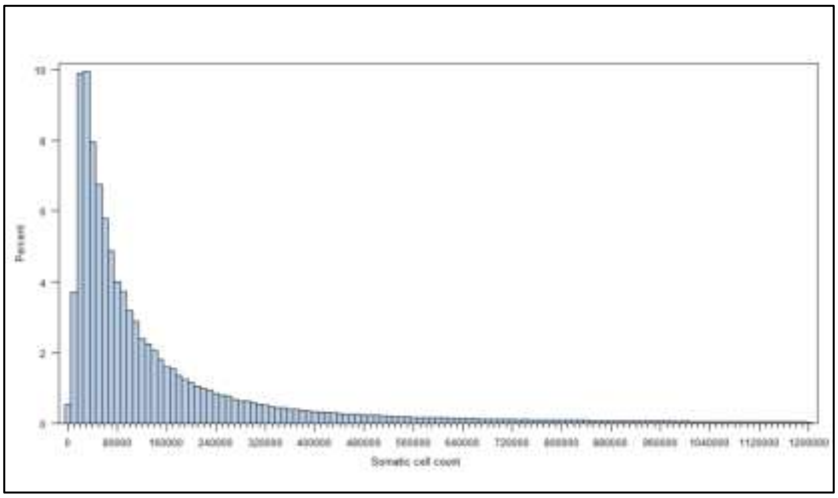


Figure 77: Distribution of somatic cell counts from 3,896,014 individual milk recordings from Danish dairy cows where data was uploaded to DCF during Objective 4 of this project

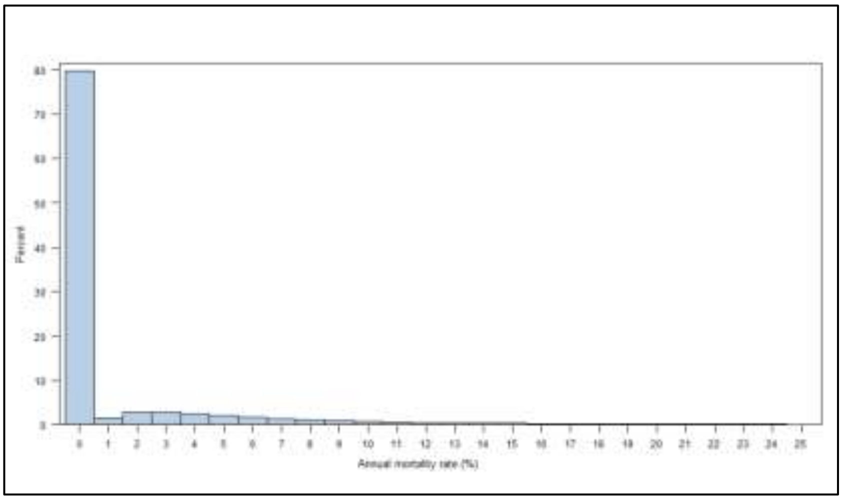


Figure 78: Distribution of annual mortality rates in 85,504 Italian dairy herds where data was uploaded to DCF during Objective 4 of this project. Mortality rate is presented in percent.

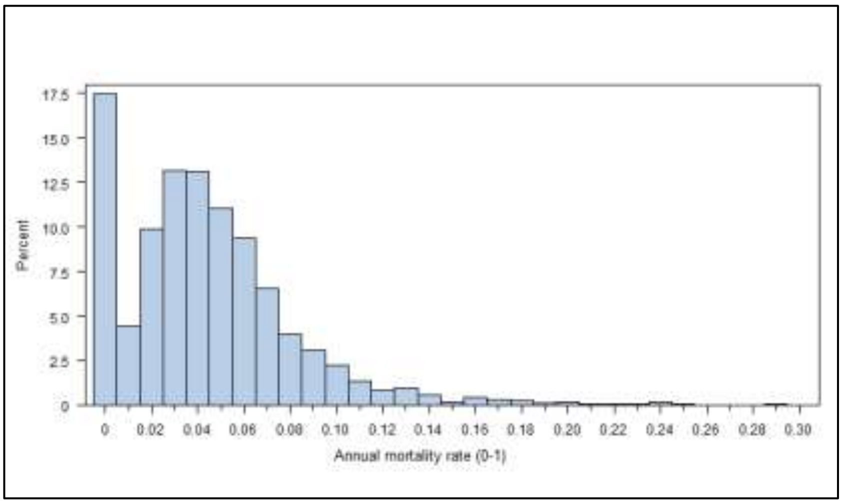


Figure 79: Distribution of annual mortality rates in 3,652 Danish dairy herds where data was uploaded to DCF during Objective 4 of this project. Mortality rate is presented on a scale from 0 to 1.

11.2. Statistical analyses

11.2.1. IZSLER/CReNBA data

In Table 49: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. Similarly, the formulas and the F-test p-values for the models with the outcome dichotomised using the P25 and the P10 are shown in Table 50: and Table 51: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 80: , Figure 81: and Figure 82: for a detailed presentation of the structure and the nodes in these models.

Table 49: Model formulas for models with outcome defined by the median. IZSLER/CReNBA data, N = 608.

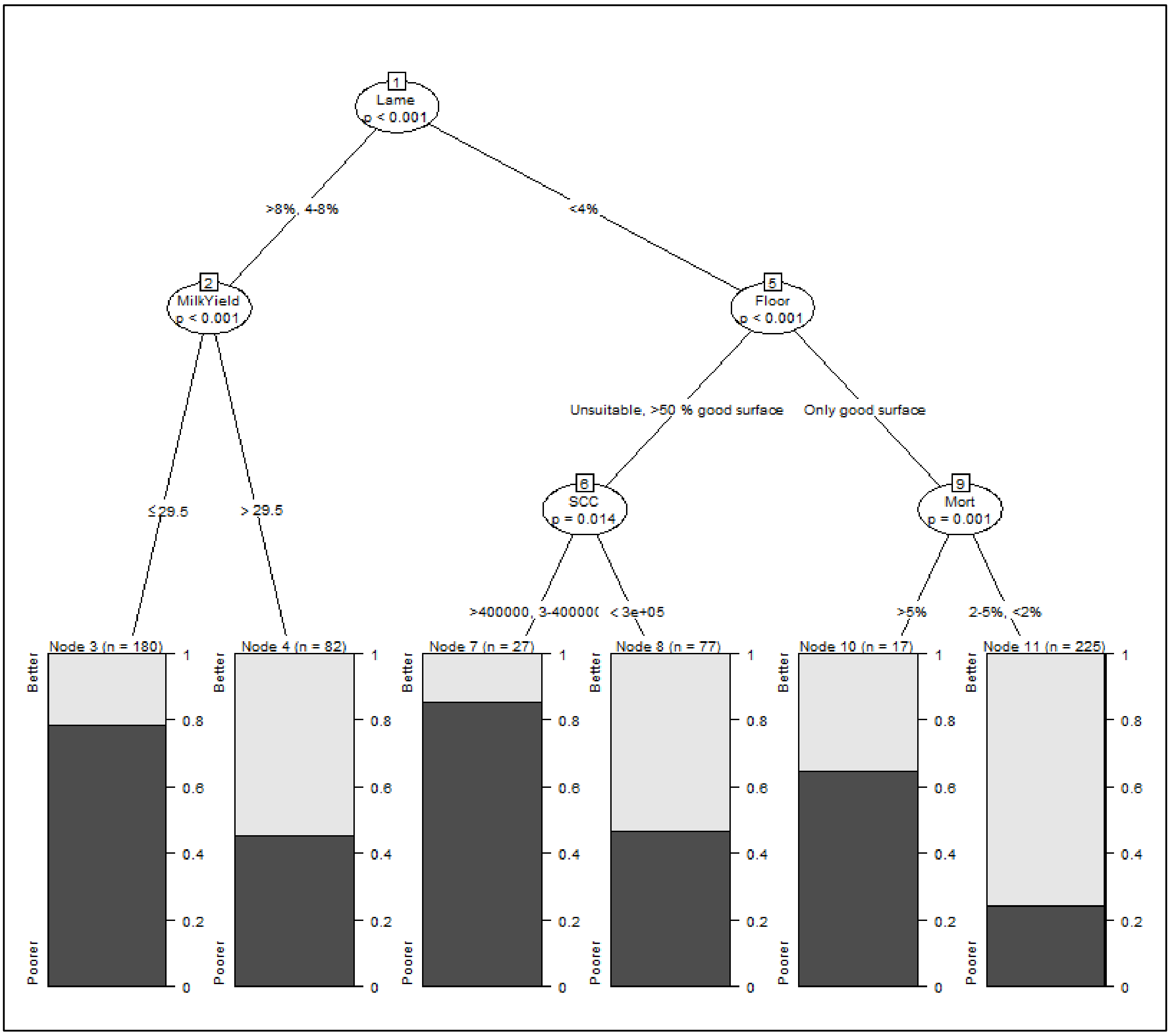
Model	Formula	F-test, P value
Model 1	Mortality	<0.0001
Model 2	SCC	<0.0001
Model 3	Lameness	<0.0001
Model 4	Mortality + SCC	<0.0001
Model 5	Mortality + Lameness	<0.0001
Model 6	SCC + Lameness	<0.0001
Model 7	Mortality + SCC + Lameness	<0.0001
Model 8	Mortality + SCC + Lameness + Herd size + Milk yield + Housing + Floor	<0.0001
Model 9	Unreliable due to rank deficiency	NA
Model 10	Lameness + Milk yield + Floor + SCC + Mortality	NA

Table 50: Model formulas for models with outcome defined by the 25th percentile. IZSLER/CReNBA data, N = 608.

Model	Formula	F-test, P value
Model 1	Mortality	<0.0001
Model 2	SCC	<0.0001
Model 3	Lameness	<0.0001
Model 4	Mortality + SCC	<0.0001
Model 5	Mortality + Lameness	<0.0001
Model 6	SCC + Lameness	<0.0001
Model 7	Mortality + SCC + Lameness	<0.0001
Model 8	Mortality + SCC + Lameness + Herd size + Milk yield + Housing + Floor	<0.0001
Model 9	Unreliable due to rank deficiency	NA
Model 10	Lameness + Milk yield + Floor + Herd size + Lameness + SCC + Mortality	NA

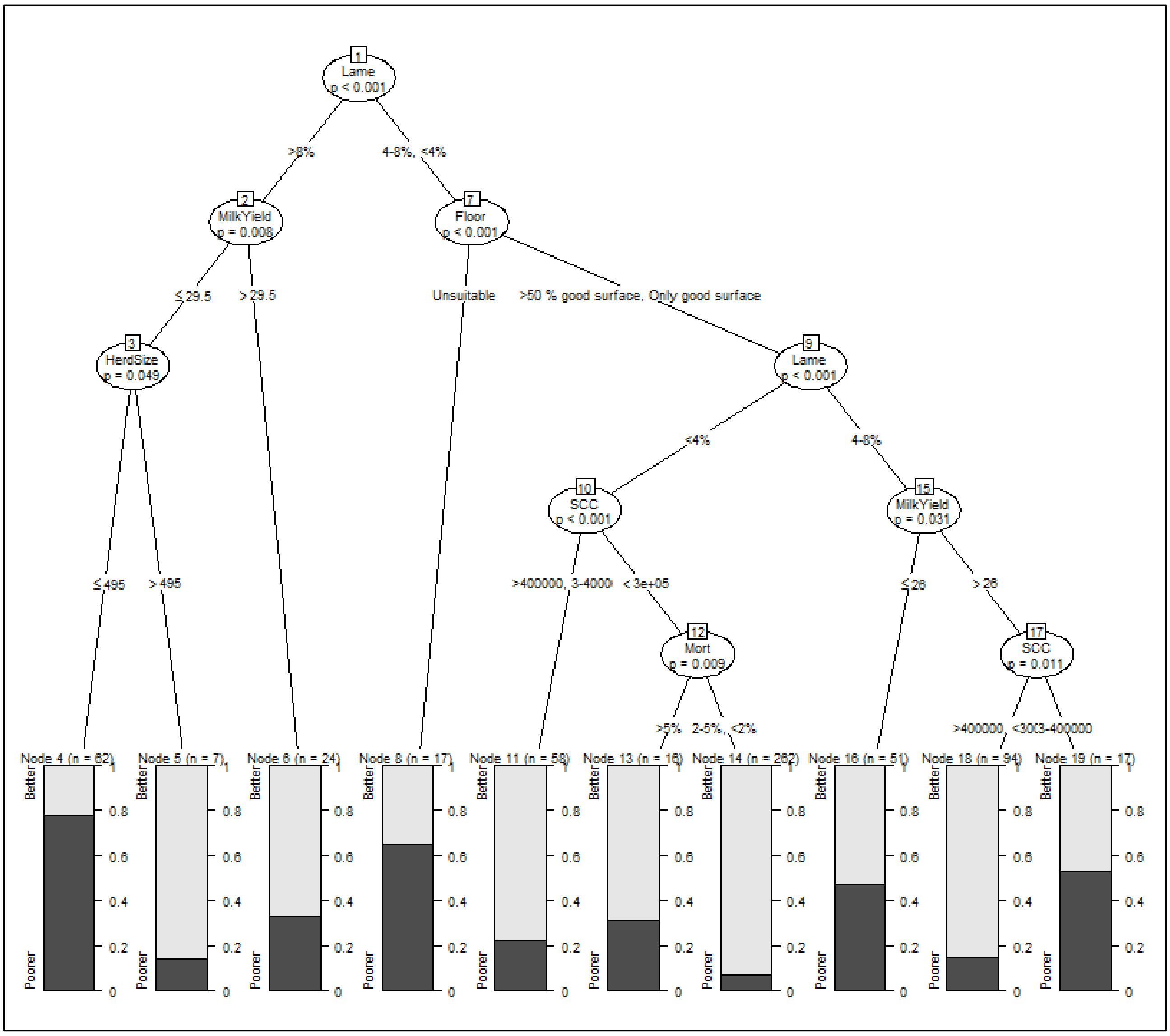
Table 51: Model formulas for models with outcome defined by the 10th percentile. IZSLER/CReNBA data, N = 608.

Model	Formula	F-test, P value
Model 1	Mortality	0.00172
Model 2	SCC	<0.0001
Model 3	Lameness	<0.0001
Model 4	Mortality + SCC	<0.0001
Model 5	Mortality + Lameness	<0.0001
Model 6	SCC + Lameness	<0.0001
Model 7	Mortality + SCC + Lameness	<0.0001
Model 8	Mortality + SCC + Lameness + Herd size + Milk yield + Housing + Floor	<0.0001
Model 9	Unreliable due to rank deficiency	NA
Model 10	Lameness + SCC + Milk yield + Floor	NA



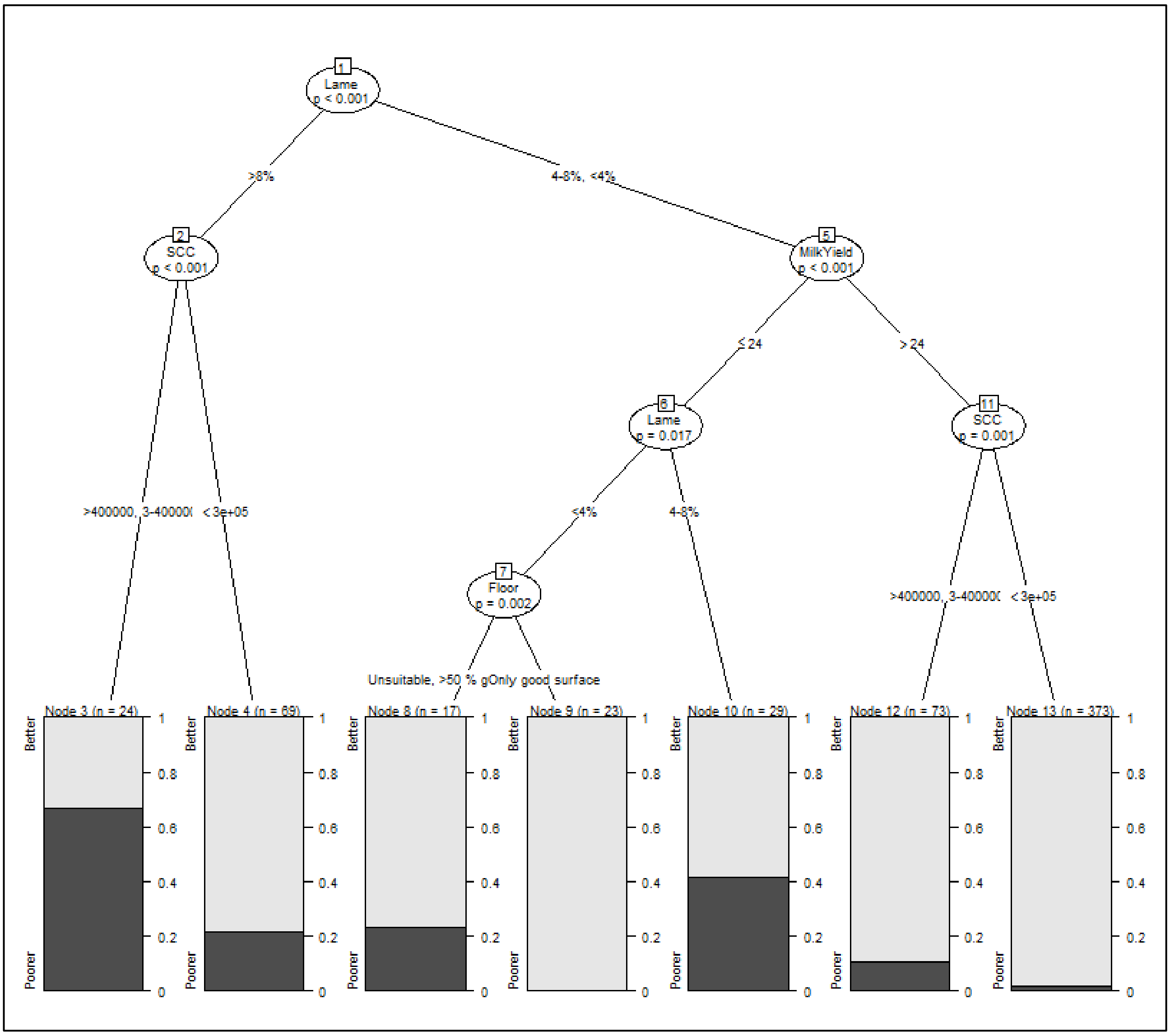
Lame: lameness, Mort: mortality

Figure 80: Conditional inference tree model with outcome defined by the median, IZSLER/CReNBA data, N=608.



Lame: lameness, Mort: mortality

Figure 81: Conditional inference tree model with outcome defined by the 25th percentile, IZSLER/CReNBA data, N=608.



Lame: lameness, Mort: mortality

Figure 82: Conditional inference tree model with outcome defined by the 10th percentile, IZSLER/CReNBA data, N=608.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 83: Figure 84: Figure 85: . Regardless of the outcome, the logistic regression models including two-way interactions were unstable due to rank deficiency and therefore, ROC curves are not presented. All of the ROC AUC were significantly larger than 0.5 and thus all models applied to the IZSLER/CReNBA data had a predictive value above chance.

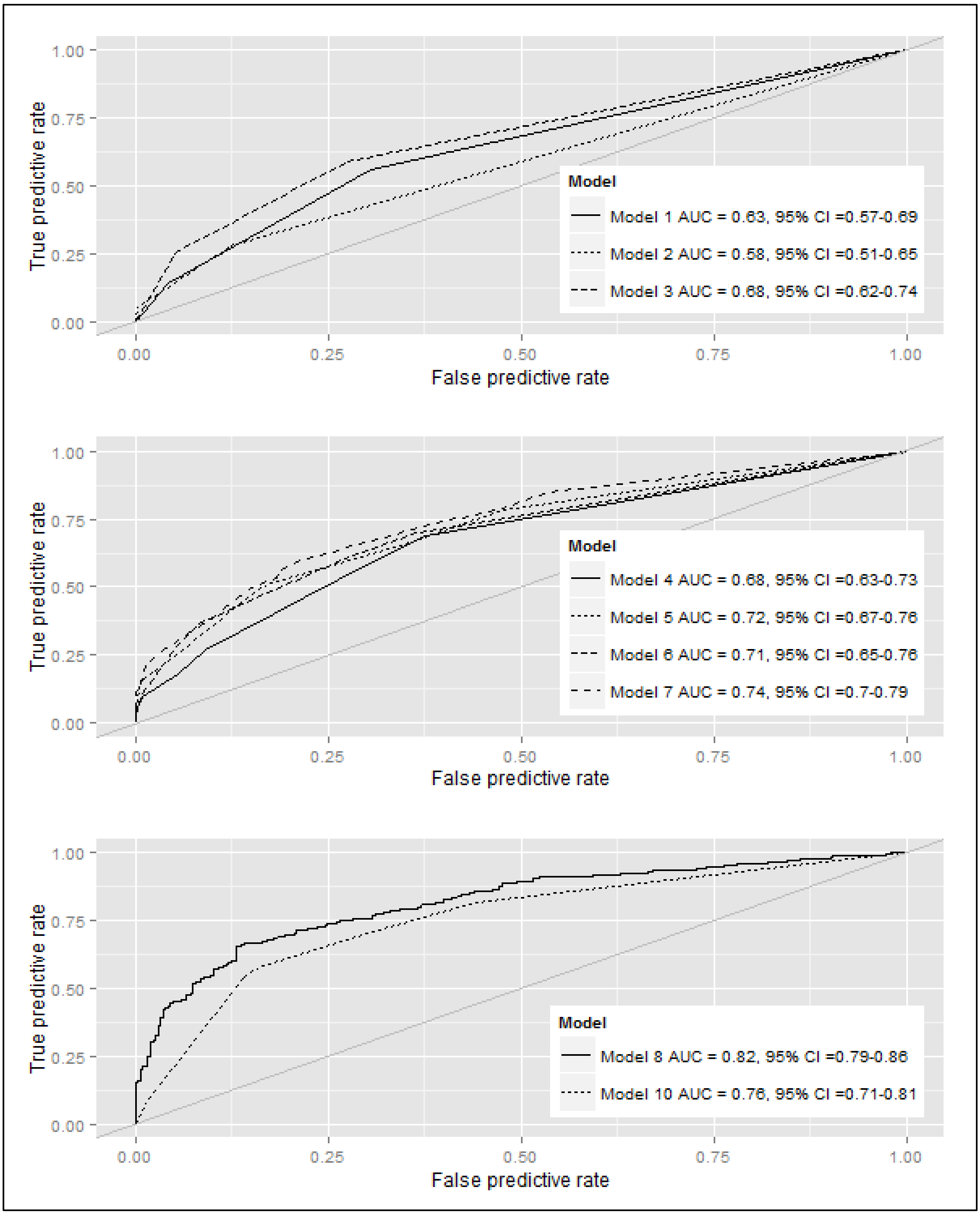


Figure 83: ROC curves and cross validated AUC (with confidence intervals) of models applied to the IZSLER/CReNBA data with outcome defined by the median. N=608.

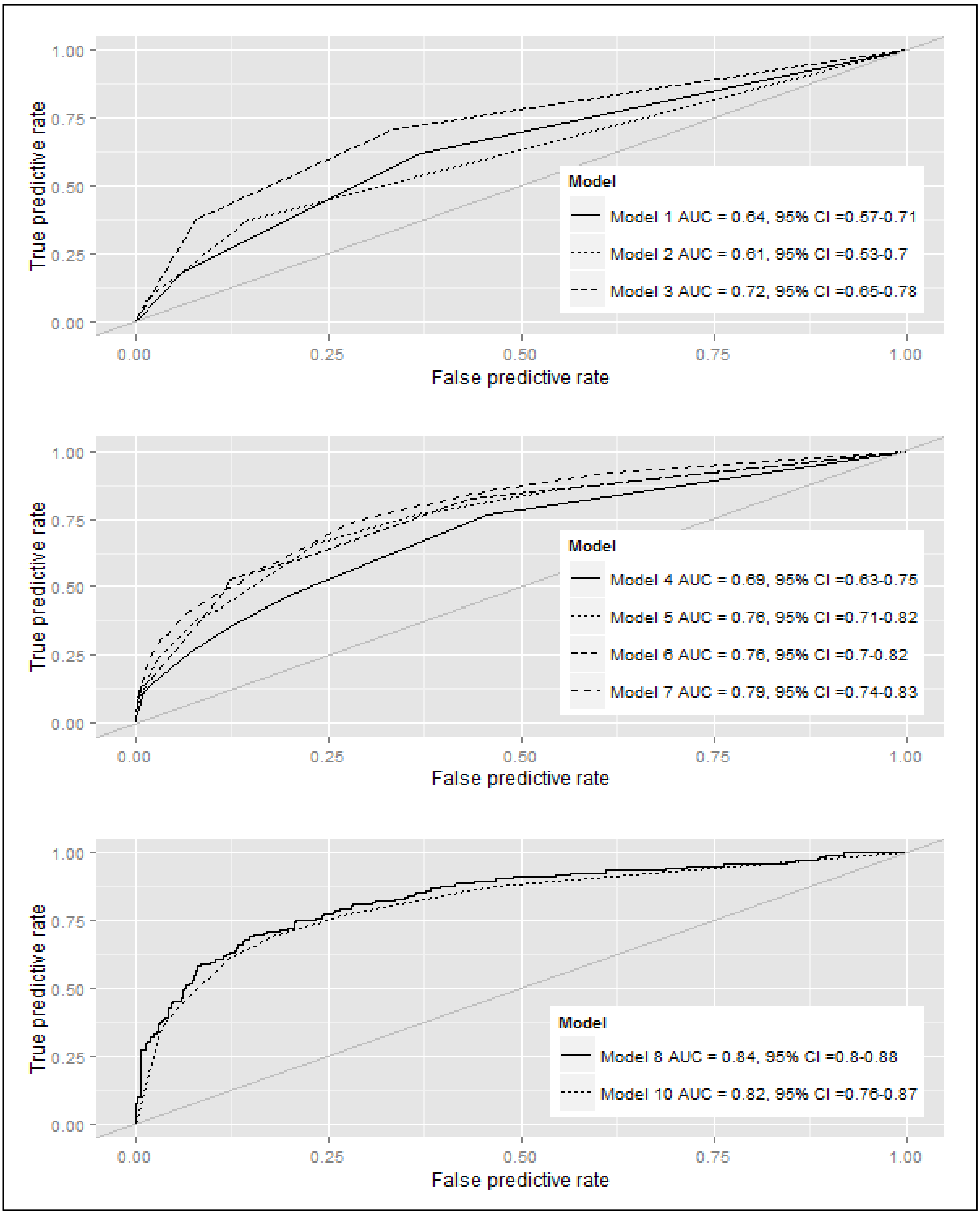


Figure 84: ROC curves and cross validated AUC (with confidence intervals) of models applied to the IZSLER/CReNBA data with outcome defined by the 25th percentile. N=608.

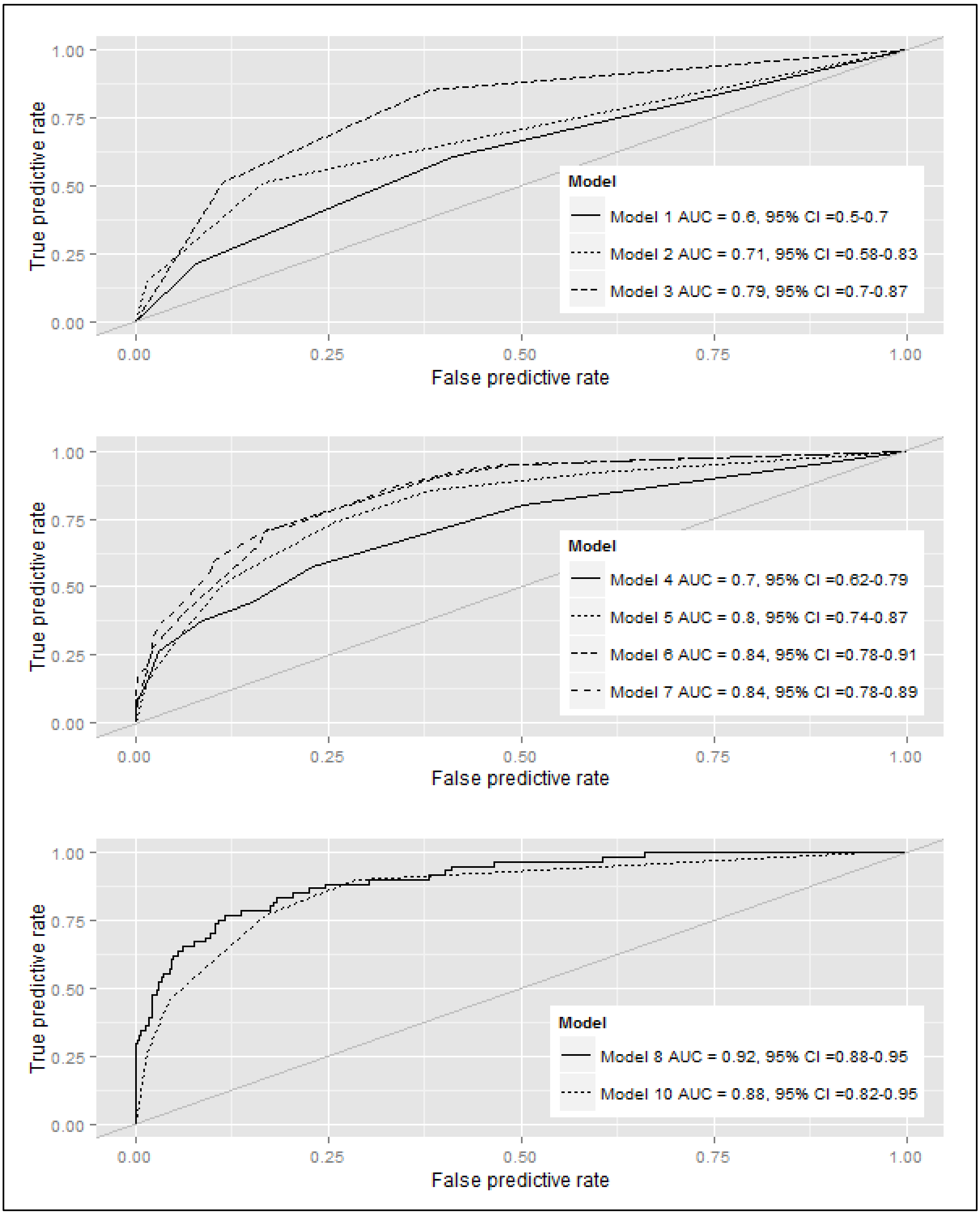


Figure 85: ROC curves and cross validated AUC (with confidence intervals) of models applied to the IZSLER/CReNBA data with outcome defined by the 10th percentile. N=608.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the

Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 52: .

Table 52: Comparing model results from ten different model applied to the IZSLER/CReNBA data using three different outcomes. N = 608. AUCs significantly larger than 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	816.1	0.2294	**0.6374**	668.6	0.1784	**0.6354**	402.7	0.0892	**0.6187**
2	SCC	832.1	0.2406	**0.5795**	667.8	0.1776	**0.6152**	375.8	0.0835	**0.7045**
3	Lameness	785.1	0.2220	**0.6761**	614.0	0.1599	**0.7292**	347.9	0.0798	**0.7912**
4	Mortality + SCC	801.1	0.2248	**0.6759**	649.1	0.1693	**0.6977**	377.9	0.0816	**0.7235**
5	Mortality + Lameness	766.7	0.2129	**0.7116**	605.9	0.1551	**0.7581**	356.0	0.0795	**0.7847**
6	SCC + Lameness	771.9	0.2141	**0.7125**	595.1	0.1520	**0.7564**	324.8	0.0732	**0.8490**
7	Mortality + SCC + Lameness	754.2	0.2055	**0.7431**	588.8	0.1478	**0.7920**	334.2	0.0729	**0.8461**
8	All factors, additive	710.6	0.1804	**0.8232**	571.8	0.1330	**0.8429**	324.2	0.0632	**0.9124**
9	All factors, incl. interactions	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	Conditional Inference Tree	NA	NA	**0.7576**	NA	NA	**0.8140**	NA	NA	**0.8743**

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve.

In the IZSLER/CReNBA data, all the applied models were highly significant (F-test, see Table 49: , Table 50: and Table 51: except from the interaction models, none of which resulted in reliable estimates due to rank deficiency. Low mortality, low SCC and low lameness prevalence were all associated with lower risk of classifying the herd in the poorer welfare group. Lameness seemed to be the ABM that alone performs best in identifying the herds with the poorest welfare at all outcome levels. However, the MSPE decreases and the AUC increase when ABMs are combined. In the more complex models, all three ABMs are significant predictors except from mortality in the conditional inference tree model with outcome based on the P10.

Regarding the factors of variation the following were identified in the data: herd size, milk yield, housing and floor. Generally, the inclusion of the factors of variation decreased the BIC and the MSPE and increased the AUC when compared to simpler model including only the ABMs. All of the additive models included at least two of these factors but milk yield was the only one consistently staying in all models. Thus, higher milk yield, larger herd size, house system with tethering and herd with unsuitable floors were associated with poorer welfare which confirmed the findings from the descriptive analysis. However, the inclusion of the two-way interactions resulted in rank deficiency and consequently unreliable estimates. The complex structure of the results of the conditional inference tree models indicated that interactions between variables could be identified in the data. Nevertheless, the predictive values of the conditional inference tree models were not better than the simpler additive logistic regression models.

11.2.2. Otten data

Results from statistical analyses of all three different welfare indexes in the Otten data are presented individually before a summation of results from the Otten data is given.

11.2.2.1 Animal based welfare measure

In Table 53: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. Similarly, the formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 54: and Table 55: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 86: and Figure 87: for a detailed presentation of the structure and the nodes in these models. When using the P90 in defining the outcome, no significant variables could be detected in the conditional inference tree.

Table 53: Model formulas for models with outcome defined by the median. Otten data, animal based welfare measure, N = 72.

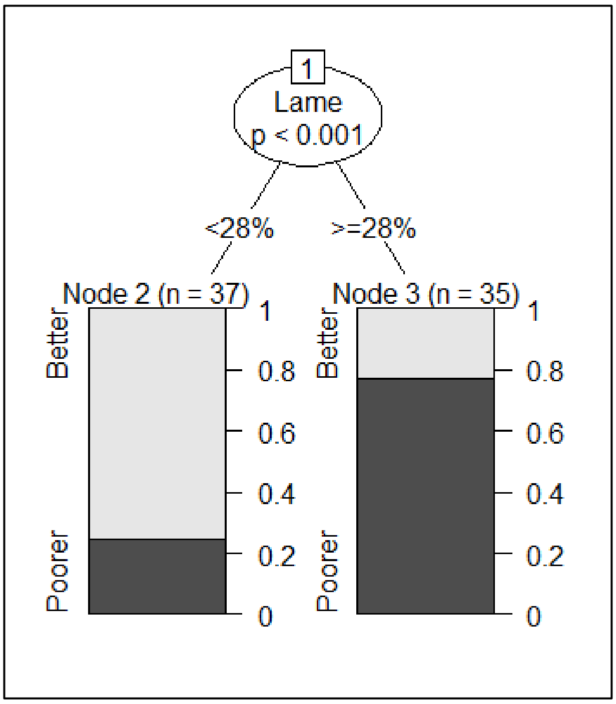
Model	Formula	F-test, P value
Model 1	Mortality	0.3562
Model 2	SCC	0.2309
Model 3	Lameness	<0.0001
Model 4	Mortality + SCC	0.1811
Model 5	Mortality + Lameness	<0.0001
Model 6	SCC + Lameness	<0.0001
Model 7	Mortality + SCC + Lameness	<0.0001
Model 8	Lameness	<0.0001
Model 9	Lameness	<0.0001
Model 10	Lameness	NA

Table 54: Model formulas for models with outcome defined by the 75th percentile. Otten data, animal based welfare measure, N = 72.

Model	Formula	F-test, P value
Model 1	Mortality	0.2322
Model 2	SCC	0.7276
Model 3	Lameness	0.0035
Model 4	Mortality + SCC	0.3593
Model 5	Mortality + Lameness	0.0095
Model 6	SCC + Lameness	0.0130
Model 7	Mortality + SCC + Lameness	0.0190
Model 8	Lameness	0.0035
Model 9	Lameness	0.0035
Model 10	Lameness	NA

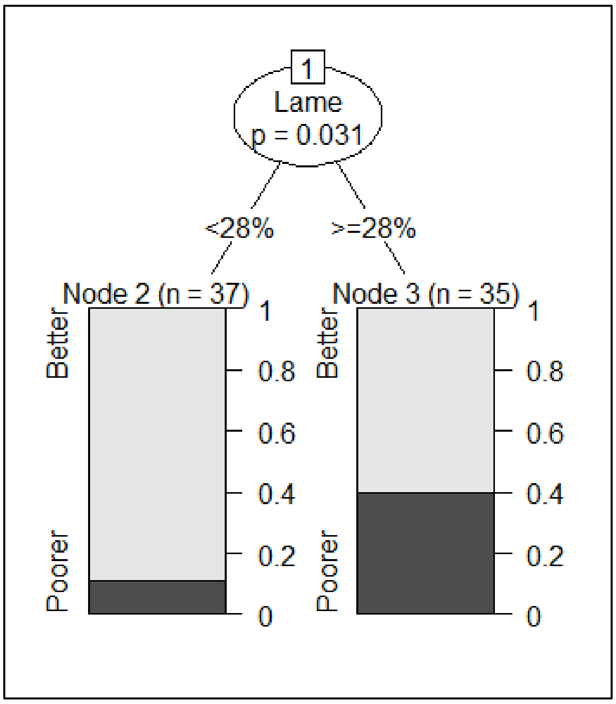
Table 55: Model formulas for models with outcome defined by the 90th percentile. Otten data, animal based welfare measure, N = 72.

Model	Formula	F-test, P value
Model 1	Mortality	0.1635
Model 2	SCC	0.5811
Model 3	Lameness	0.1068
Model 4	Mortality + SCC	0.2121
Model 5	Mortality + Lameness	0.1298
Model 6	SCC + Lameness	0.2300
Model 7	Mortality + SCC + Lameness	0.1498
Model 8	Unreliable due to rank deficiency	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	No inner nodes	NA



Lame: lameness

Figure 86: Conditional inference tree model with outcome defined by the median. Otten data, animal based welfare measure, N =72.



Lame: lameness

Figure 87: Conditional inference tree model with outcome defined by the 25th percentile. Otten data, animal based welfare measure, N =72.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 88: ,Figure 89: and Figure 90: . ROC curves are missing where models were unreliable due to rank deficiency (P90 model 8 and 9). When using the median or the P75 for the definition of the outcome, models 8-10 contained the same variables and the ROC curves were therefore identical. As there were no significant nodes in the tree model at the P90 outcome, there was no ROC curve for this outcome.

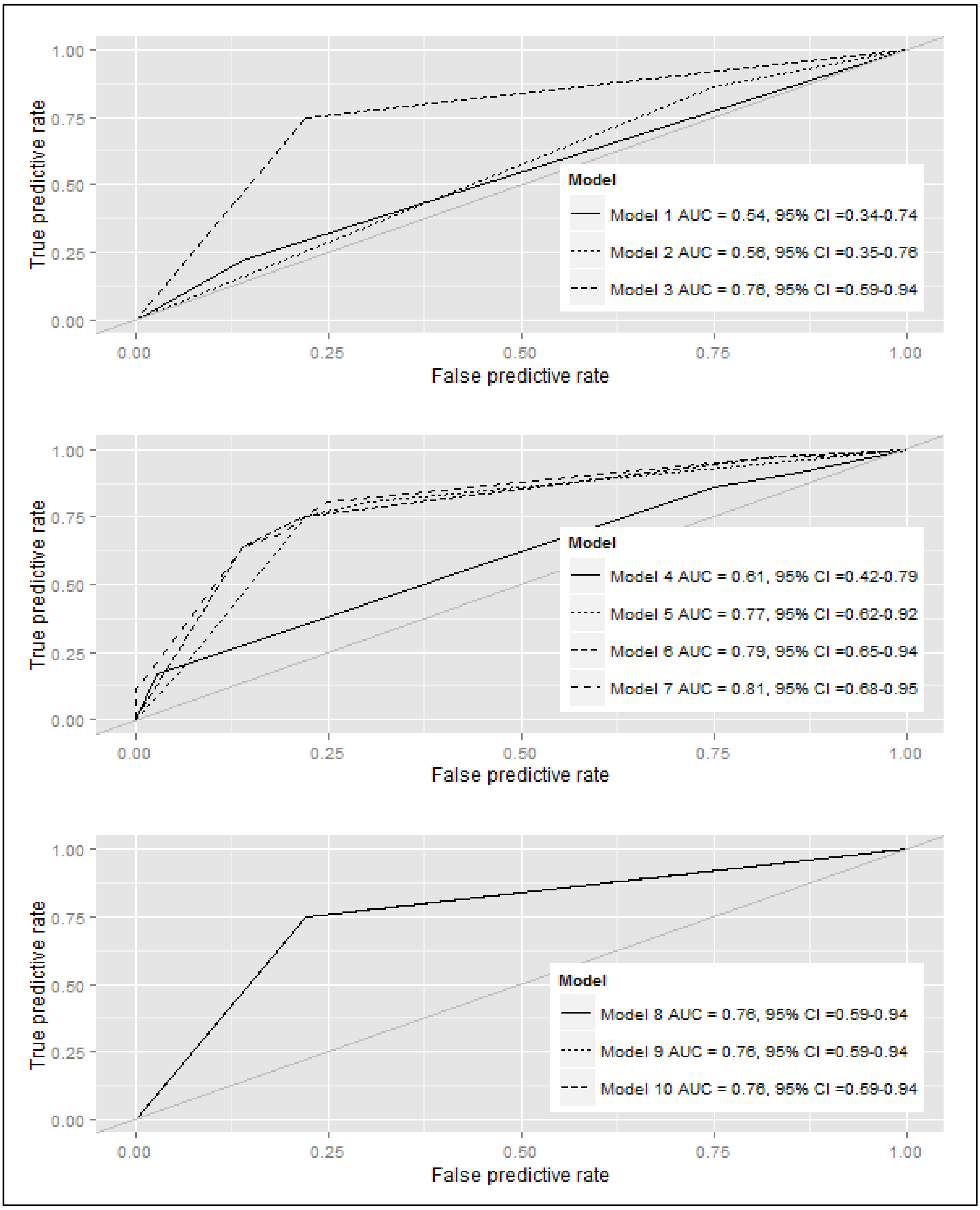


Figure 88: ROC curves and cross validated AUC (with confidence intervals) of models applied to the Otten data with outcome defined by the median. Otten data, animal based welfare measure, N =72.

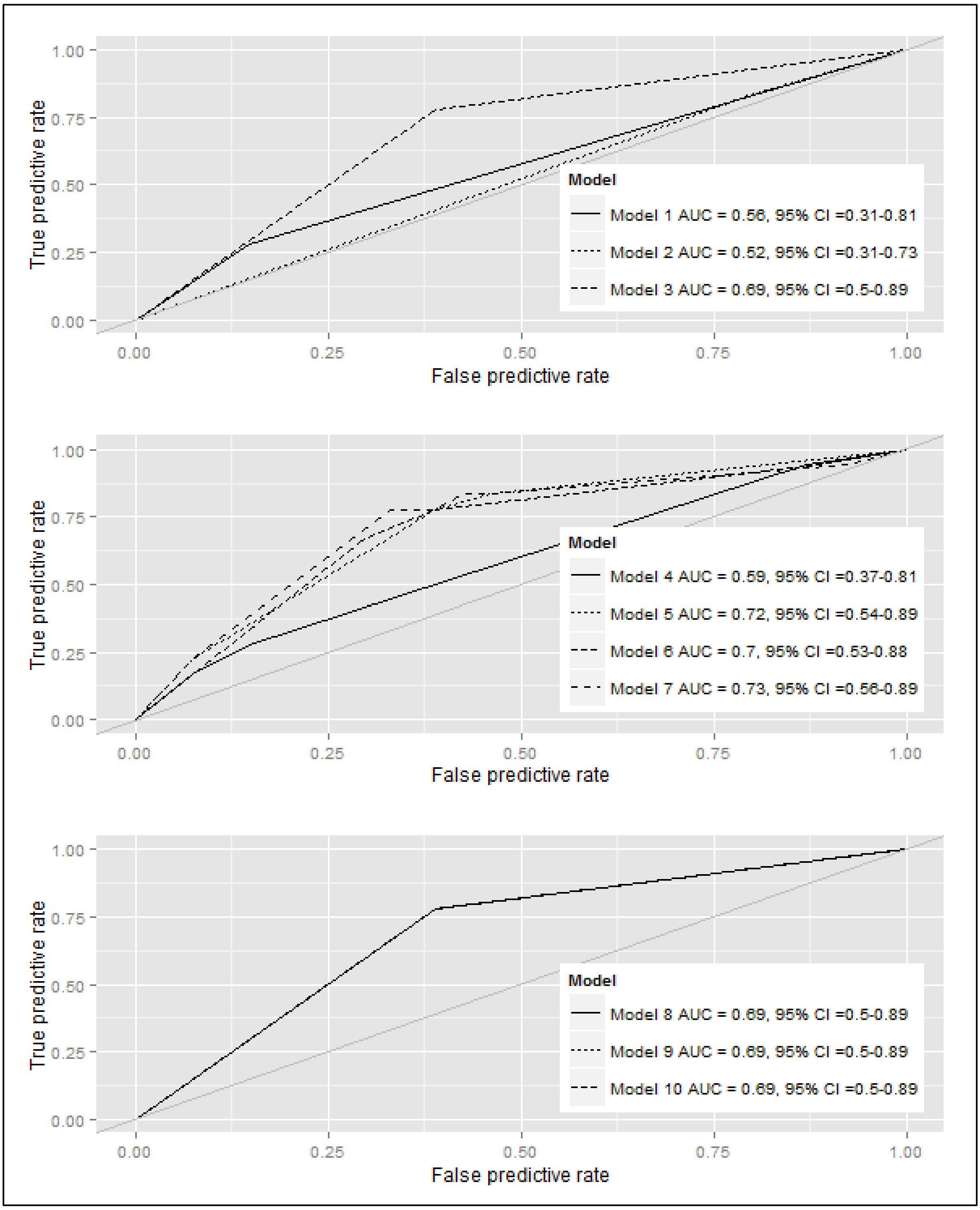


Figure 89: ROC curves and cross validated AUC (with confidence intervals) of models applied to the Otten data with outcome defined by the 75th percentile. Otten data, animal based welfare measure, N =72.

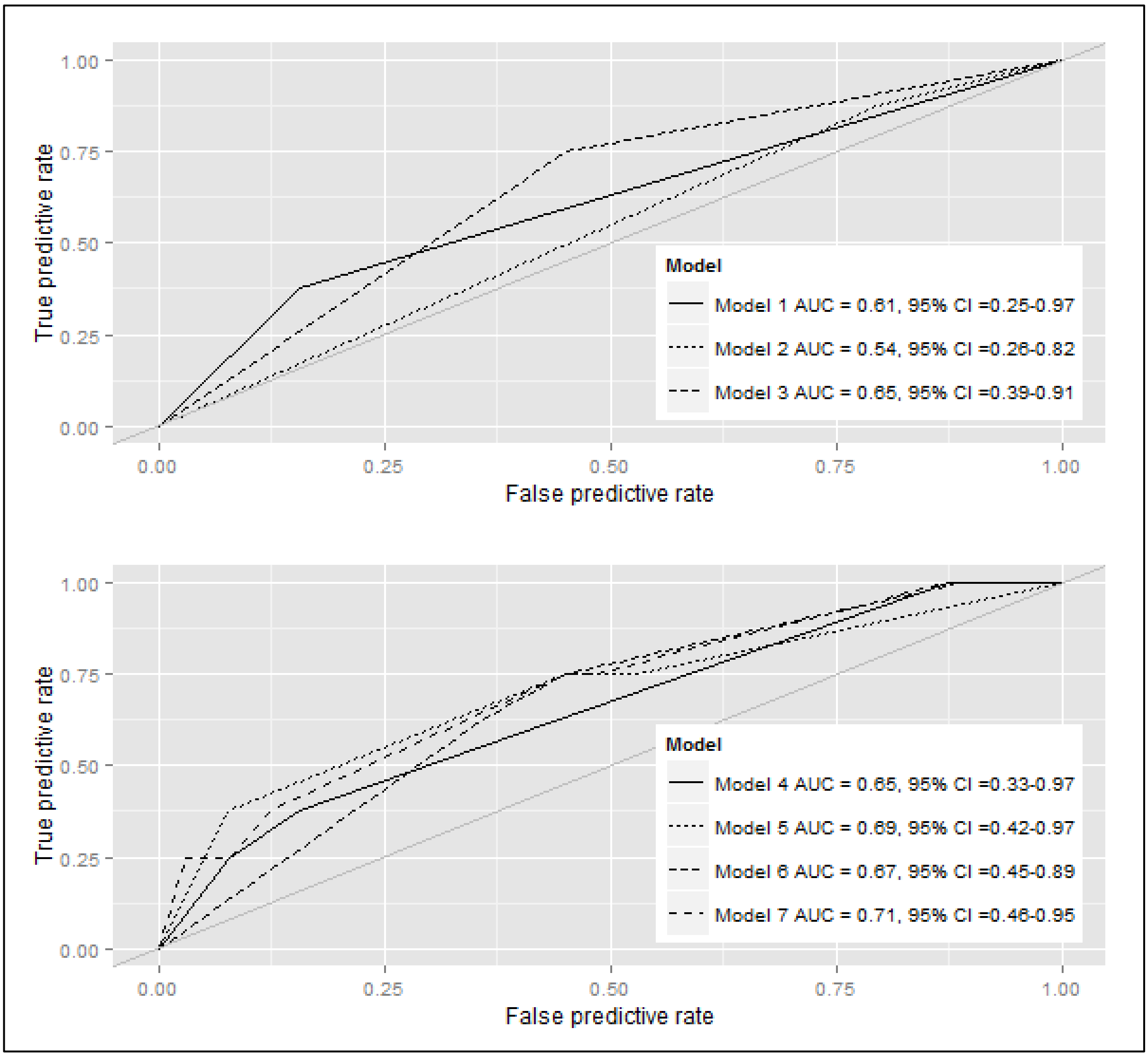


Figure 90: ROC curves and cross validated AUC (with confidence intervals) of models applied to the Otten data with outcome defined by the 90th percentile. Otten data, animal based welfare measure, N =72.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 56: .

Table 56: Comparing model results from ten different models applied to the animal based welfare measure using three different outcomes. Otten data, N =72. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	107.6	0.2614	0.5417	88.1	0.1957	0.5648	56.8	0.1033	0.6094
2	SCC	106.9	0.2592	0.5556	89.4	0.1978	0.5185	58.4	0.1034	0.5391
3	Lameness	87.2	0.1907	**0.7639**	81.0	0.1759	0.6944	56.2	0.1009	0.6484
4	Mortality + SCC	109.2	0.2591	0.6053	91.8	0.2000	0.5900	60.0	0.1067	0.6533
5	Mortality + Lameness	91.3	0.1965	**0.7728**	84.5	0.1809	**0.7176**	59.0	0.1016	0.6943
6	SCC + Lameness	89.3	0.1909	**0.7928**	85.1	0.1809	**0.7042**	60.1	0.1040	0.6689
7	Mortality + SCC + Lameness	92.5	0.1955	**0.8129**	88.1	0.1837	**0.7253**	62.0	0.1051	0.7051
8	All factors, additive	87.2	0.1907	**0.7639**	81.0	0.1759	0.6944	NA	NA	NA
9	All factors, incl. interactions	87.2	0.1907	**0.7639**	81.0	0.1759	0.6944	NA	NA	NA
10	Conditional Inference Tree	NA	NA	**0.7639**	NA	NA	0.6944	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

With the animal based welfare measure, high levels of lameness were the ABM that most consistently was associated with poorer welfare. With the outcome defined by the median or the P75, all models including lameness were significant. For the models using the median outcome, the corresponding ROC AUC were all significantly larger than 0.5, however in the P75 models, only models 5-7 had ROC AUC significantly larger than 0.5. In model 6 and 7 where SCC was included, a higher level of SCC was associated with better welfare.

None of the factors of variation were significantly associated with the animal based welfare measure.

11.2.2.2. System based welfare measure

In Table 57: , the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The model containing the additive effects of the ABMs as well as the factors of variation (model 8) revealed no significant factors. The formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 58: and Table 59: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. None of the conditional inference tree models were able to detect any significant nodes in these data.

Table 57: Model formulas for models with outcome defined by the median. Otten data, system based welfare measure, N = 72.

Model	Explanatory variables	F-test, P value
Model 1	Mortality	0.2526
Model 2	SCC	0.1529
Model 3	Lameness	0.8031
Model 4	Mortality + SCC	0.2734
Model 5	Mortality + Lameness	0.4808
Model 6	SCC + Lameness	0.3468
Model 7	Mortality + SCC + Lameness	0.4356
Model 8	No significant explanatory variables	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	No inner nodes	NA

Table 58: Model formulas for models with outcome defined by the 75th percentile. Otten data, system based welfare measure, N = 72.

Model	Explanatory variables	F-test, P value
Model 1	Mortality	0.3564
Model 2	SCC	0.0993
Model 3	Lameness	0.4959
Model 4	Mortality + SCC	0.0737
Model 5	Mortality + Lameness	0.4755
Model 6	SCC + Lameness	0.2053
Model 7	Mortality + SCC + Lameness	0.1081
Model 8	Production type	0.0154
Model 9	Production type	0.0154
Model 10	No inner nodes	NA

Table 59: Model formulas for models with outcome defined by the 90th percentile. Otten data, system based welfare measure, N = 72.

Model	Explanatory variables	F-test, P value
Model 1	Unreliable due to rank deficiency	NA
Model 2	SCC	0.1321
Model 3	Lameness	0.0308
Model 4	Unreliable due to rank deficiency	NA
Model 5	Unreliable due to rank deficiency	NA
Model 6	SCC + Lameness	0.0301
Model 7	Unreliable due to rank deficiency	NA
Model 8	Lameness	0.0308
Model 9	Did not converge	NA
Model 10	No inner nodes	NA

The ROC curves and the cross validated AUC (with confidence intervals) from the models with outcome defined by the median are presented in Figure 91: , Figure 92: and Figure 93: . ROC curves are missing where models were unreliable due to rank deficiency, where models did not converge or where no significant effects were found. When P75 was defining the outcome, the reduced models 8 and 9 contained the same variables and therefore the ROC curves were identical. The combination of lameness and SCC showed some predictive potential, whereas none of the other model's ROC AUC were significantly larger than 0.5.

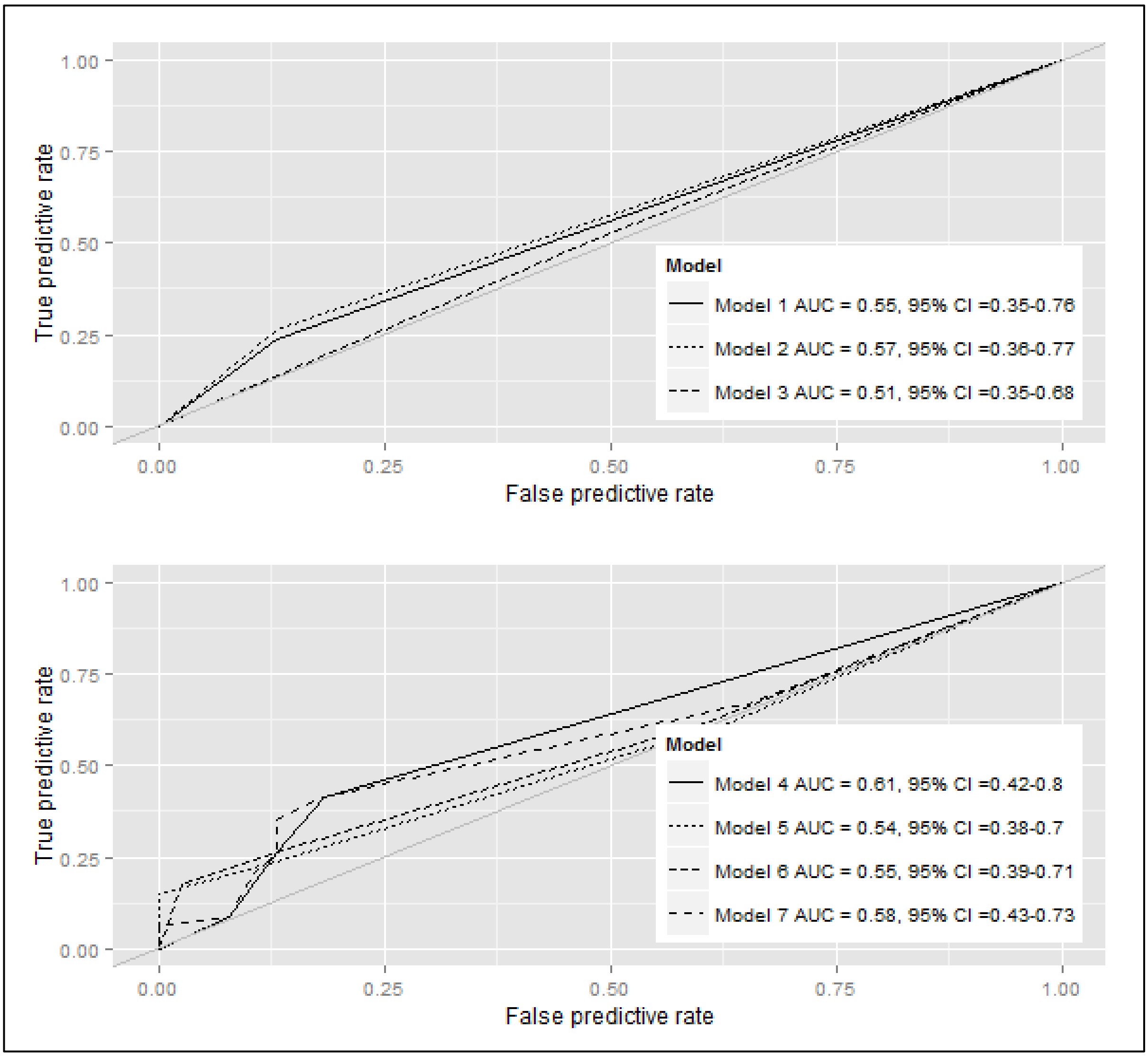


Figure 91: ROC curves and cross validated AUC (with confidence intervals) of models applied to the system based welfare measure from the Otten data with outcome defined by the median. N = 72.

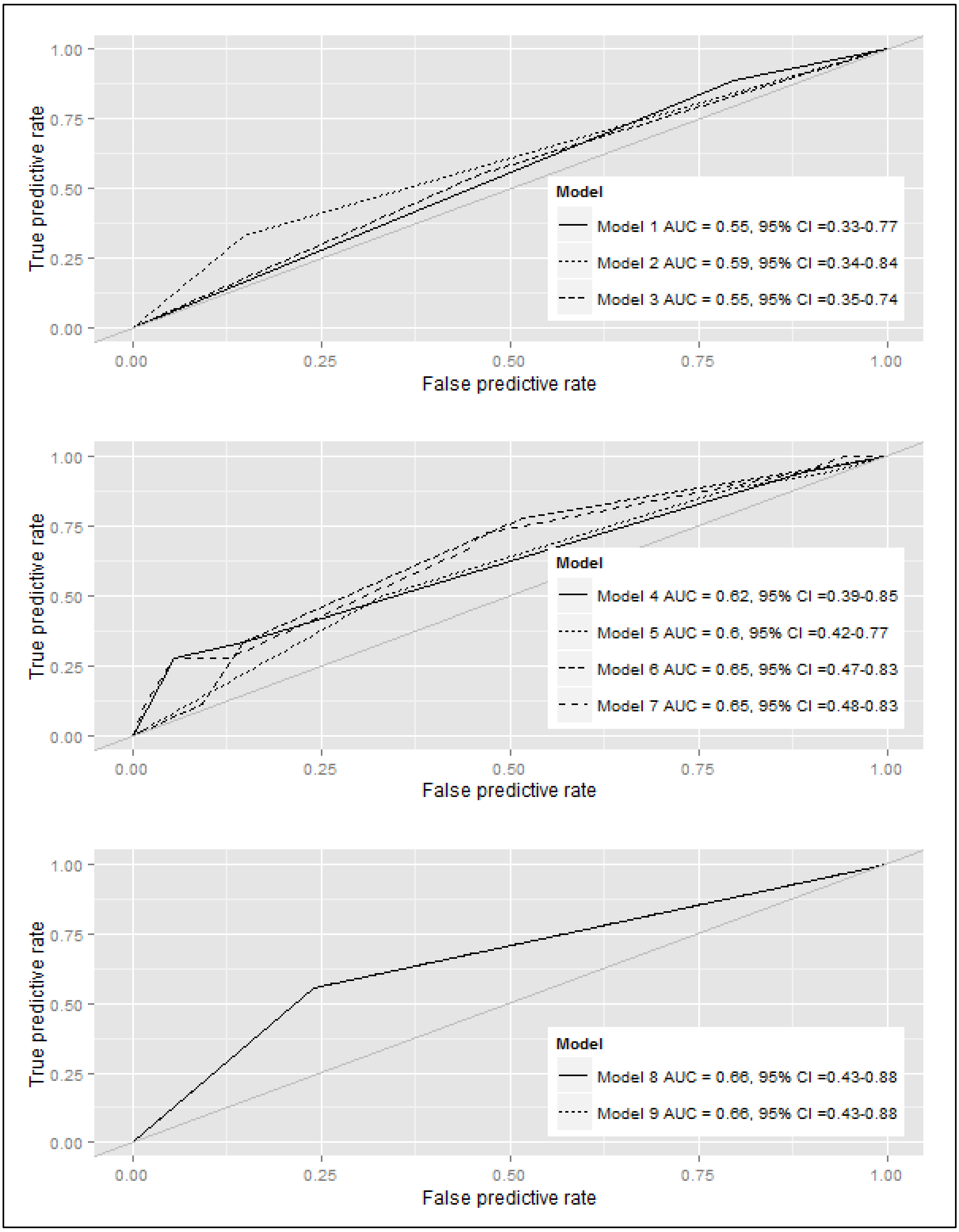


Figure 92: ROC curves and cross validated AUC (with confidence intervals) of models applied to the system based welfare measure from the Otten data with outcome defined by the 75th percentile. N = 72.

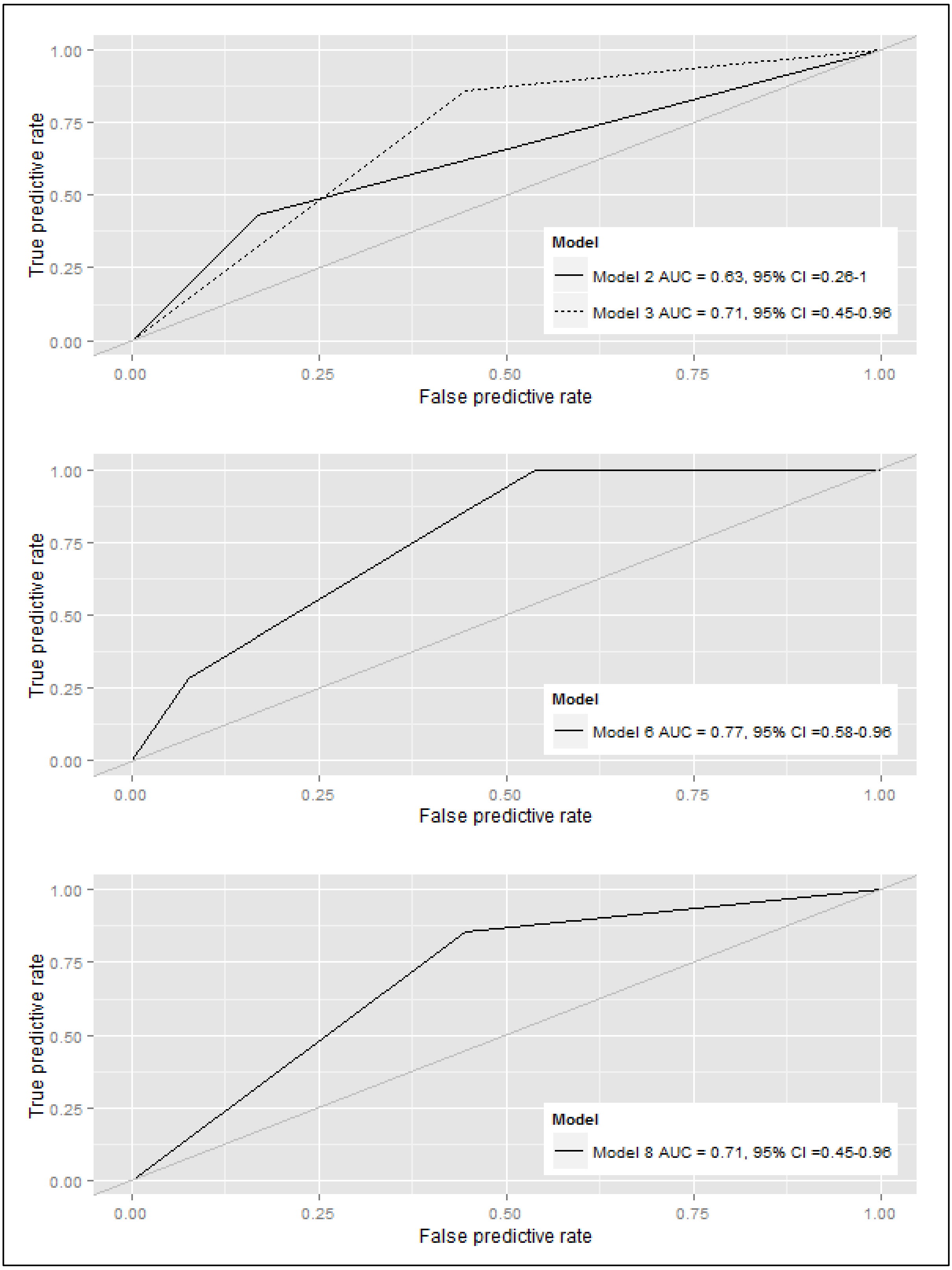


Figure 93: ROC curves and cross validated AUC (with confidence intervals) of models applied to the system based welfare measure from Otten data with outcome defined by the 90th percentile. N = 72.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 60: .

Table 60: Comparing model results from ten different models applied to the system based welfare measure using three different outcomes. Otten data, N = 72. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			**Outcome P25**			**Outcome P10**		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	106.8	0.2591	0.5519	88.7	0.1951	0.5463	NA	NA	NA
2	SCC	106.1	0.2562	0.5666	86.8	0.1920	0.5926	52.2	0.0915	0.6297
3	Lameness	108.1	0.2633	0.5147	89.1	0.1970	0.5463	49.8	0.0874	0.7055
4	Mortality + SCC	109.8	0.2623	0.6064	88.6	0.1874	0.6173	NA	NA	NA
5	Mortality + Lameness	111.0	0.2657	0.5364	92.3	0.1986	0.5957	NA	NA	NA
6	SCC + Lameness	110.3	0.2633	0.5538	90.6	0.1983	0.6512	51.7	0.0922	**0.7692**
7	Mortality + SCC + Lameness	114.0	0.2692	0.5805	92.0	0.1918	0.6548	NA	NA	NA
8	All factors, additive	NA	NA	NA	83.7	0.1826	0.6574	49.8	0.0874	0.7055
9	All factors, incl. interactions	NA	NA	NA	83.7	0.1826	0.6574	49.8	NA	NA
10	Conditional Inference Tree	NA	NA	NA	NA	NA	NA	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

The only models capable of predicting the system based welfare score was the combination of Lamenessness and SCC in P90 model 6 combining the ABM (models 5-7) when outcome was defined by the P90. In the logistic regressions of the outcome defined by the P90, higher lameness prevalence was found to be associated with poorer welfare and also had the lowest MSPE. However, the predictive value assessed by the ROC AUC was non-significant. Based on the P75 logistic regressions and the MSPE of these models, organic herds seemed to have a lower risk of classifying as poorer welfare herds when compared to conventional herds. Nevertheless, the predictive value was low.

11.2.2.3. Register based welfare measure

In Table 61: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 62: and Table 63: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 94: , Figure 95: and Figure 96: for a detailed presentation of the structure and the nodes in these models.

Table 61: Model formulas for models with outcome defined by the median. Otten data, register based welfare measure, N = 72.

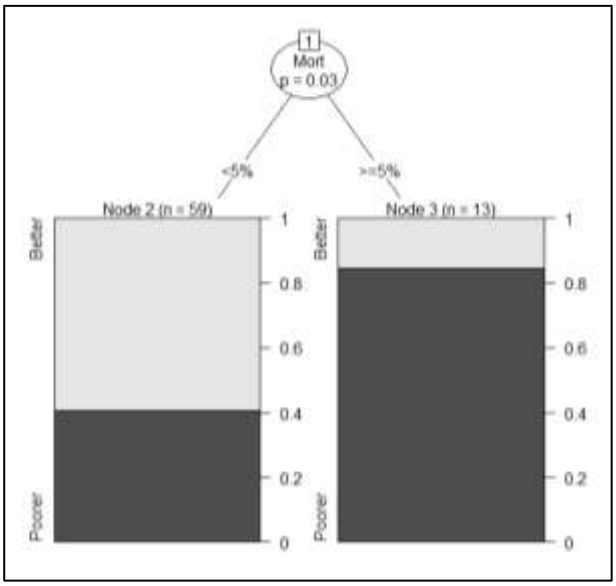
Model	Formula	F-test, P value
Model 1	Mortality	0.0029
Model 2	SCC	0.0542
Model 3	Lameness	0.6417
Model 4	Mortality + SCC	0.0061
Model 5	Mortality + Lameness	0.0117
Model 6	SCC + Lameness	0.1415
Model 7	Mortality + SCC + Lameness	0.0168
Model 8	Mortality + Milk yield	0.0016
Model 9	Mortality + Milk yield	0.0016
Model 10	Mortality	NA

Table 62: Model formulas for models with outcome defined by the 75th percentile. Otten data, register based welfare measure, N = 72.

Model	Formula	F-test, P value
Model 1	Mortality	0.0079
Model 2	SCC	0.0719
Model 3	Lameness	0.2058
Model 4	Mortality + SCC	0.0172
Model 5	Mortality + Lameness	0.0063
Model 6	SCC + Lameness	0.0822
Model 7	Mortality + SCC + Lameness	0.0107
Model 8	Mortality	0.0079
Model 9	Mortality	0.0079
Model 10	Mortality	NA

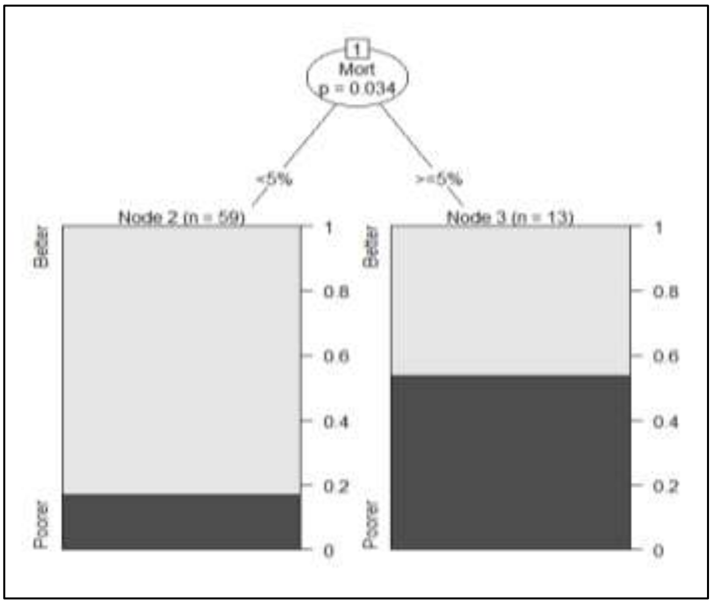
Table 63: Model formulas for models with outcome defined by the 90th percentile. Otten data, register based welfare measure, N = 72.

Model	Formula	F-test, P value
Model 1	Mortality	0.0131
Model 2	SCC	0.0184
Model 3	Lameness	0.2564
Model 4	Mortality + SCC	0.0118
Model 5	Mortality + Lameness	0.0126
Model 6	SCC + Lameness	0.0289
Model 7	Mortality + SCC + Lameness	0.0080
Model 8	Mortality	0.0131
Model 9	Mortality + SCC*Milk yield	0.0061
Model 10	Mortality + SCC	NA



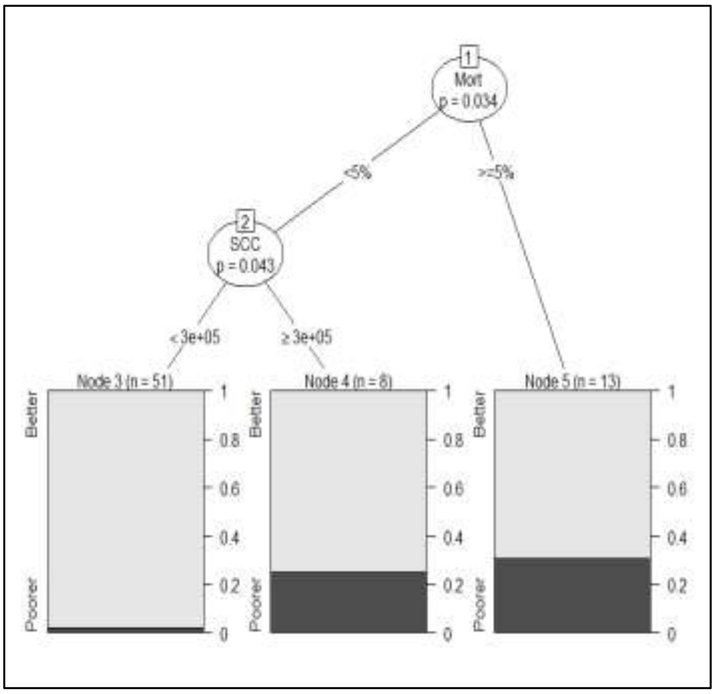
Mort: mortality

Figure 94: Conditional inference tree model with outcome defined by the median. Otten data, register based welfare measure, N = 72.



Mort: mortality

Figure 95: Conditional inference tree model with outcome defined by the 75th percentile. Otten data, register based welfare measure, N =72.



Mort: mortality

Figure 96: Conditional inference tree model with outcome defined by the 90th percentile. Otten data, register based welfare measure, N =72.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 97: , Figure 98: and Figure 99: . When defining outcome by the median, models 8 and 9 contained the same variables and when P75 was defining the outcome, the reduced models 8 and 9 and 10 contained the same variables and therefore the ROC curves were identical. Though all model defined by the median except model 3, 6 and 10 had a significant P-value from the F-test, only model 7 and 8 (equal to model 9) had a ROC AUC larger than 0.5. For the P75 models only model 5 and 7 had a significant predictive value though also model 1, 4 and 8/9 had F-test P-values smaller than 0.05. For the P90 models, all but model 3 had significant F-tests but only model 4, 5, 7, 9 and 10 had a ROC AUC larger than 0.5.

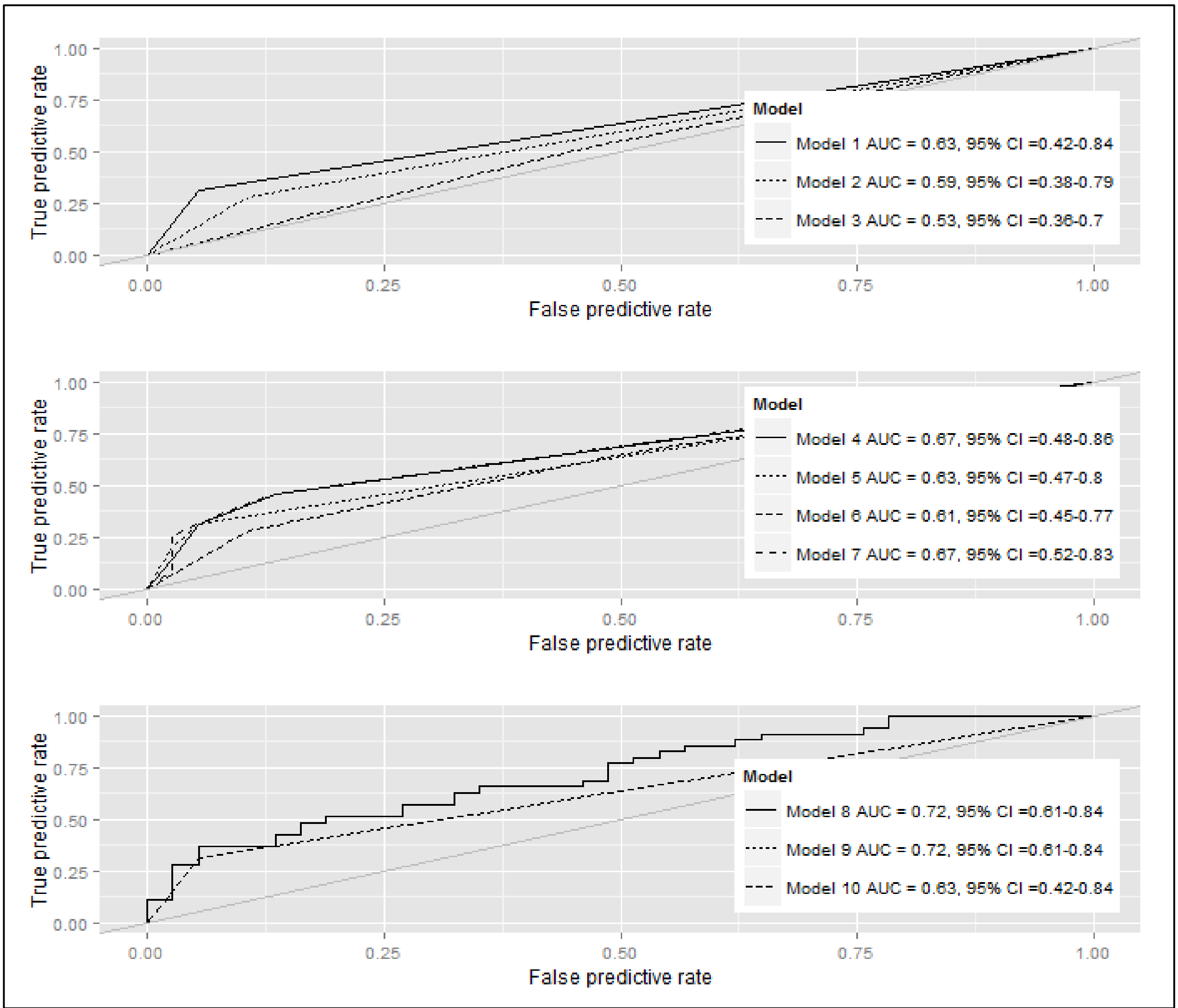


Figure 97: ROC curves and cross validated AUC (with confidence intervals) of models applied to the register based welfare measure from the Otten data with outcome defined by the median. N = 72.

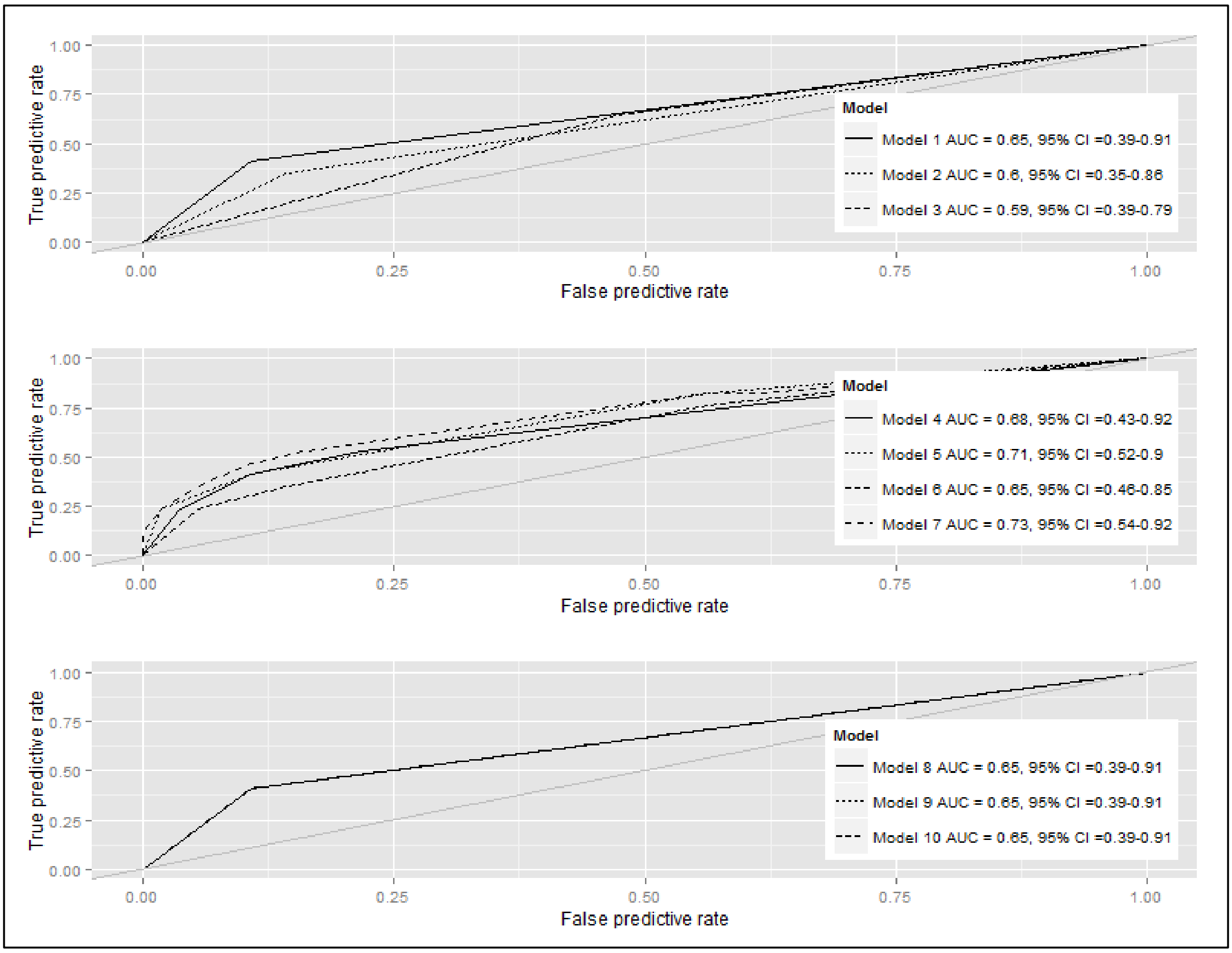


Figure 98: ROC curves and cross validated AUC (with confidence intervals) of models applied to the register based welfare measure from the Otten data with outcome defined by the 75th percentile. N = 72.

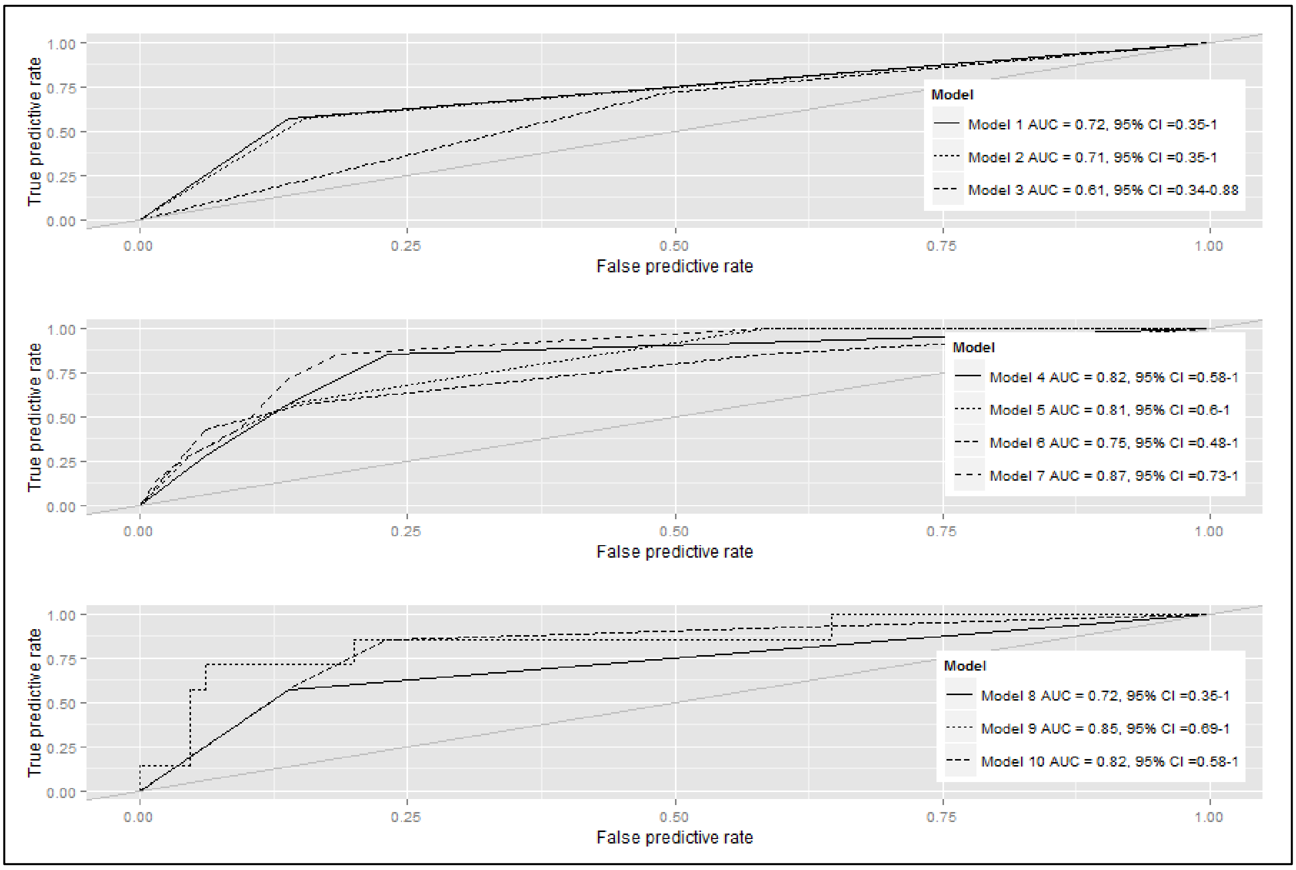


Figure 99: ROC curves and cross validated AUC (with confidence intervals) of models applied to the system based welfare measure from Otten data with outcome defined by the 90th percentile. N = 72.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 72.

Table 64: Comparing model results from ten different models applied to the animal based welfare measure using three different outcomes. Otten data, N = 72. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	99.4	0.2321	0.6301	80.2	0.1719	0.6513	48.3	0.0860	0.7165
2	SCC	104.6	0.2505	0.5888	84.0	0.1833	0.6037	48.9	0.0869	0.7088
3	Lameness	108.1	0.2634	0.5274	85.7	0.1865	0.5872	53.2	0.0912	0.6110
4	Mortality + SCC	102.4	0.2338	0.6695	83.4	0.1751	0.6770	49.9	0.0926	**0.8220**
5	Mortality + Lameness	103.7	0.2399	0.6317	81.4	0.1669	**0.7102**	50.0	0.0909	**0.8121**
6	SCC + Lameness	108.7	0.2569	0.6097	86.5	0.1832	0.6519	51.7	0.0869	0.7527
7	Mortality + SCC + Lameness	106.7	0.2405	**0.6718**	84.6	0.1691	**0.7273**	51.2	0.0946	**0.8736**
8	All factors, additive	99.7	0.2249	**0.7205**	80.2	0.1719	0.6513	48.3	0.0860	0.7165
9	All factors, incl. interactions	99.7	0.2249	**0.7205**	80.2	0.1719	0.6513	50.6	0.0910	**0.8505**
10	Conditional Inference Tree	NA	NA	0.6301	NA	NA	0.6513	NA	NA	**0.8198**

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

The combination of the three ABMs was useful for the prediction of the welfare status defined by the Otten register based measure regardless of which of the threshold for the dichotomisation of the outcome were used. Generally, high mortality showed the strongest association with poorer welfare regardless of how the outcome was defined. All models containing mortality were significant deemed by the F-test and mortality was also found significant by all three conditional inference tree models. However, evaluation of the ROC AUC showed that not all these models were capable of predicting the true welfare status above chance.

Among the identified factors of variation (production type, herd size, milk yield), only higher milk yield showed any association with the outcome. Furthermore, the inclusion of the milk yield in the logistic regression reduced the MSPE and increased the ROC AUC. However, the importance of milk yield was not confirmed by the conditional inference tree models.

11.2.2.4. Summarising results on Otten data

With the Otten data, three different welfare measurements were tested: The first one was based on animal based indicators, the second was based on resource based indicators and the third one was based on indicators identified in register data. None of the ABMs were significantly associated with all three welfare measurements; still, all three ABMs did show some predictive potential in at least one of the outcomes.

Five out of the ten factors of variation were identified in the data. However, only production type, herd size and milk yield were used for the analyses. Breed was omitted because of the much skewed distribution of data with the vast majority being Danish Holsteins. In Denmark, all organic farms are obliged to use pasture grazing during summer and therefore pasture access was omitted because it was confounded by production type. Analyses of the remaining factors of variation indicated that organic herds were more likely to be classified in the better welfare group in the system based welfare definition and that higher milk yield was associated with an increased risk of poorer welfare based on the register data based welfare definition.

11.2.3. **Burow data**

In Table 65: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 66: and Table 67: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 100: and Figure 101: for a detailed presentation of the structure and the nodes in these models. Generally, due to the small sample size in these data there were problems with rank deficiency which affected the model estimates resulting in unrealistic model coefficients

Table 65: Model formulas for models with outcome defined by the median. Burow data, AWI Summer, N = 31.

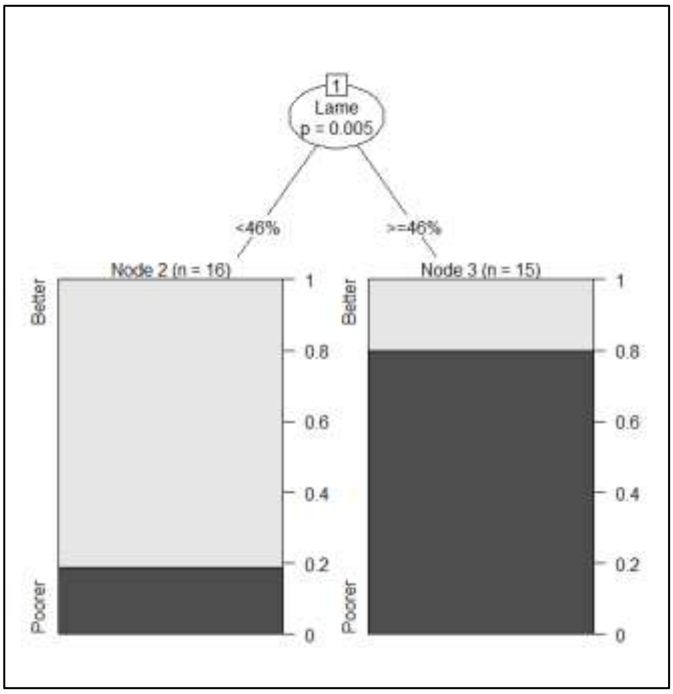
Model	Formula	F-test, P value
Model 1	Mortality	0.8527
Model 2	SCC	0.4723
Model 3	Lameness	0.0004
Model 4	Mortality + SCC	0.7706
Model 5	Mortality + Lameness	0.0008
Model 6	SCC + Lameness	0.0018
Model 7	Mortality + SCC + Lameness	0.0027
Model 8	Lameness	0.0004
Model 9	Did not converge	NA
Model 10	Lameness	NA

Table 66: Model formulas for models with outcome defined by the 75th percentile. Burow data, AWI Summer, N = 31.

Model	Formula	F-test, P value
Model 1	Mortality	0.9156
Model 2	SCC	0.9516
Model 3	Unreliable due to rank deficiency	NA
Model 4	Mortality + SCC	0.9913
Model 5	Unreliable due to rank deficiency	NA
Model 6	Unreliable due to rank deficiency	NA
Model 7	Unreliable due to rank deficiency	NA
Model 8	Did not converge	NA
Model 9	Did not converge	NA
Model 10	Production type	NA

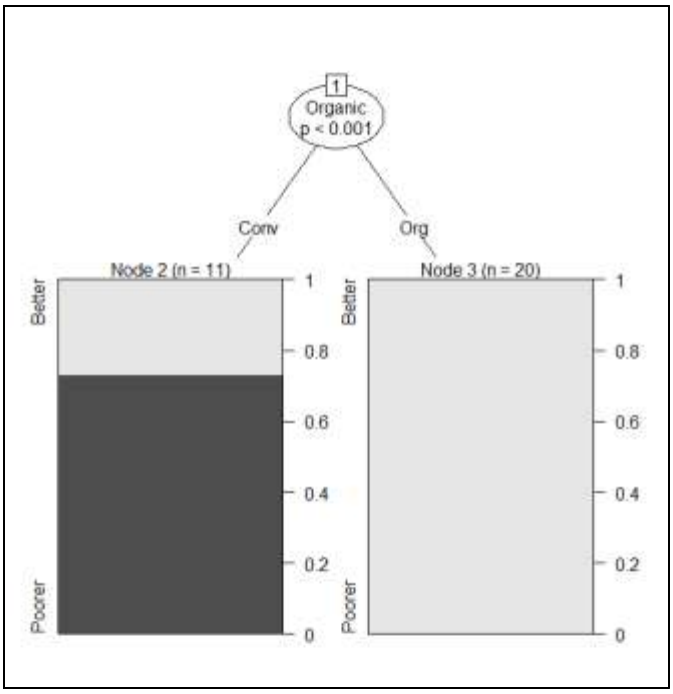
Table 67: Model formulas for models with outcome defined by the 90th percentile. Burow data, AWI Summer, N = 31.

Model	Formula	F-test, P value
Model 1	Mortality	0.5792
Model 2	SCC	0.7597
Model 3	Unreliable due to rank deficiency	NA
Model 4	Mortality + SCC	0.7884
Model 5	Unreliable due to rank deficiency	NA
Model 6	Unreliable due to rank deficiency	NA
Model 7	Unreliable due to rank deficiency	NA
Model 8	Unreliable due to rank deficiency	NA
Model 9	Did not converge	NA
Model 10	No inner nodes	NA



Lame: lameness

Figure 100: Inference tree model with outcome defined by the median. Burow data, N =31.



Conv: conventional, Org: organic

Figure 101: Conditional inference tree model with outcome defined by the 75th percentile. Burow data, N =31

The ROC curves and the cross validated AUC (with confidence intervals) from the models with outcome defined by the median are presented in Figure 102: . In Figure 103: the ROC curve from the conditional inference tree model applied to the P75 defined outcome. Due to the general rank deficiency problems in the other models defined by the P75 and in most of the P90 model, ROC curves for these modes are not presented. With the outcome defined by the median, model 8 and 10 contained the same variables and the ROC curves therefore were identical.

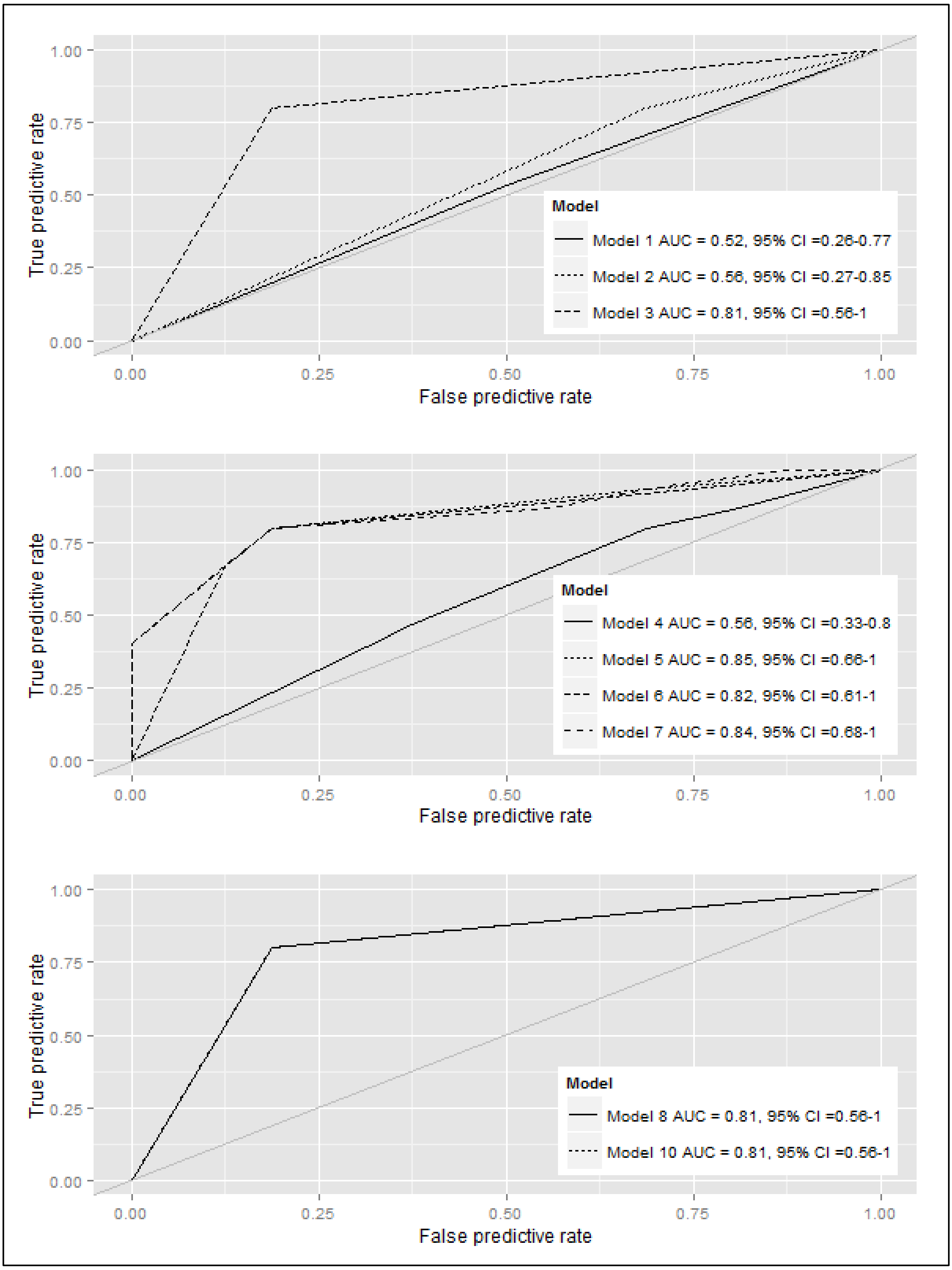


Figure 102: ROC curve and cross validated AUC (with confidence intervals) of conditional inference tree model applied to the AWI Summer welfare measure from the Burow data with outcome defined by the median. N = 31.

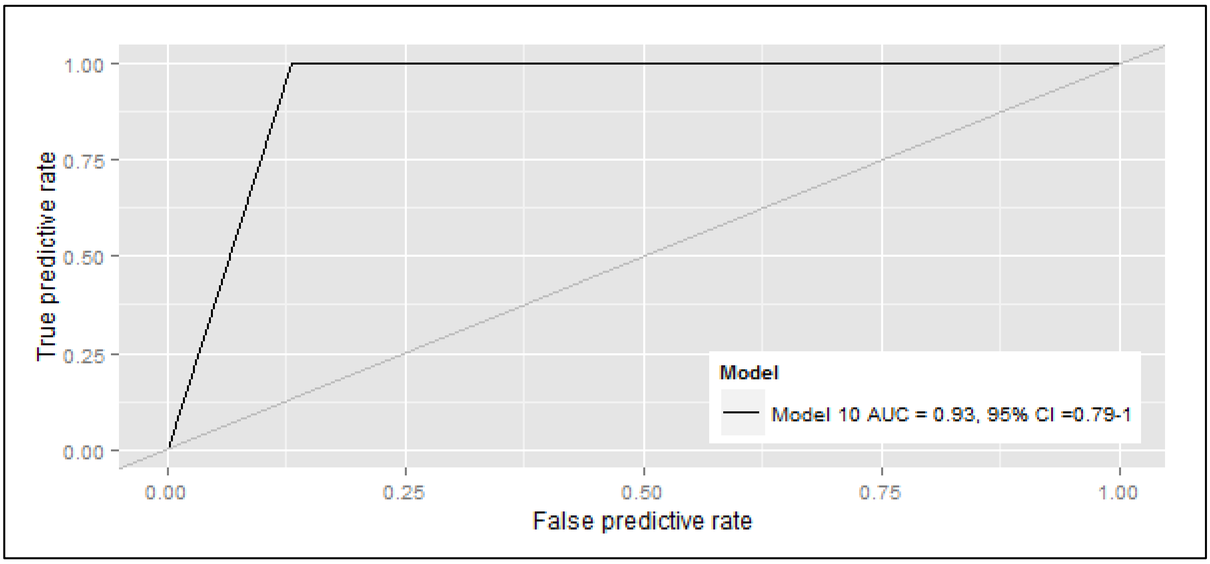


Figure 103: ROC curves and cross validated AUC (with confidence intervals) of models applied to the AWI Summer welfare measure from the Burow data with outcome defined by the 75th percentile. N = 31.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 68: . The many missing values are due to the problems with rank deficiency in many of the models.

Table 68: Comparing model results from ten different models applied to the AWI, Summer welfare measure using three different outcomes. Burow data, N = 31. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			**Outcome P25**			**Outcome P10**		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	49.8	0.2845	0.5167	42.3	0.2183	0.5109	26.3	0.0986	0.5833
2	SCC	49.3	0.2807	0.5562	42.3	0.2191	0.5054	26.5	0.1010	0.5417
3	Lameness	37.3	0.1780	**0.8063**	NA	NA	NA	NA	NA	NA
4	Mortality + SCC	52.7	0.3005	0.5646	NA	NA	NA	29.5	0.1036	0.6071
5	Mortality + Lameness	39.1	0.1789	**0.8479**	NA	NA	NA	NA	NA	NA
6	SCC + Lameness	40.6	0.1902	**0.8167**	NA	NA	NA	NA	NA	NA
7	Mortality + SCC + Lameness	42.5	0.1953	**0.8438**	NA	NA	NA	NA	NA	NA
8	All factors, additive	37.3	0.1780	**0.8063**	NA	NA	NA	NA	NA	NA
9	All factors, incl. interactions	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	Conditional Inference Tree	NA	NA	**0.8063**	NA	NA	**0.9348**	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

Due to the rather small sample size in the Burow data, there were problems with rank deficiency in many of the analyses and results were therefore rather inconclusive. Lameness was the only ABM that was associated with the overall welfare measure in the logistic regression models. In the logistic regression models suffering from rank deficiency, lameness also was deemed to be important. However, coefficients were unrealistic high due to the small number of observations (data not shown) and therefore, caution should be taken in the interpretation. Lameness was also identified in the conditional inference tree model when the outcome was defined by the median.

Milk yield, herd size and production type were included as factors of variation. When defining the outcome by the P75, the conditional inference model identified production type as a significant predictor where organic herds were most likely to have a better welfare. However, from the descriptive analysis we could see that the proportion of herds with the high level of lameness seemed to be much lower in organic herds that in conventional herds (Figure 33:) and it is therefore possible that the effect of production type actually is an effect of the lower lameness prevalence in the organic herds compared to the conventional herds.

11.2.4. **The French Welfare Quality® data**

Results from statistical analyses of the four principle welfare scores in the French Welfare Quality® data are presented individually before a summation of all results is given.

11.2.4.1. Health principle

In Table 69: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P25 and the P10 are shown in Table 70: and Table 71: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 104: and Figure 105: for a detailed presentation of the structure and the nodes in these models. For the outcome defined by the P10 no significant nodes were identified in the conditional inference tree model.

Table 69: Model formulas for models with outcome defined by the median. French WQ® data, health principle, N = 129.

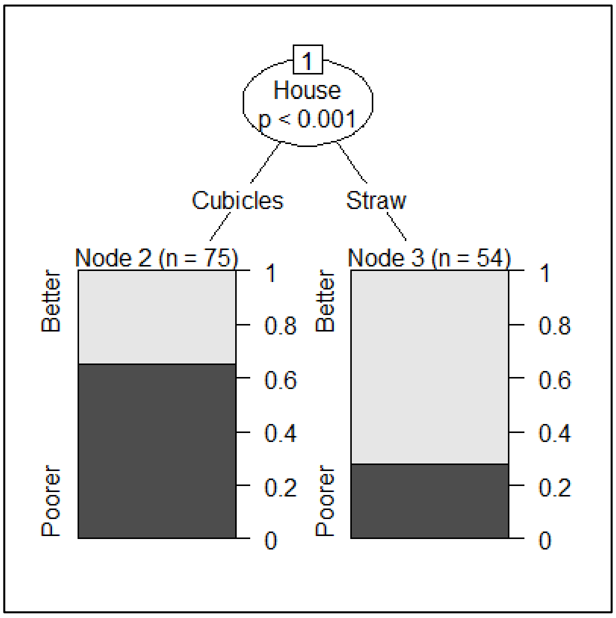
Model	Formula	F-test, P value
Model 1	Mortality	0.3938
Model 2	SCC	0.6602
Model 3	Lameness	0.9477
Model 4	Mortality + SCC	0.5819
Model 5	Mortality + Lameness	0.6891
Model 6	SCC + Lameness	0.9066
Model 7	Mortality + SCC + Lameness	0.7811
Model 8	Herd size + Pasture + Housing	<0.0001
Model 9	Herd size + Pasture + Housing	<0.0001
Model 10	Housing system	NA

Table 70: Model formulas for models with outcome defined by the 25th percentile. French WQ® data, health principle, N = 129.

Model	Formula	F-test, P value
Model 1	Mortality	0.5011
Model 2	SCC	0.0914
Model 3	Lameness	0.1310
Model 4	Mortality + SCC	0.2225
Model 5	Mortality + Lameness	0.2732
Model 6	SCC + Lameness	0.1300
Model 7	Mortality + SCC + Lameness	0.2392
Model 8	Herd size + Housing system + Breed	0.0006
Model 9	Breed + Housing system*Herd size	<0.0001
Model 10	Housing system	NA

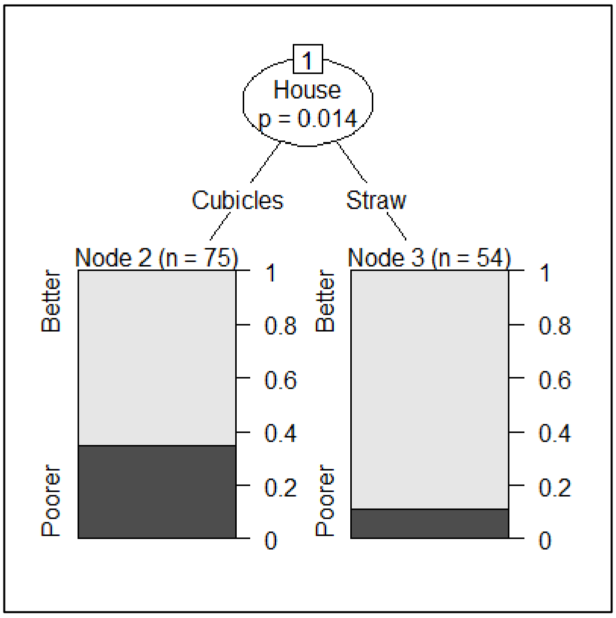
Table 71: Model formulas for models with outcome defined by the 10th percentile. French WQ® data, behaviour principle, N = 128.

Model	Formula	F-test, P value
Model 1	Mortality	0.7074
Model 2	SCC	0.1314
Model 3	Lameness	0.4120
Model 4	Mortality + SCC	0.2581
Model 5	Mortality + Lameness	0.6470
Model 6	SCC + Lameness	0.2905
Model 7	Mortality + SCC + Lameness	0.4047
Model 8	Herd size + Breed	0.0015
Model 9	Did not converge	NA
Model 10	No inner nodes	NA



House: housing system

Figure 104: Conditional inference tree model with outcome defined by the median. French WQ® health principle score, N =129.



House: housing system

Figure 105: Conditional inference tree model with outcome defined by the 25th percentile. French WQ® health principle score, N =129.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 106: , Figure 107: and Figure 108: . When defining the outcome by the P10, model 9 did not converge and the conditional inference tree model did not identify any significant nodes. Therefore, no ROC curves from these models are presented.

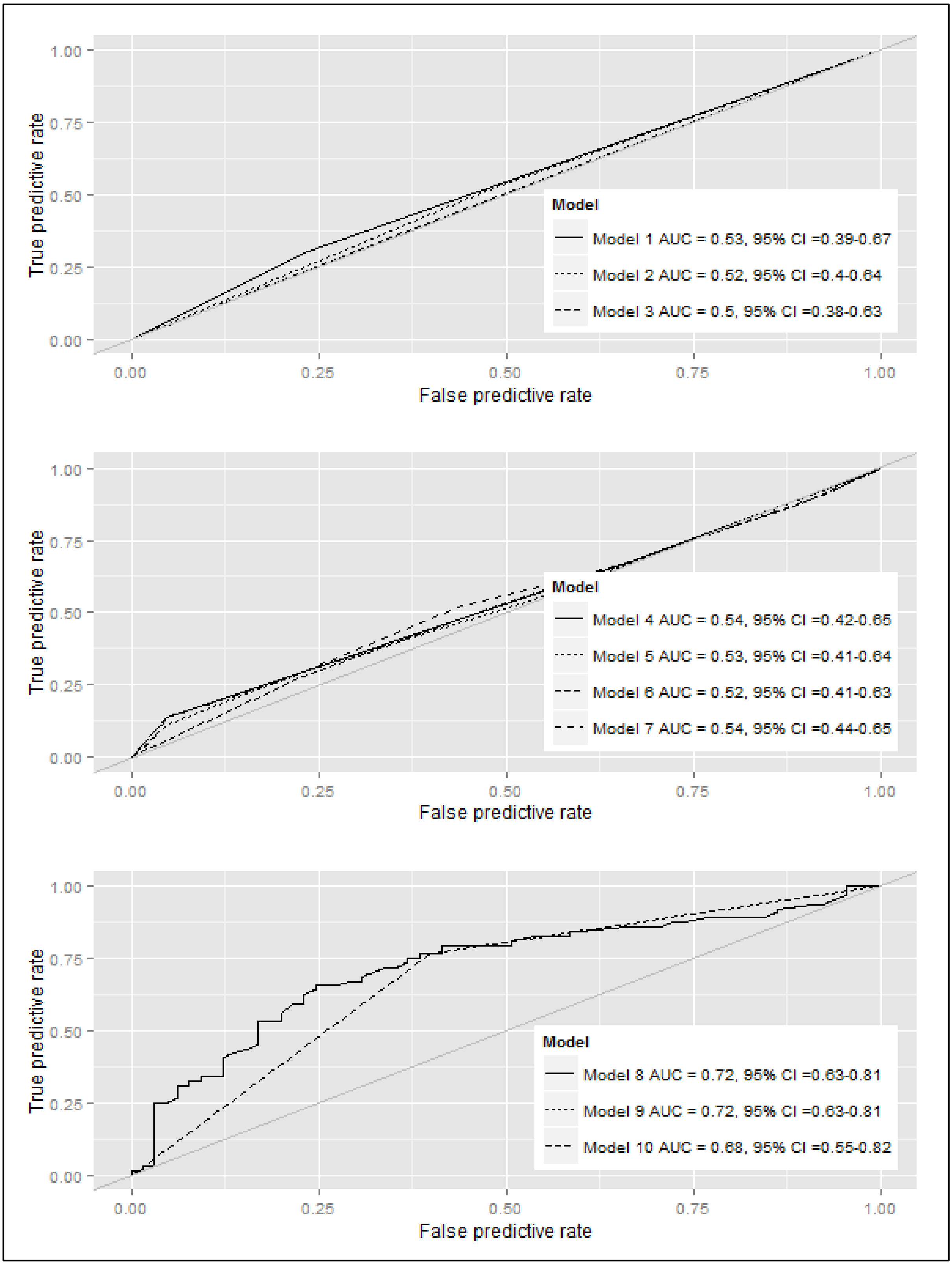


Figure 106: ROC curves and cross validated AUC (with confidence intervals) of models applied to the health principle welfare measure from the French WQ® data with outcome defined by the median. N = 129.

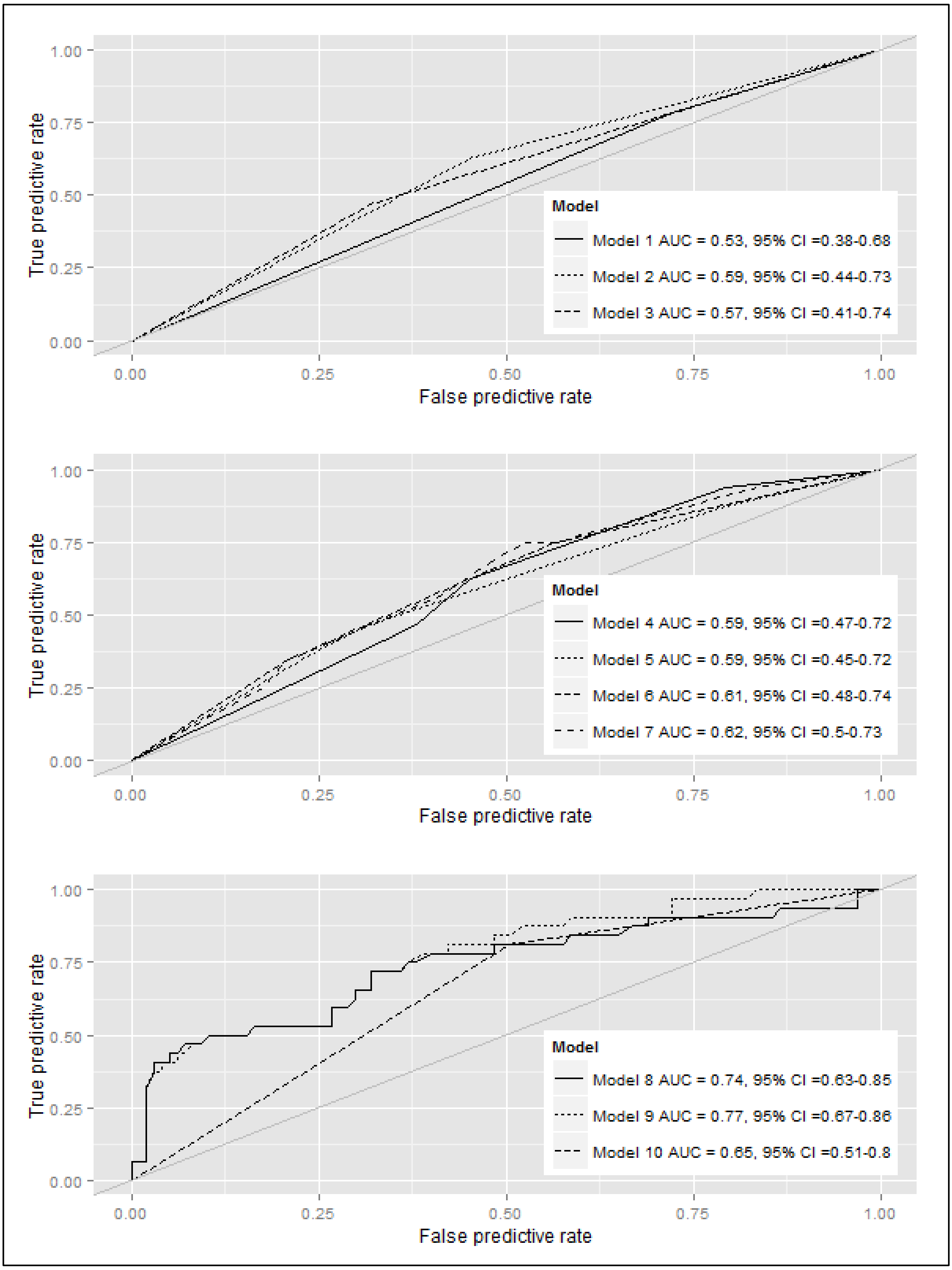


Figure 107: ROC curves and cross validated AUC (with confidence intervals) of models applied to the health principle welfare measure from the French WQ® data with outcome defined by the 25th percentile. N = 129.

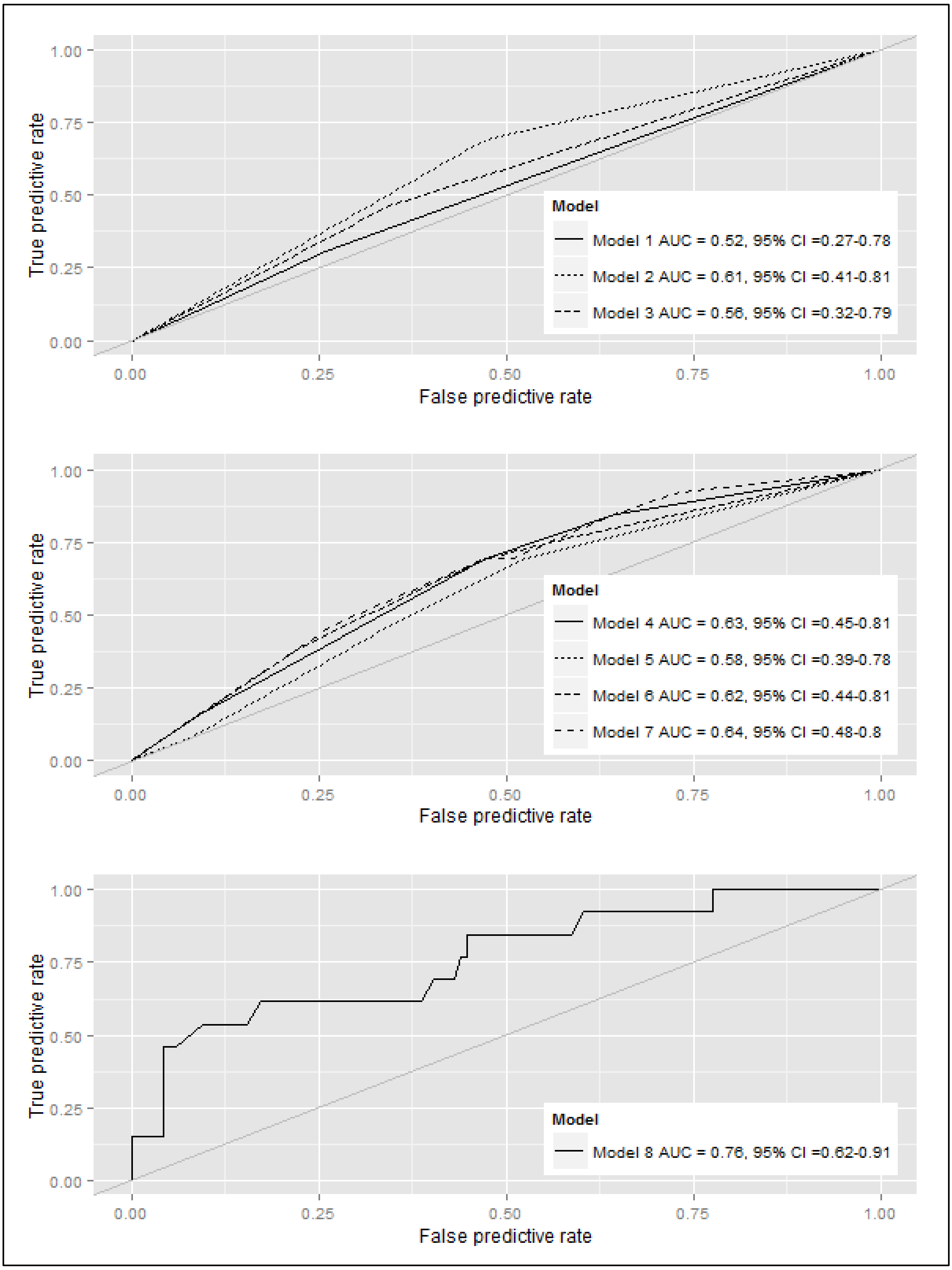


Figure 108: ROC curves and cross validated AUC (with confidence intervals) of models applied to the health principle welfare measure from the French WQ® data with outcome defined by the 10th percentile. N = 129.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 72: .

Table 72: Comparing model results from ten different models applied to the WQ® health principle welfare measure using three different outcomes. French WQ® data, N = 129. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			**Outcome P25**			**Outcome P10**		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	187.8	0.2565	0.5331	153.8	0.1916	0.5298	93.9	0.0935	0.5245
2	SCC	187.3	0.2555	0.5320	154.3	0.1926	0.5006	93.7	0.0929	0.5262
3	Lameness	188.5	0.2579	0.5028	152.0	0.1892	0.5746	93.4	0.0932	0.5584
4	Mortality + SCC	191.4	0.2580	0.5439	158.7	0.1948	0.5284	98.4	0.0943	0.5554
5	Mortality + Lameness	192.7	0.2605	0.5268	156.5	0.1917	0.5888	98.0	0.0947	0.5806
6	SCC + Lameness	192.2	0.2595	0.5352	156.7	0.1924	0.5855	97.6	0.0941	0.5723
7	Mortality + SCC + Lameness	196.3	0.2621	0.5518	161.3	0.1949	0.5958	102.3	0.0956	0.5905
8	All factors, additive	176.5	0.2228	**0.7238**	146.5	0.1746	**0.7176**	89.8	0.0904	0.6369
9	All factors, incl. interactions	176.5	0.2228	**0.7238**	148.5	0.1653	**0.7862**	89.8	NA	NA
10	Conditional Inference Tree	NA	NA	**0.6828**	NA	NA	**0.6537**	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

None of the models containing only the ABMs were able to predict the true WQ® health class regardless of the definition of the outcome. With the outcome defined by the median, herds with straw yards were at a lower risk of being in the poorer welfare class than herds with cubicles and herds with pasture access were less likely to categorise in the poorer welfare class than herds with no pasture access. Also, herd size seemed to be of importance with increasing herd size linked to an increased odds ratio of being classified in the poorer welfare class. However, herd size was not deemed significant by any of the conditional inference tree models. When the P25 was defining the outcome, herds with dual purpose breeds were less likely to be in the poorer welfare group compared to herds with milking breeds. Also, increasing herd size and cubicle housing were associated with a higher risk of classifying in the poorer welfare group. In the P10 models, herd size and breed were again found as significant predictors in the logistic regression. However, the ROC AUC was not significantly larger than 0.5 and the conditional inference tree model did not confirm the importance of these variables.

11.2.4.2. Housing principle

In Table 73: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P25 and the P10 are shown in Table 74: and Table 75: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 109: , Figure 110: and Figure 111: for a detailed presentation of the structure and the nodes in these models.

Table 73: Model formulas for models with outcome defined by the median. French WQ® data, housing principle, N = 129.

Model	Formula	F-test, P value
Model 1	Mortality	0.8089
Model 2	SCC	0.5388
Model 3	Lameness	0.0996
Model 4	Mortality + SCC	0.8200
Model 5	Mortality + Lameness	0.2401
Model 6	SCC + Lameness	0.1402
Model 7	Mortality + SCC + Lameness	0.2646
Model 8	Herd size + Pasture + Housing	<0.0001
Model 9	Pasture + Housing system	<0.0001
Model 10	Housing system	NA

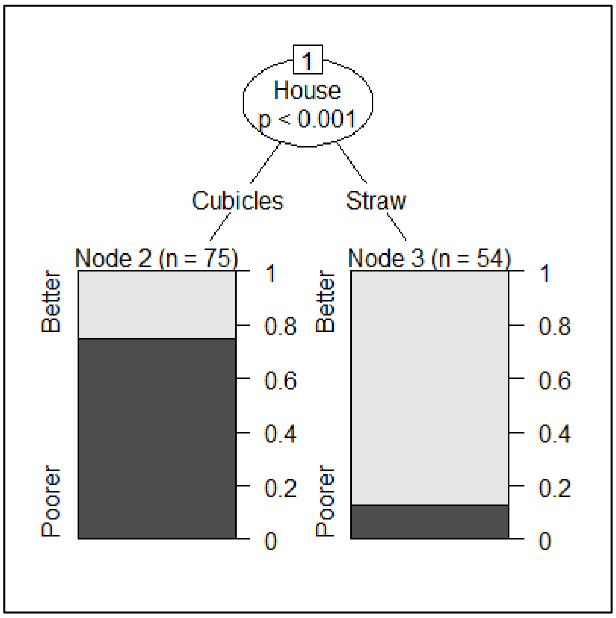
Table 74: Model formulas for models with outcome defined by the 25th percentile. French WQ® data, housing principle, N = 129.

Model	Formula	F-test, P value
Model 1	Mortality	0.2988
Model 2	SCC	0.8756
Model 3	Lameness	0.0734
Model 4	Mortality + SCC	0.5509
Model 5	Mortality + Lameness	0.0995
Model 6	SCC + Lameness	0.1914
Model 7	Mortality + SCC + Lameness	0.2009
Model 8	Housing system + Breed	<0.0001
Model 9	Housing system + Breed	<0.0001
Model 10	Housing system	NA

Table 75: Model formulas for models with outcome defined by the 10th percentile. French WQ® data, housing principle, N = 129.

Model	Formula	F-test, P value
Model 1	Mortality	0.6255
Model 2	SCC	0.9796
Model 3	Lameness	0.0532
Model 4	Mortality + SCC	0.8863
Model 5	Mortality + Lameness	0.1267
Model 6	SCC + Lameness	0.1356
Model 7	Mortality + SCC + Lameness	0.2314
Model 8	Unrealiable due to rank deficiency	NA
Model 9	Did not converge	NA

Model 10 Housing system NA



House: housing system

Figure 109: Conditional inference tree model with outcome defined by the median. French WQ® Housing principle score, N =129.

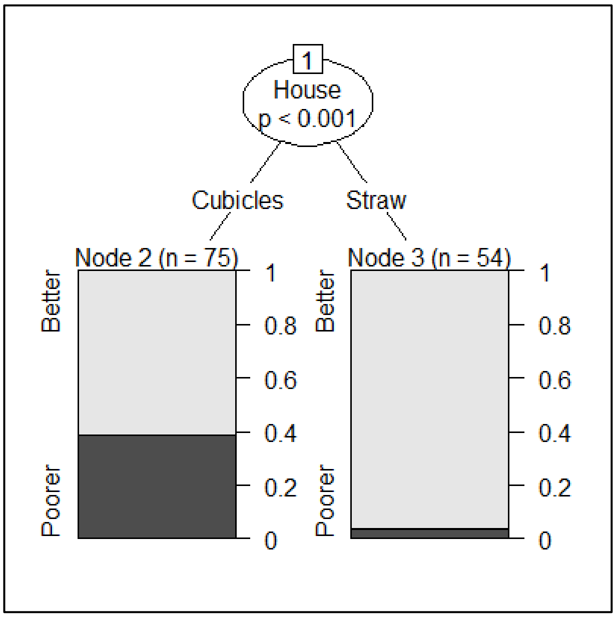


Figure 110: Conditional inference tree model with outcome defined by the 25th percentile. French WQ® Housing principle score, N =129.

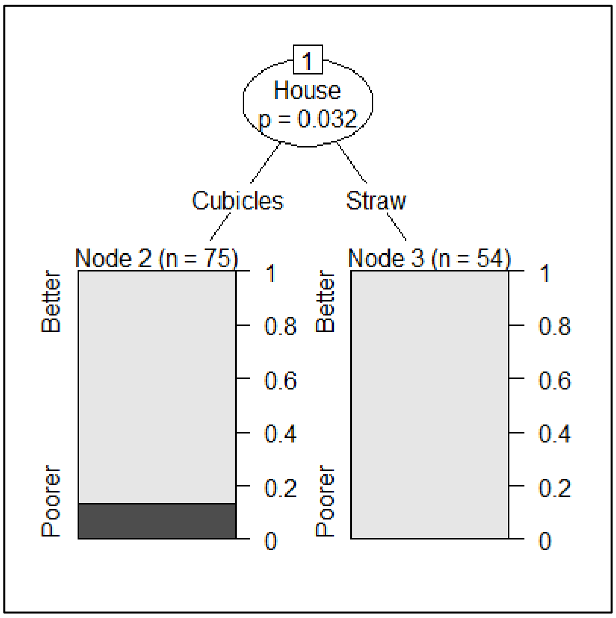


Figure 111: Conditional inference tree model with outcome defined by the 10th percentile. French WQ® Housing principle score, N =129.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 112: , Figure 113: and Figure 114: .

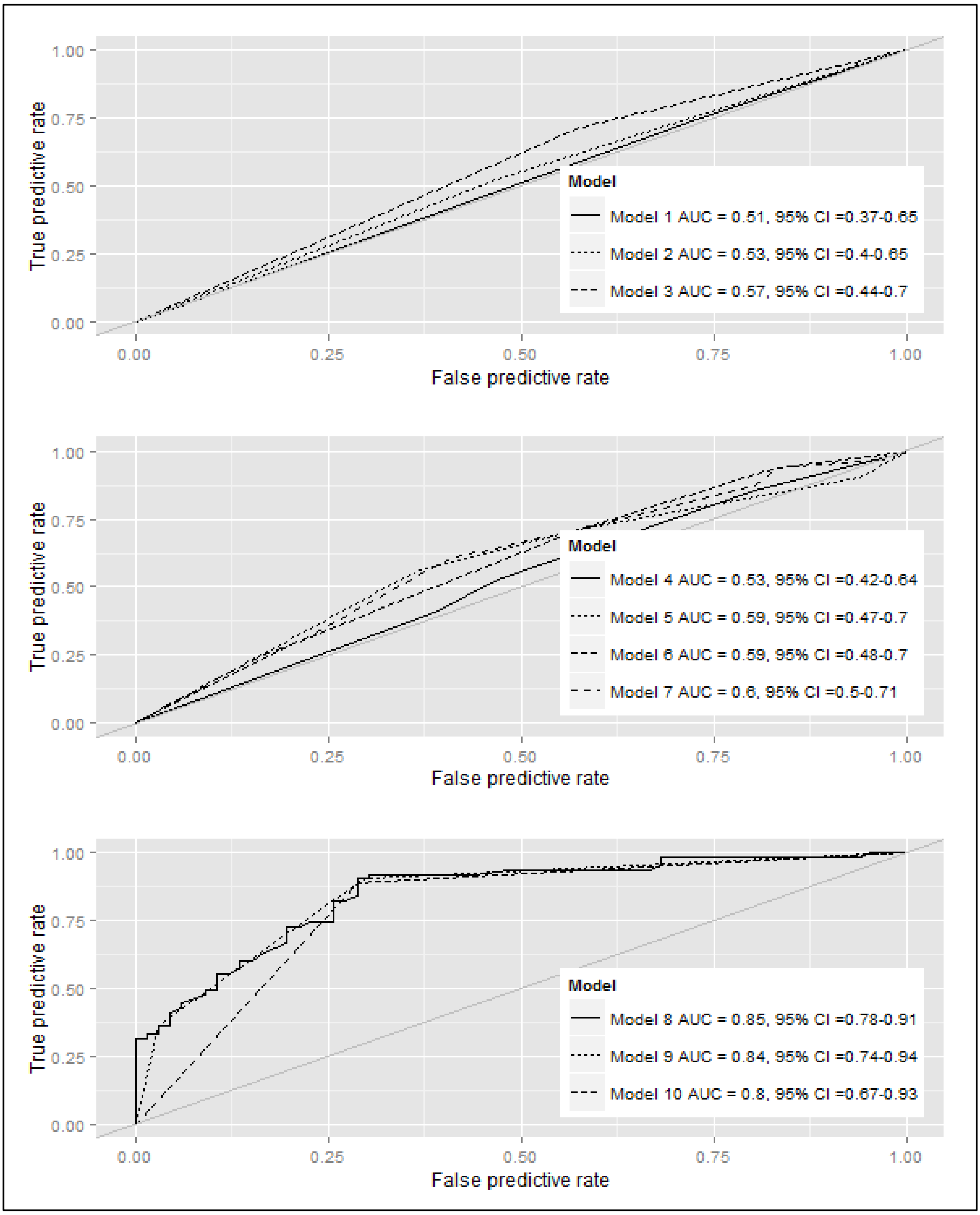


Figure 112: ROC curves and cross validated AUC (with confidence intervals) of models applied to the housing principle welfare measure from the French WQ® data with outcome defined by the median. N = 129.

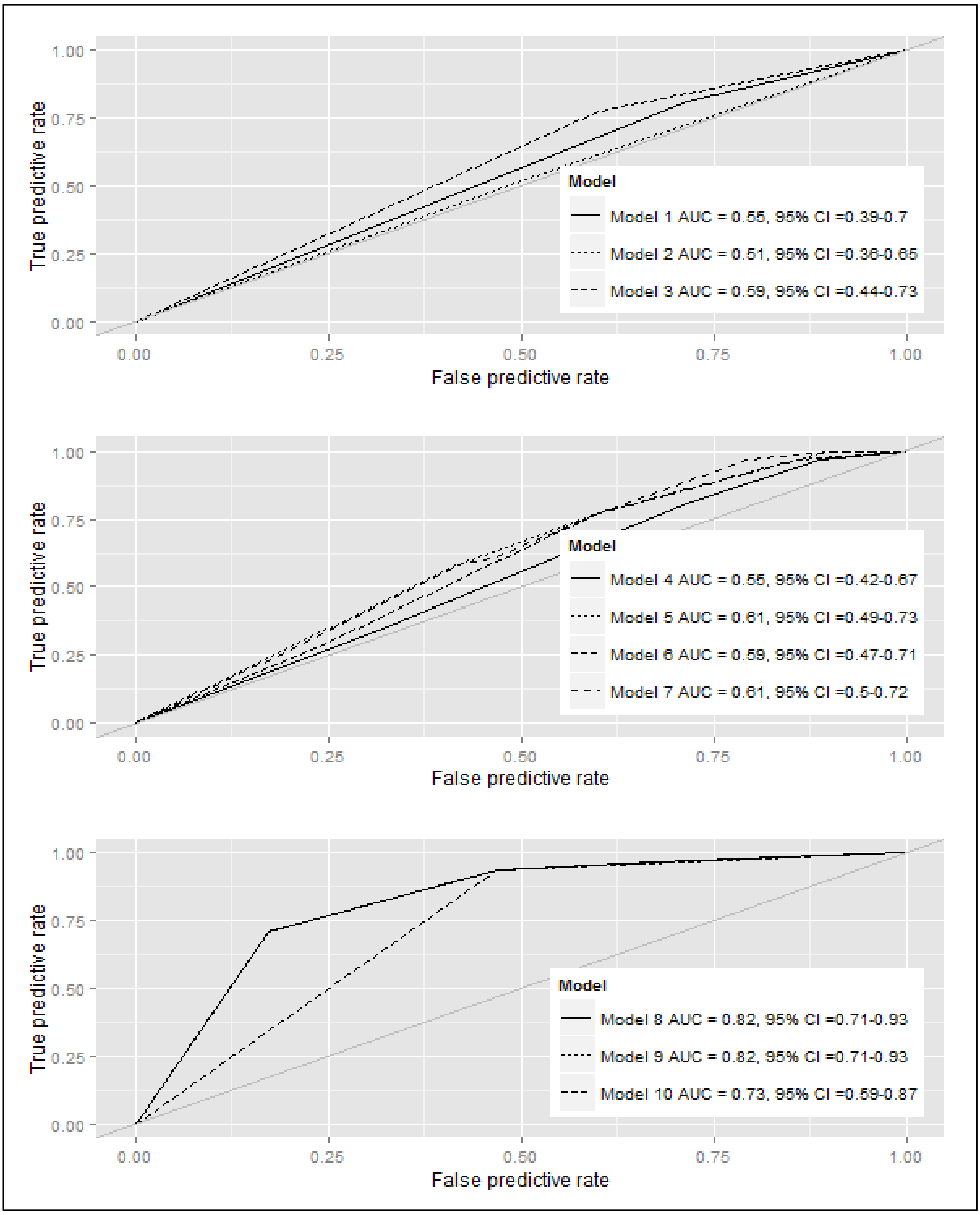


Figure 113: ROC curves and cross validated AUC (with confidence intervals) of models applied to the housing principle welfare measure from the French WQ® data with outcome defined by the 25th percentile. N = 129.

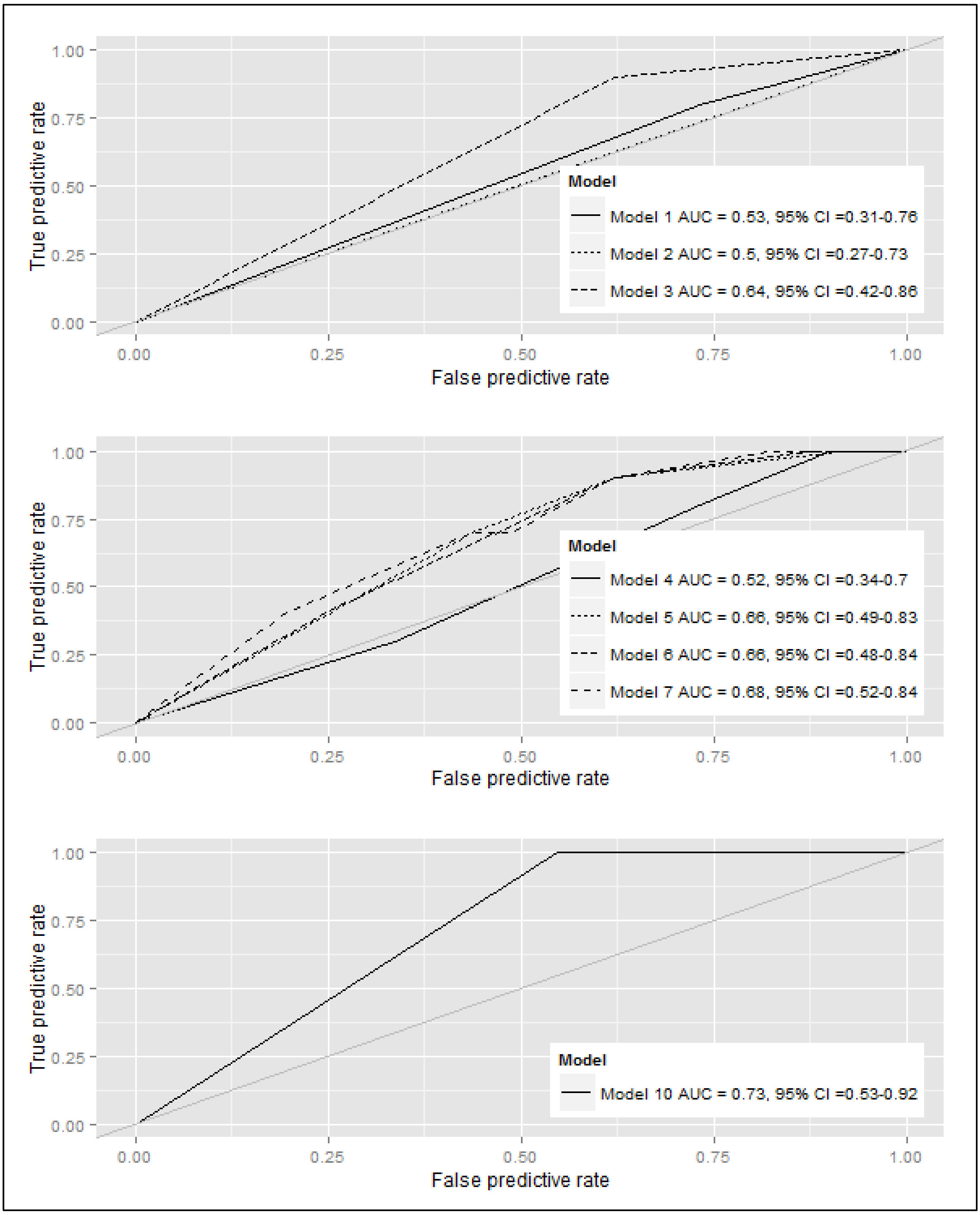


Figure 114: ROC curves and cross validated AUC (with confidence intervals) of models applied to the housing principle welfare measure from the French WQ® data with outcome defined by the 10th percentile. N = 129.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the

Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 76: .

Table 76: Comparing model results from ten different models applied to the WQ® housing principle welfare measure using three different outcomes. French WQ® data, N = 129. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			**Outcome P25**			**Outcome P10**		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	188.4	0.2577	0.5094	150.9	0.1865	0.5461	79.8	0.0735	0.5345
2	SCC	188.1	0.2570	0.5271	152.0	0.1883	0.5081	80.1	0.0738	0.5021
3	Lameness	185.8	0.2524	0.5693	148.8	0.1836	0.5861	76.3	0.0716	0.6391
4	Mortality + SCC	192.9	0.2611	0.5298	155.7	0.1893	0.5474	84.7	0.0745	0.5197
5	Mortality + Lameness	190.5	0.2563	0.5878	152.2	0.1851	0.6104	80.8	0.0728	0.6643
6	SCC + Lameness	189.4	0.2539	0.5907	153.5	0.1865	0.5884	80.9	0.0731	0.6601
7	Mortality + SCC + Lameness	194.2	0.2581	0.6002	157.1	0.1880	0.6129	85.5	0.0740	**0.6836**
8	All factors, additive	137.1	0.1578	**0.8484**	121.1	0.1401	**0.8204**	NA	NA	NA
9	All factors, incl. interactions	132.6	0.1536	**0.8444**	121.1	0.1401	**0.8204**	NA	NA	NA
10	Conditional Inference Tree	NA	NA	**0.8005**	NA	NA	**0.7330**	NA	NA	**0.7269**

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

The combination of all three ABMs had some predictive value when the outcome was defined by the P10. However, the ROC AUCs were still small. When combining the ABMs with the factors of variation, housing was deemed important and herds with straw yards had a lower risk of being in the poor welfare group. Furthermore, breed was of importance with dual purpose herds at a lower risk of poorer welfare than the milk breed herds. Also, it should be noted, that when using the Housing principle as the outcome, increasing herd size was associated with better welfare – not poorer, as when using the Health principle.

11.2.4.3. Feeding principle

In Table 77: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the feeding principle outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P25 and the P10 are shown in Table 78: and Table 79: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; only when using the median as the outcome threshold the conditional tree model found any significant variables. Please see Figure 115: for a detailed presentation of the structure and the nodes in this model.

Table 77: Model formulas for models with outcome defined by the median. French WQ® data, feeding principle, N = 129.

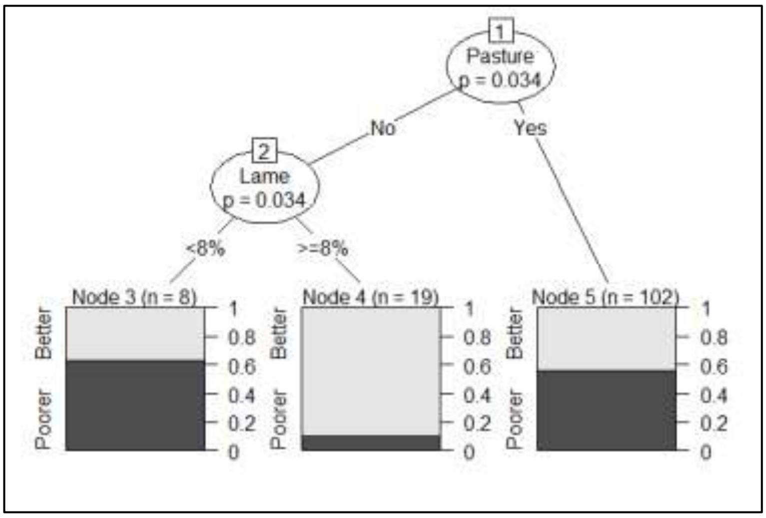
Model	Formula	F-test, P value
Model 1	Mortality	0.6509
Model 2	SCC	0.5371
Model 3	Lameness	0.4230
Model 4	Mortality + SCC	0.7768
Model 5	Mortality + Lameness	0.6344
Model 6	SCC + Lameness	0.4998
Model 7	Mortality + SCC + Lameness	0.6736
Model 8	Pasture + Breed	0.0007
Model 9	Breed + Pasture:Lameness	0.0002
Model 10	Pasture + Lameness	NA

Table 78: Model formulas for models with outcome defined by the 25th percentile. French WQ® data, feeding principle, N = 129.

Model	Formula	F-test, P value
Model 1	Mortality	0.8833
Model 2	SCC	0.3248
Model 3	Lameness	0.4074
Model 4	Mortality + SCC	0.6157
Model 5	Mortality + Lameness	0.6934
Model 6	SCC + Lameness	0.3203
Model 7	Mortality + SCC + Lameness	0.5167
Model 8	Unreliable due to rank deficiency	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	No inner nodes	NA

Table 79: Model formulas for models with outcome defined by the 90th percentile. French WQ® data, feeding principle, N = 129.

Model	Formula	F-test, P value
Model 1	Mortality	0.7744
Model 2	SCC	0.7924
Model 3	Lameness	0.8248
Model 4	Mortality + SCC	0.9130
Model 5	Mortality + Lameness	0.9408
Model 6	SCC + Lameness	0.9227
Model 7	Mortality + SCC + Lameness	0.9663
Model 8	Unreliable due to rank deficiency	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	No inner nodes	NA



Lame: lameness

Figure 115: Conditional inference tree model with outcome defined by the median. French WQ® feeding principle score, N =129.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 116: , Figure 117: and Figure 118: . Only when defining the outcome by the median significant effects were found by the conditional inference tree. Therefore, there were no ROC curves for outcome P75 and P90, model 10. When P75 or P90 was defining the outcome, the reduced models 8 and 9 contained the same variables and therefore the ROC curves were identical. Regardless of which threshold was used for the outcome, none of the models had ROC AUC that were significantly larger than 0.5.

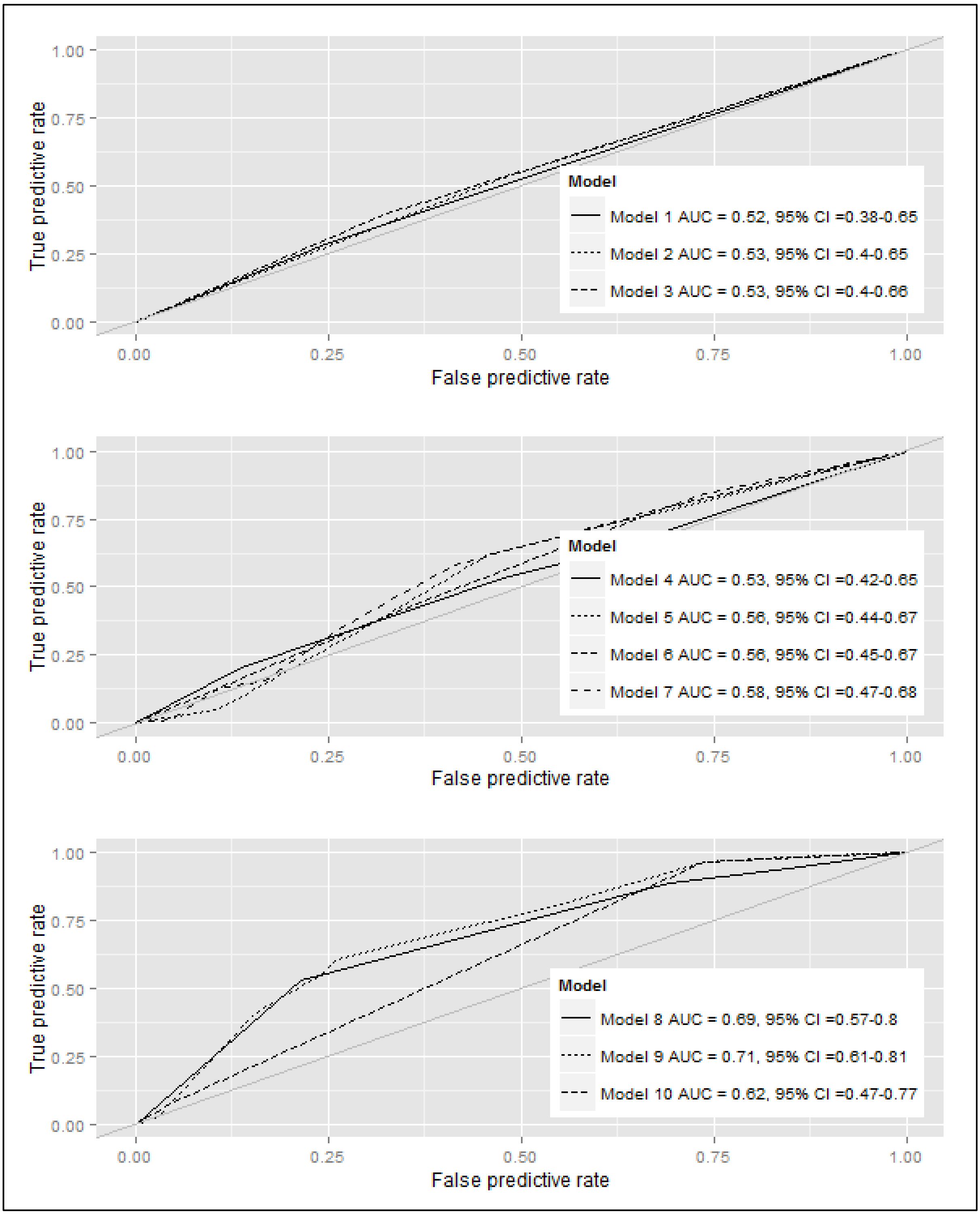


Figure 116: ROC curves and cross validated AUC (with confidence intervals) of models applied to the feeding principle welfare measure from the French WQ® data with outcome defined by the median. N = 128.

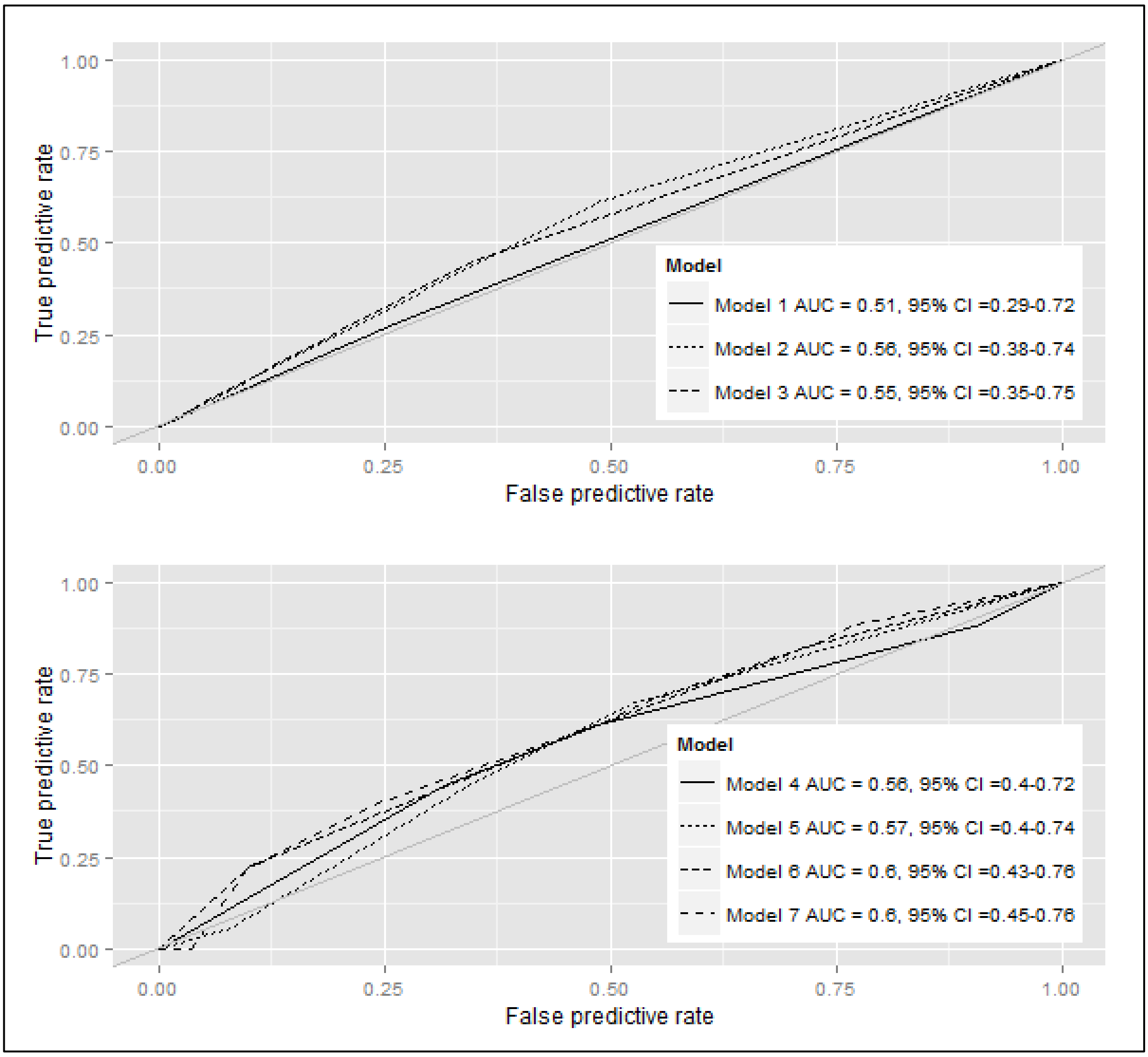


Figure 117: ROC curves and cross validated AUC (with confidence intervals) of models applied to the feeding principle welfare measure from the French WQ® data with outcome defined by the 25th percentile. N = 128.

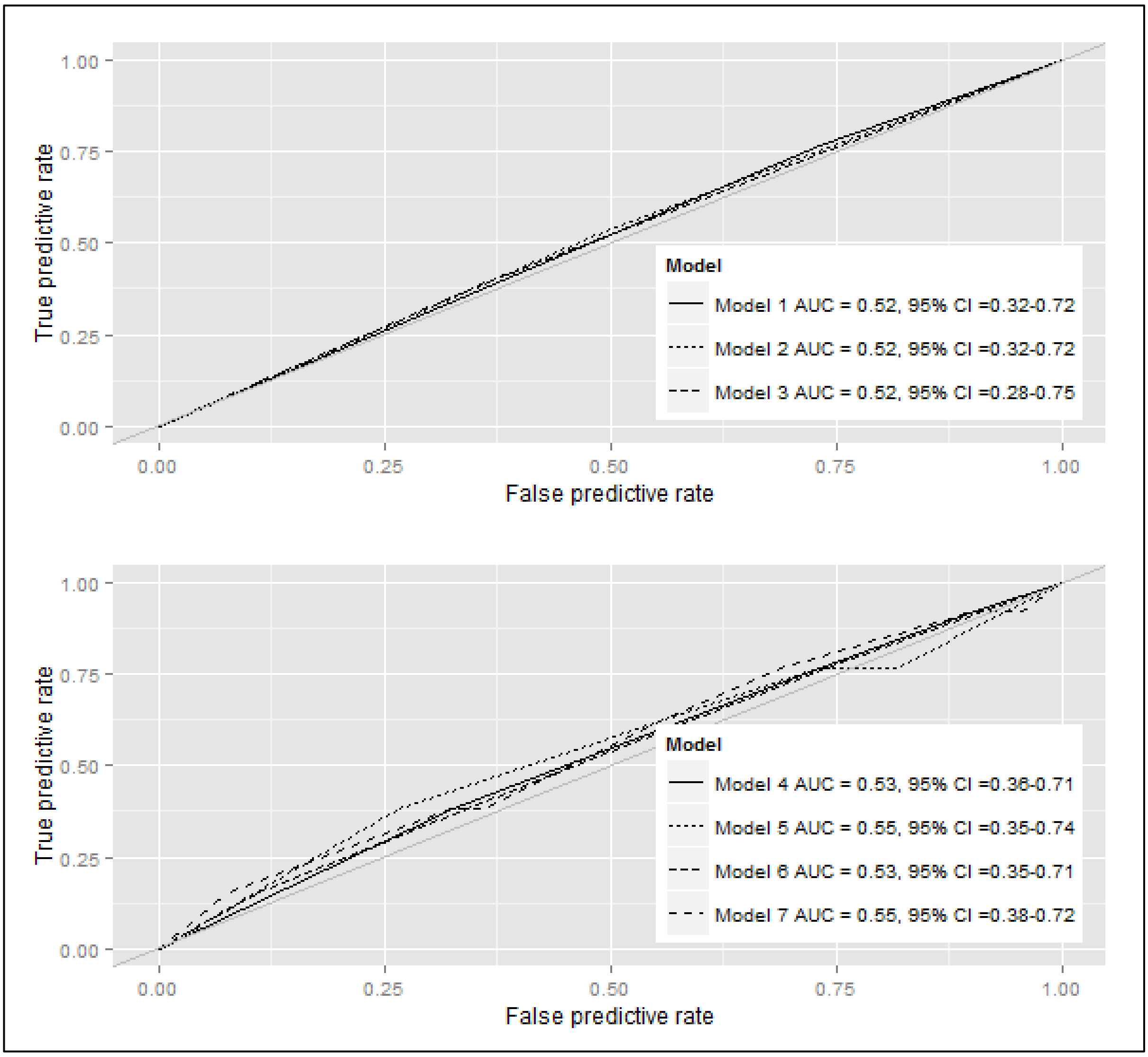


Figure 118: ROC curves and cross validated AUC (with confidence intervals) of models applied to the feeding principle welfare measure from the French WQ® data with outcome defined by the 10th percentile. N = 129.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 80: .

Table 80: Comparing model results from ten different models applied to the WQ® behaviour principle welfare measure using three different outcomes. French WQ® data, N = 128. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	188.3	0.2575	0.5175	114.0	0.1239	0.5083	93.9	0.0933	0.5182
2	SCC	188.2	0.2571	0.5272	113.0	0.1229	0.5623	94.0	0.0934	0.5192
3	Lameness	187.9	0.2566	0.5338	113.3	0.1234	0.5511	94.0	0.0935	0.5156
4	Mortality + SCC	192.9	0.2610	0.5340	117.9	0.1250	0.5631	98.8	0.0947	0.5338
5	Mortality + Lameness	192.5	0.2604	0.5575	118.1	0.1253	0.5656	98.8	0.0947	0.5464
6	SCC + Lameness	192.0	0.2592	0.5579	116.6	0.1238	0.5958	98.7	0.0949	0.5295
7	Mortality + SCC + Lameness	196.7	0.2632	0.5767	121.4	0.1259	0.6034	103.5	0.0960	0.5484
8	All factors, additive	178.8	0.2324	0.6868	NA	NA	NA	NA	NA	NA
9	All factors, incl. interactions	181.0	0.2283	0.7112	NA	NA	NA	NA	NA	NA
10	Conditional Inference Tree	NA	NA	0.6216	NA	NA	NA	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

The analyses on the WQ® feeding principle score were rather inconclusive. Pasture seemed to be associated with the feeding principle score but the predictive value based on the ROC AUC was low and insignificant.

11.2.4.4. Behaviour principle

In Table 81: , the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 82: and Table 83: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 119: and Figure 120: for a detailed presentation of the structure and the nodes in these models. When defining the outcome by the P10 the conditional inference tree model found no significant variables.

Table 81: Model formulas for models with outcome defined by the median. French WQ® data, behaviour principle, N = 128.

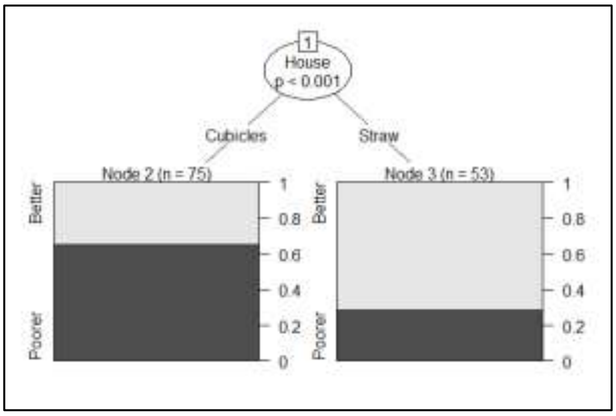
Model	Formula	F-test, P value
Model 1	Mortality	0.3116
Model 2	SCC	0.7237
Model 3	Lameness	0.8531
Model 4	Mortality + SCC	0.5222
Model 5	Mortality + Lameness	0.5754
Model 6	SCC + Lameness	0.9356
Model 7	Mortality + SCC + Lameness	0.7238
Model 8	Herd size + Pasture + Housing system	0.0001
Model 9	Herd size + Pasture + Housing system	0.0001
Model 10	Housing system	NA

Table 82: Model formulas for models with outcome defined by the 25th percentile. French WQ® data, behaviour principle, N = 128.

Model	Formula	F-test, P value
Model 1	Mortality	0.6365
Model 2	SCC	0.1473
Model 3	Lameness	0.0804
Model 4	Mortality + SCC	0.3395
Model 5	Mortality + Lameness	0.2070
Model 6	SCC + Lameness	0.1306
Model 7	Mortality + SCC + Lameness	0.2505
Model 8	Herd size + Housing system + Breed	0.0008
Model 9	Housing system*Herd size + Herd size*Breed	<0.0001
Model 10	Housing system	NA

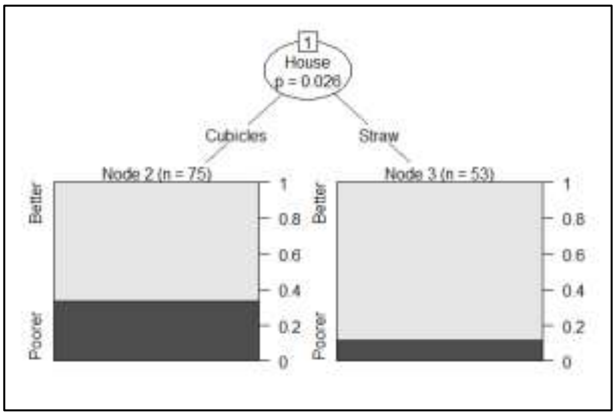
Table 83: Model formulas for models with outcome defined by the 10th percentile. French WQ® data, behaviour principle, N = 128.

Model	Formula	F-test, P value
Model 1	Mortality	0.6698
Model 2	SCC	0.1388
Model 3	Lameness	0.3888
Model 4	Mortality + SCC	0.2641
Model 5	Mortality + Lameness	0.6031
Model 6	SCC + Lameness	0.2988
Model 7	Mortality + SCC + Lameness	0.4054
Model 8	Herd size + Breed	0.0016
Model 9	Did not converge	NA
Model 10	No inner nodes	NA



House: housing system

Figure 119: Conditional inference tree model with outcome defined by the median. French WQ® Behaviour principle score, N =128.



House: housing system

Figure 120: Conditional inference tree model with outcome defined by the 25th percentile. French WQ® Behaviour principle score, N =128.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 121: , Figure 122: and Figure 123: . When using the median as the outcome threshold, models 8 and 9 contained the same variables and thus their ROC curves were identical. With the P10 outcome, no significant variables were detected in the conditional inference tree model and therefore no ROC curve is presented. Also, at this threshold the model 9 did not converge.

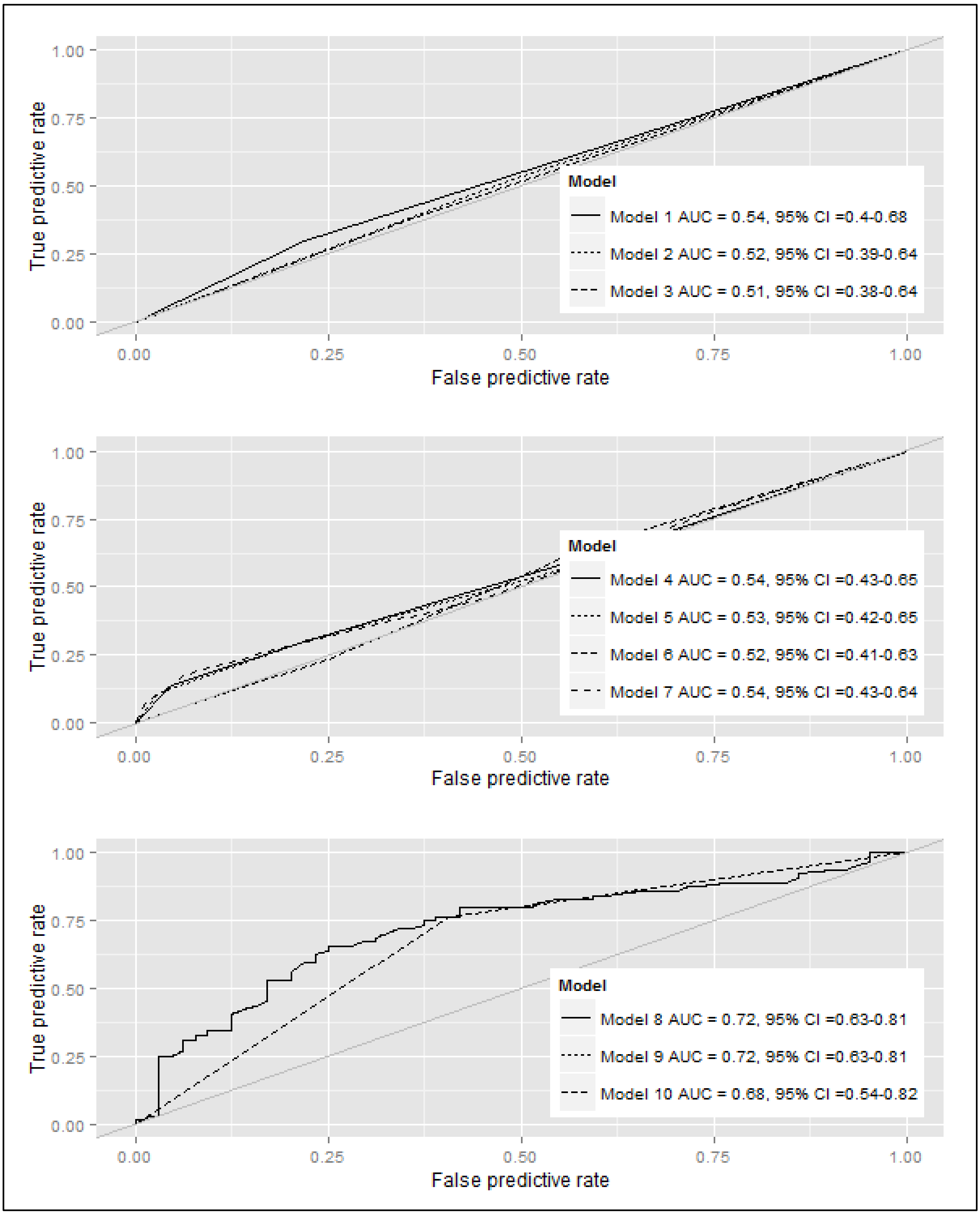


Figure 121: ROC curves and cross validated AUC (with confidence intervals) of models applied to the behaviour principle welfare measure from the French WQ® data with outcome defined by the median. N = 128.

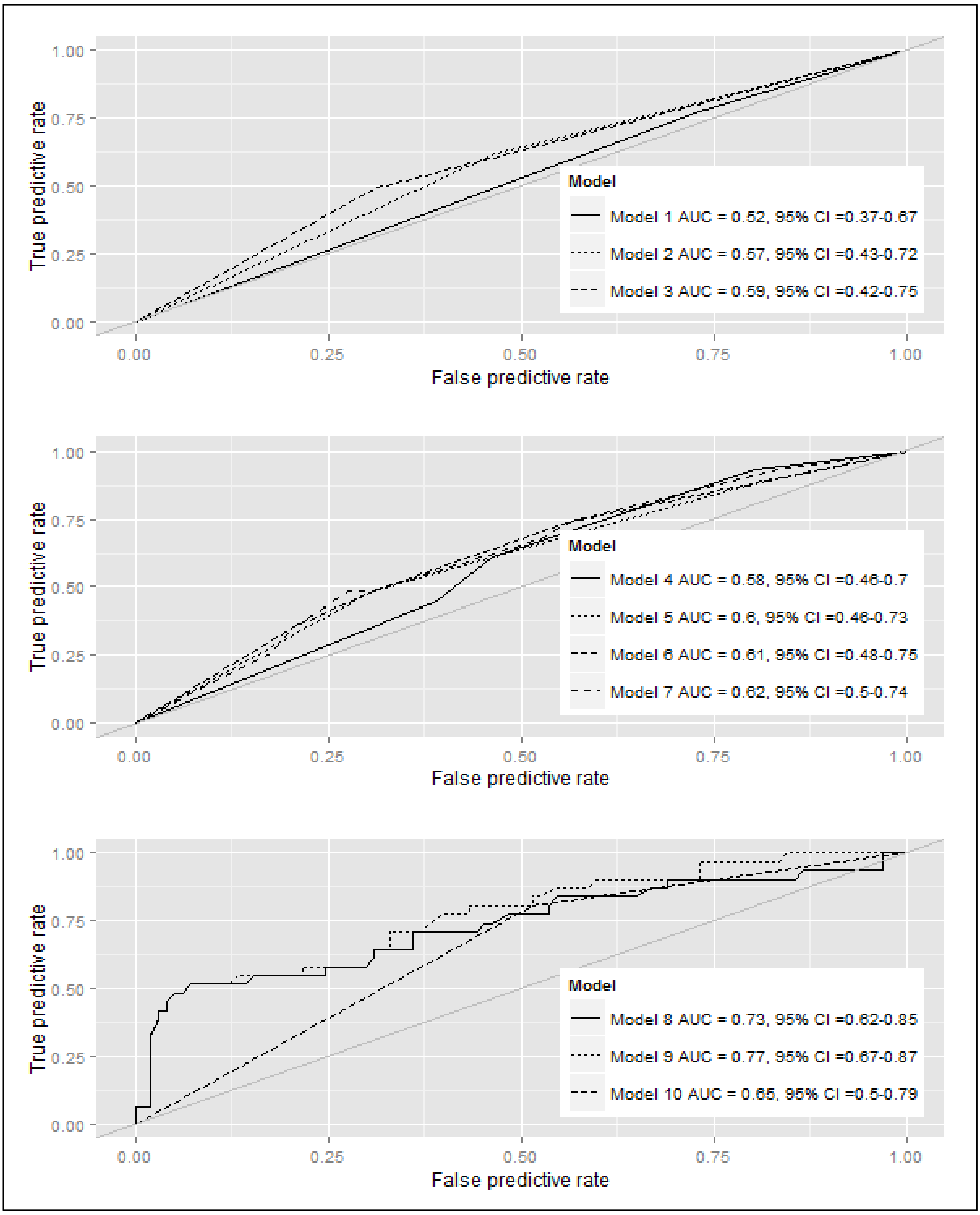


Figure 122: ROC curves and cross validated AUC (with confidence intervals) of models applied to the behaviour principle welfare measure from the French WQ® data with outcome defined by the 75th percentile. N = 128.

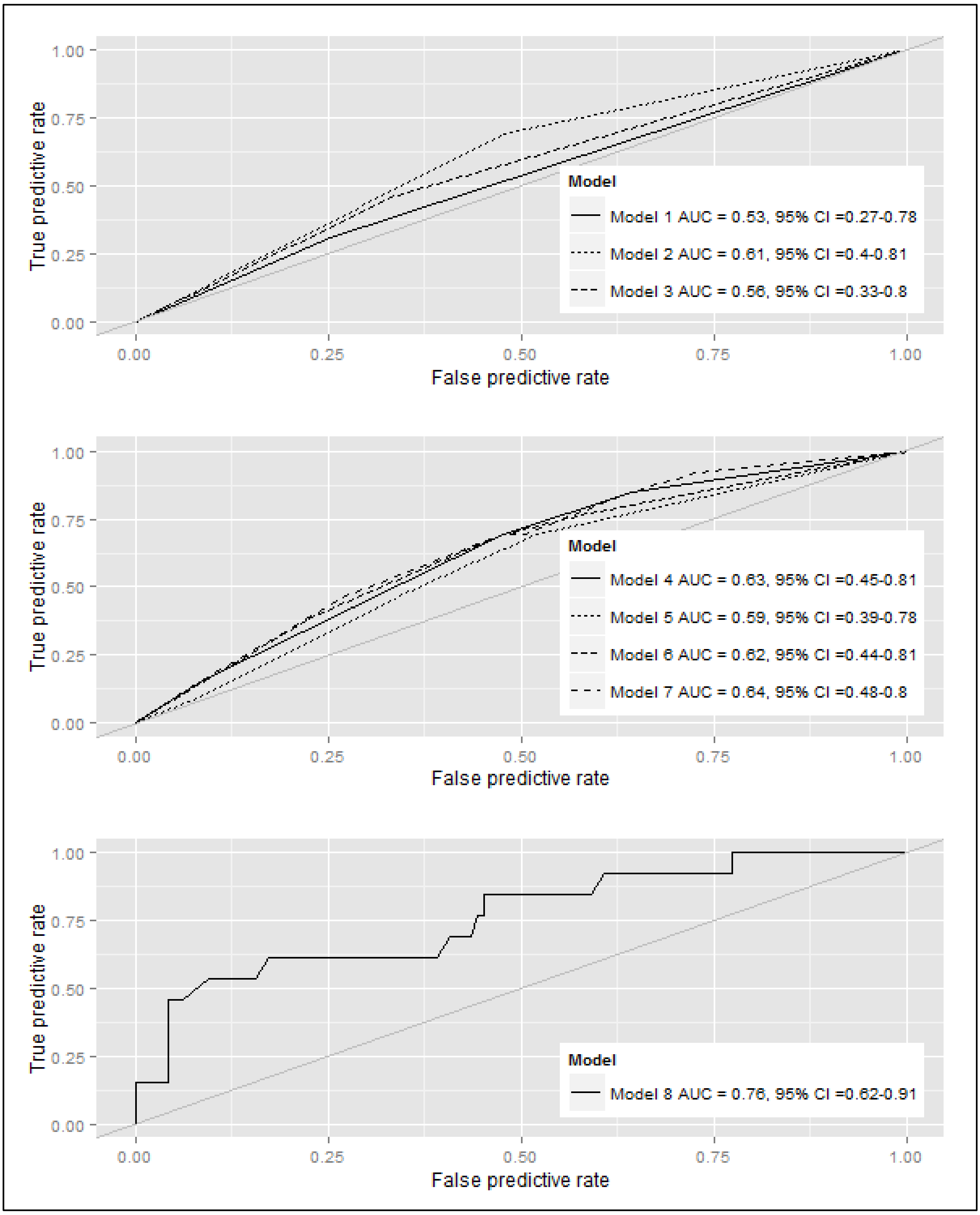


Figure 123: ROC curves and cross validated AUC (with confidence intervals) of models applied to the behaviour principle welfare measure from the French WQ® data with outcome defined by the 90th percentile. N = 129.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the

Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 84: .

Table 84: Comparing model results from ten different models applied to the WQ® behaviour principle welfare measure using three different outcomes. French WQ® data, N = 128. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			**Outcome P25**			**Outcome P10**		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	186.1	0.2559	0.5391	151.2	0.1889	0.5211	93.6	0.0942	0.5278
2	SCC	187.0	0.2577	0.5156	149.3	0.1863	0.5745	91.6	0.0926	0.6070
3	Lameness	187.1	0.2579	0.5078	148.4	0.1851	0.5873	93.1	0.0938	0.5612
4	Mortality + SCC	190.7	0.2594	0.5402	154.1	0.1895	0.5798	96.0	0.0941	0.6284
5	Mortality + Lameness	190.9	0.2598	0.5342	153.1	0.1879	0.5959	97.6	0.0953	0.5853
6	SCC + Lameness	191.9	0.2619	0.5175	152.2	0.1867	0.6149	96.2	0.0941	0.6234
7	Mortality + SCC + Lameness	195.5	0.2635	0.5359	157.0	0.1899	0.6167	100.6	0.0957	0.6445
8	All factors, additive	175.8	0.2239	**0.7218**	144.4	0.1652	**0.7328**	85.8	0.0830	**0.7629**
9	All factors, incl. interactions	175.8	0.2239	**0.7218**	137.3	0.1548	**0.7687**	NA	NA	NA
10	Conditional Inference Tree	NA	NA	0.6797	NA	NA	0.6455	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

None of the ABMs had any significant effect on the risk of the herds to be in the poorer welfare group as defined by the behavioural principle score in the French WQ® data. Increasing herd size was associated with an increased risk of poorer welfare regardless of the threshold of the outcome. In the models defined by the median and the P25, herds with cubicles had higher odds of classifying in the poorer welfare class compared to herds with straw yards. In the median model also pasture access was associated with better welfare, whereas herds with dual purpose breeds were less likely to categorise as poorer welfare herds than herds with milking breeds in the P25 and P10 models.

11.2.4.5. Summarising results on the French Welfare Quality® data

The association between the ABMs and the factors of variation with the four different principles (Health, Housing, Behaviour and Feeding) defined in the Welfare Quality® protocol were analysed. Regardless of how the outcome was defined, none of the ABMs were associated with any of the Welfare Quality® principles in the French Welfare Quality® data. Instead, the factors of variation increased herd size, milk breed and no access to pasture seemed to be correlated to the poorer welfare status as defined by the Welfare Quality® principles.

11.2.5. **The Belgian Welfare Quality® data**

Results from statistical analyses of the four different principle scores in the Belgian WQ® data are presented individually before a summation of results is given.

.

11.2.5.1. Health principle

In Table 85: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P25 and the P10 are shown in Table 86: and Table 87: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 124: and Figure 125: for a detailed presentation of the structure and the nodes in these models. The conditional inference tree model did not identify any significant variables when outcome was defined by the P10.

Table 85: Model formulas for models with outcome defined by the median. Belgian WQ® data, health principle, N = 63.

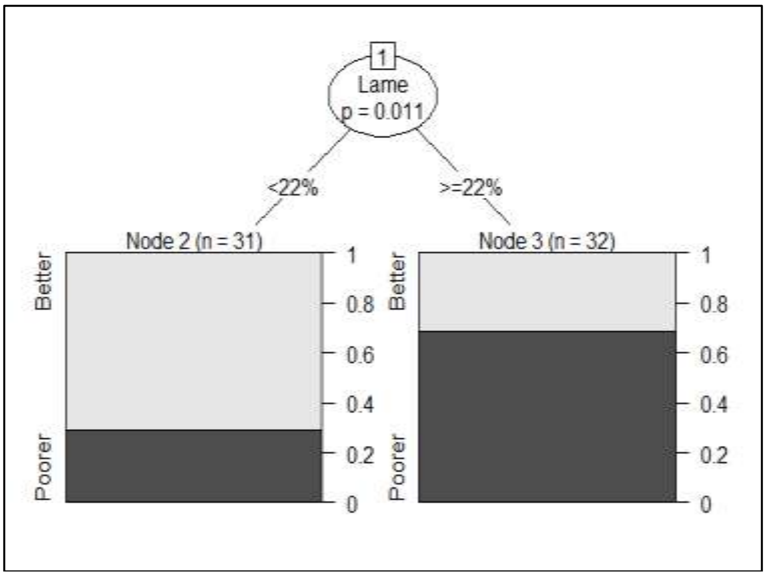
Model	Formula	F-test, P value
Model 1	Mortality	0.0034
Model 2	SCC	0.0578
Model 3	Lameness	0.0014
Model 4	Mortality + SCC	0.0023
Model 5	Mortality + Lameness	0.0003
Model 6	SCC + Lameness	0.0024
Model 7	Mortality + SCC + Lameness	0.0004
Model 8	Mortality + Pasture	0.0002
Model 9	Mortality + Lameness + Pasture	<0.0001
Model 10	Lameness	NA

Table 86: Model formulas for models with outcome defined by the 25th percentile. Belgian WQ® data, health principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.0023
Model 2	SCC	0.0678
Model 3	Lameness	0.0224
Model 4	Mortality + SCC	0.0018
Model 5	Mortality + Lameness	0.0018
Model 6	SCC + Lameness	0.0262
Model 7	Mortality + SCC + Lameness	0.0018
Model 8	Mortality	0.0023
Model 9	Mortality	0.0023
Model 10	Mortality	NA

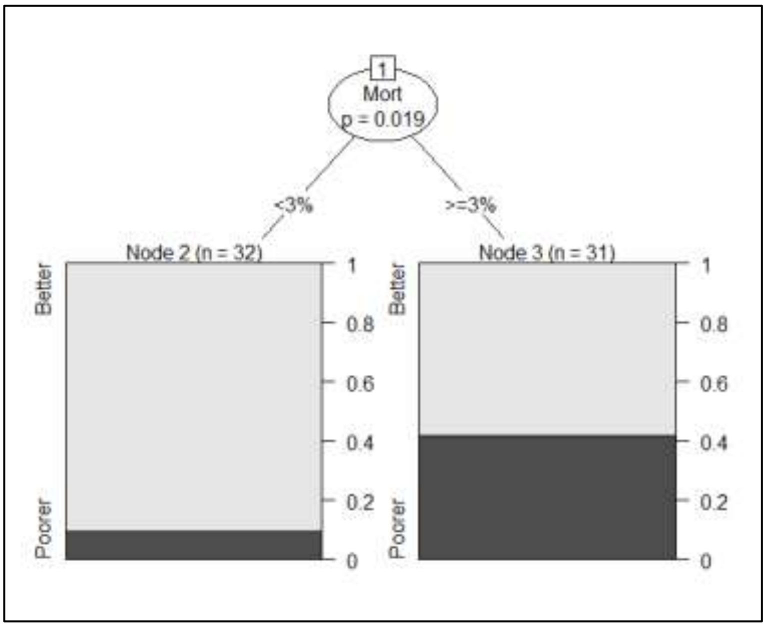
Table 87: Model formulas for models with outcome defined by the 10th percentile. Belgian WQ® data, behaviour principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.0322
Model 2	SCC	0.2061
Model 3	Lameness	0.7211
Model 4	Mortality + SCC	0.0485
Model 5	Mortality + Lameness	0.1005
Model 6	SCC + Lameness	0.4473
Model 7	Mortality + SCC + Lameness	0.1053
Model 8	Mortality	0.0322
Model 9	Mortality:Herd size	0.0421
Model 10	No inner nodes	NA



Lame: lameness

Figure 124: Conditional inference tree model with outcome defined by the median. Belgian WQ® Health principle score, N =63.



Mort: mortality

Figure 125: Conditional inference tree model with outcome defined by the P25. Belgian WQ® Health principle score, N =63.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 126: ,Figure 127: and Figure 128: . For the P25 outcome, models 8, 9 and 10 were identical and thus also had identical ROC curves.

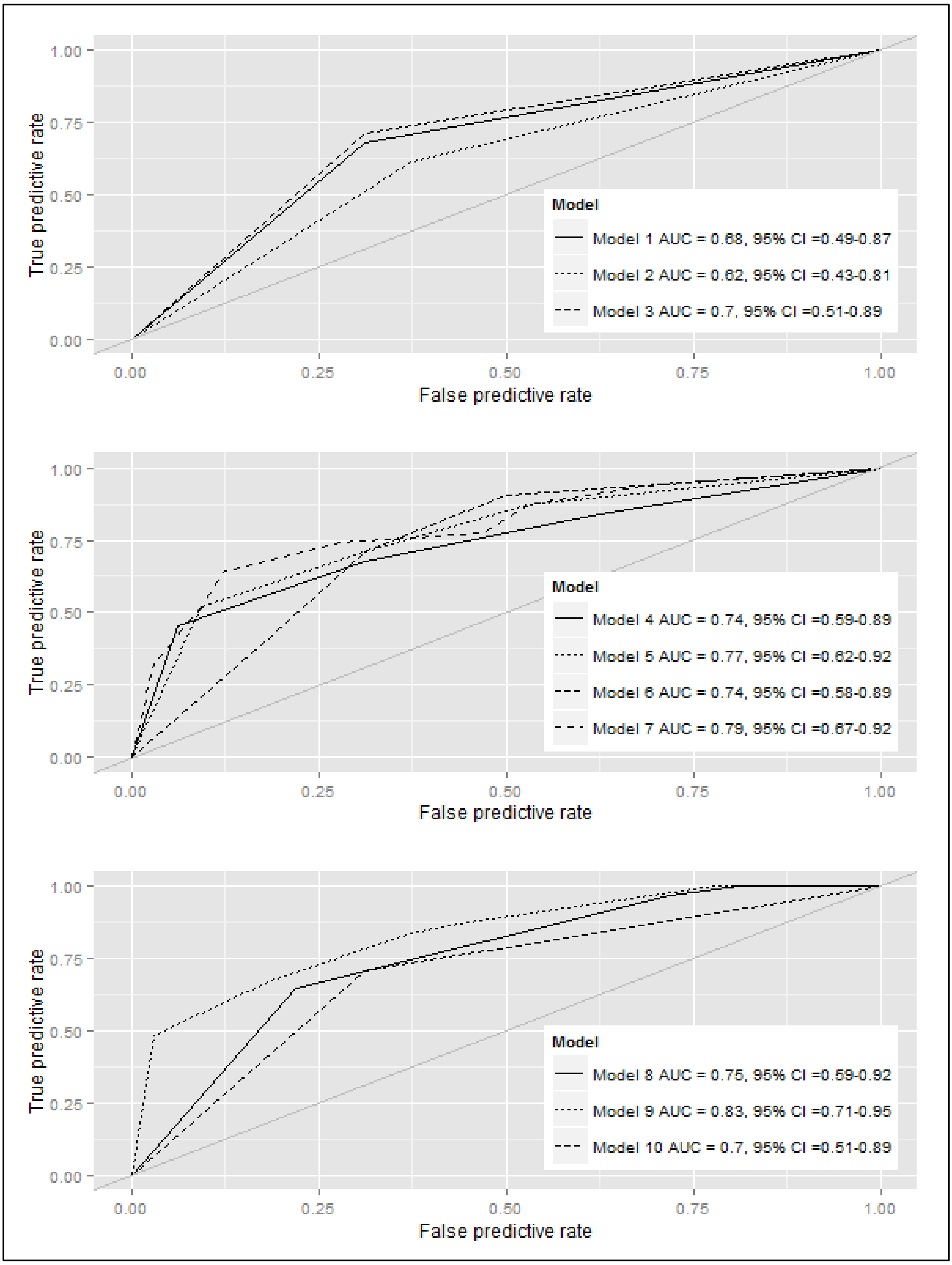


Figure 126: ROC curves and cross validated AUC (with confidence intervals) of models applied to the health principle welfare measure from the Belgian WQ® data with outcome defined by the median. N = 63.

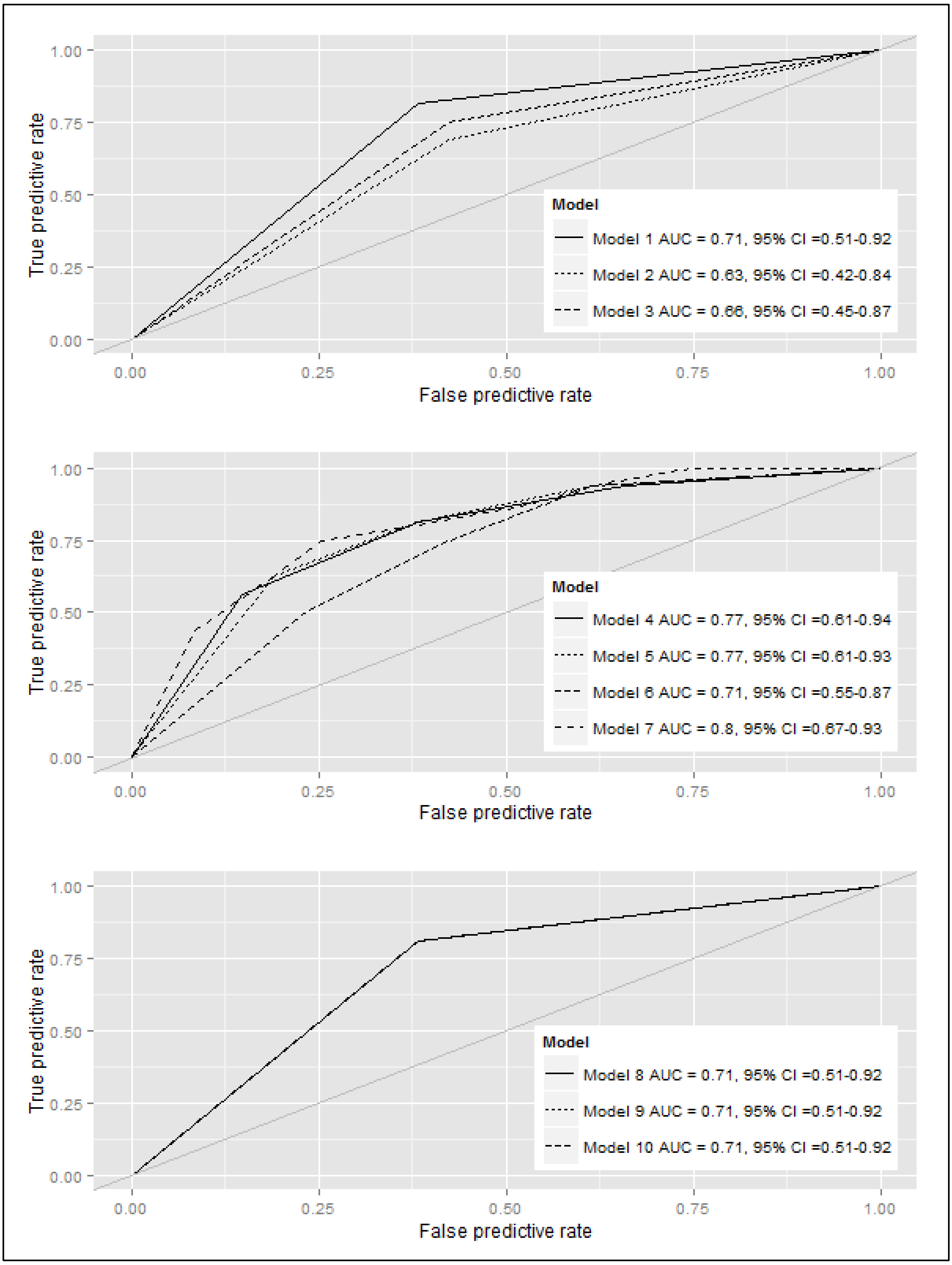


Figure 127: ROC curves and cross validated AUC (with confidence intervals) of models applied to the health principle welfare measure from the Belgian WQ® data with outcome defined by the 25th percentile. N = 63.

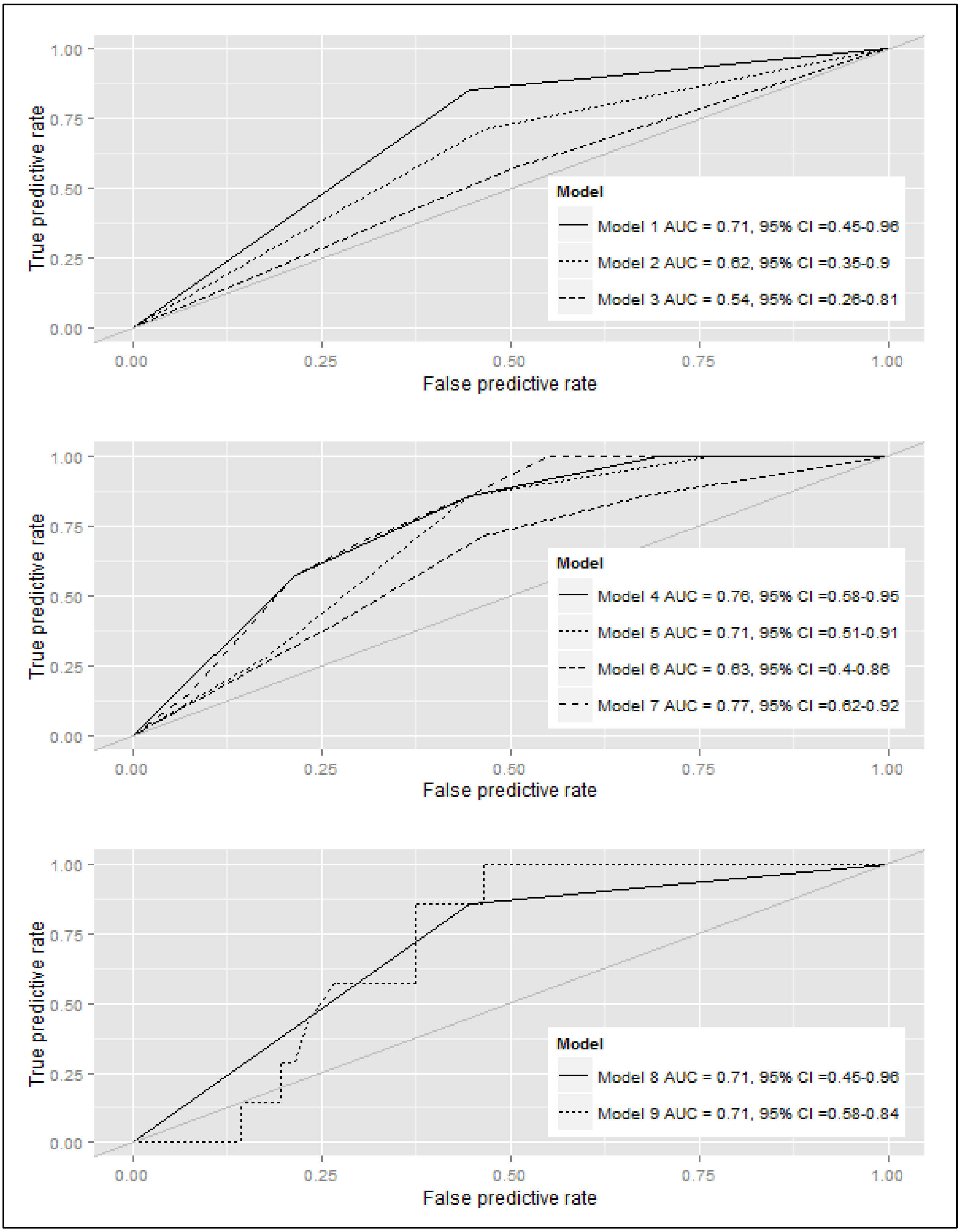


Figure 128: ROC curves and cross validated AUC (with confidence intervals) of models applied to the health principle welfare measure from the Belgian WQ® data with outcome defined by the 10th percentile. N = 63.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 88: .

Table 88: Comparing model results from ten different models applied to the WQ® health principle welfare measure using three different outcomes. Belgian WQ® data, N = 63. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	87.0	0.3175	0.6825	70.4	0.1738	**0.7148**	47.6	0.0983	0.7054
2	SCC	92.0	0.2514	0.6190	76.4	0.1915	0.6310	50.6	0.1027	0.6250
3	Lameness	85.4	0.2244	**0.6986**	74.5	0.1858	0.6622	52.1	0.1051	0.5357
4	Mortality + SCC	87.6	0.2254	**0.7369**	71.2	0.1687	**0.7746**	50.3	0.1014	**0.7628**
5	Mortality + Lameness	83.3	0.2099	**0.7697**	71.2	0.1698	**0.7733**	51.8	0.1030	**0.7092**
6	SCC + Lameness	87.7	0.2258	**0.7369**	76.5	0.1888	**0.7108**	54.8	0.1064	0.6301
7	Mortality + SCC + Lameness	85.4	0.2105	**0.7949**	72.9	0.1696	**0.7992**	54.4	0.1068	**0.7691**
8	All factors, additive	83.1	0.2124	**0.7535**	70.4	0.1738	**0.7148**	47.6	0.0983	0.7054
9	All factors, incl. interactions	77.6	0.1920	**0.8342**	70.4	0.1738	**0.7148**	52.3	0.1183	**0.7105**
10	Conditional Inference Tree	NA	NA	**0.6986**	NA	NA	**0.7148**	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

In the Belgian WQ® data, all the ABMs seemed to be correlated to the risk of the herds to be classified in the poor welfare group according to the health principle, whereas the effect of the factors of variation were minimal. Also, based on the ROC AUC it seemed as if the ABMs had a reasonable predictive value. For mortality and lameness, the risk of being classified as in the poor welfare group increased with the high level of the ABM. Pasture showed a significant effect in the logistic regressions with the threshold at the median. Surprisingly, access to pasture was associated with an increased risk of being classified in the poor welfare group. However, due to small sample size caution should be taken in the interpretation of this (see Table 44:).

11.2.5.2. Housing principle

In Table 89: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P25 and the P10 are shown in Table 90: and Table 91: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 129: , Figure 130: and Figure 131: for a detailed presentation of the structure and the nodes in these models.

Table 89: Model formulas for models with outcome defined by the median. Belgian WQ® data, housing principle, N = 63.

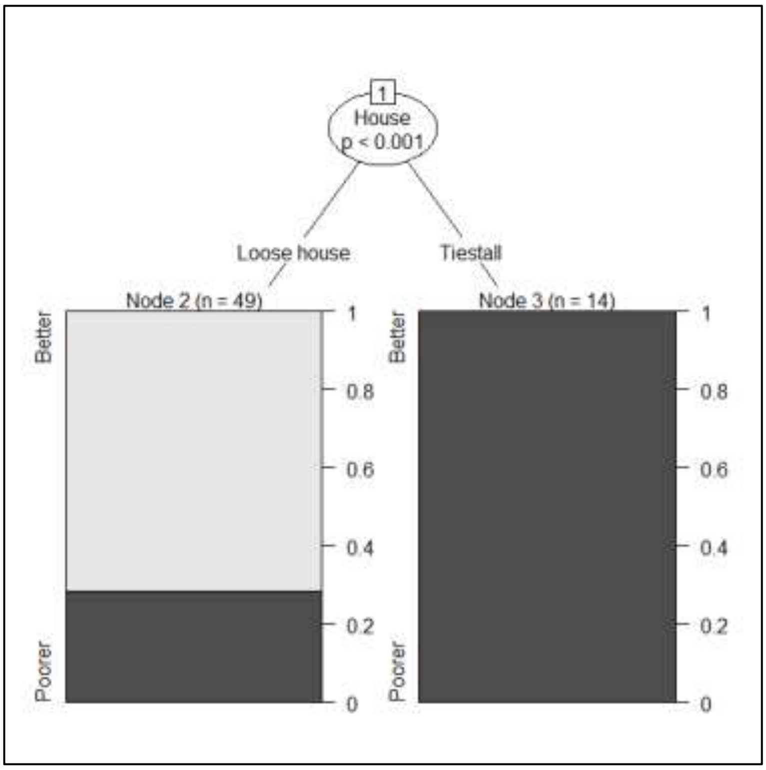
Model	Formula	F-test, P value
Model 1	Mortality	0.3667
Model 2	SCC	0.2590
Model 3	Lameness	0.3667
Model 4	Mortality + SCC	0.3310
Model 5	Mortality + Lameness	0.3537
Model 6	SCC + Lameness	0.4168
Model 7	Mortality + SCC + Lameness	0.3852
Model 8	Unreliable due to rank deficiency	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	Housing system	NA

Table 90: Model formulas for models with outcome defined by the 25th percentile. Belgian WQ® data, housing principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.4125
Model 2	SCC	0.0050
Model 3	Lameness	0.7141
Model 4	Mortality + SCC	0.0117
Model 5	Mortality + Lameness	0.7002
Model 6	SCC + Lameness	0.0117
Model 7	Mortality + SCC + Lameness	0.0220
Model 8	Unreliable due to rank deficiency	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	Housing system	NA

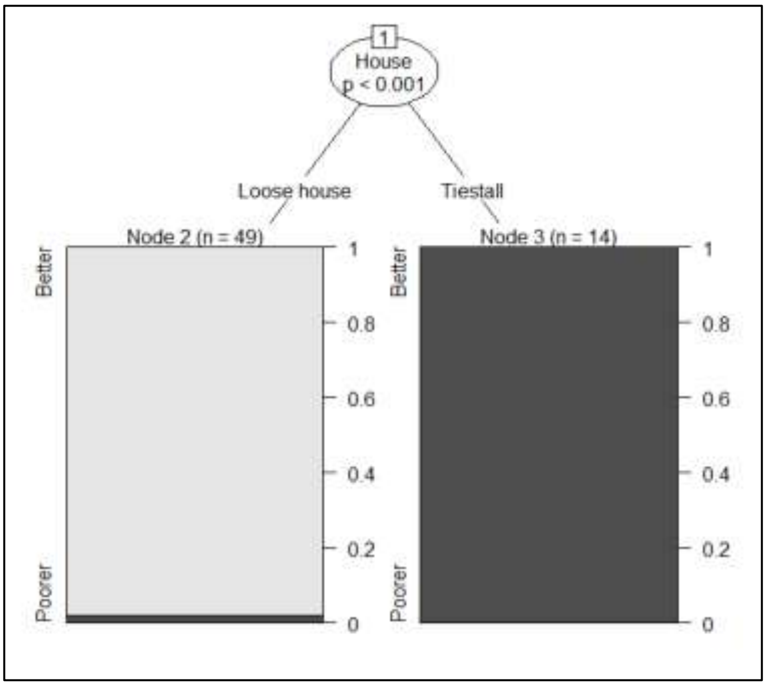
Table 91: Model formulas for models with outcome defined by the 10th percentile. Belgian WQ® data, housing principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.4092
Model 2	SCC	0.3647
Model 3	Lameness	0.3647
Model 4	Mortality + SCC	0.4524
Model 5	Mortality + Lameness	0.5343
Model 6	SCC + Lameness	0.3528
Model 7	Mortality + SCC + Lameness	0.4653
Model 8	Unreliable due to rank deficiency	NA
Model 9	Did not converge	NA
Model 10	Housing system	NA



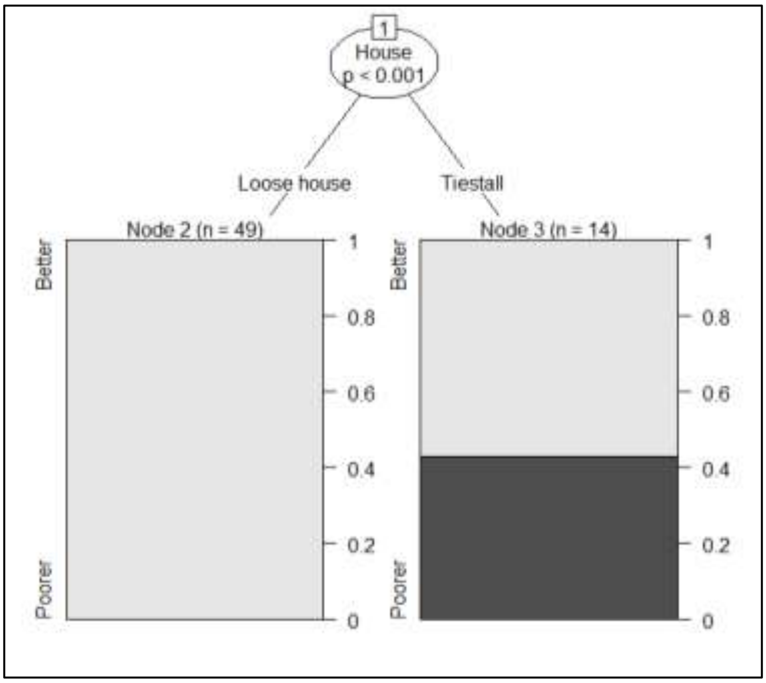
House: housing system

Figure 129: Conditional inference tree model with outcome defined by the median. Belgian WQ® Housing principle score, N =63.



House: housing system

Figure 130: Conditional inference tree model with outcome defined by the P25. Belgian WQ® housing principle score, N =63.



House: housing system

Figure 131: Conditional inference tree model with outcome defined by the P10. Belgian WQ® housing principle score, N =63.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 132: , Figure 133: and Figure 134: . When defining outcome by P10, the logistic regression including interactions (model 9) did not converge and therefore no ROC curve is presented.

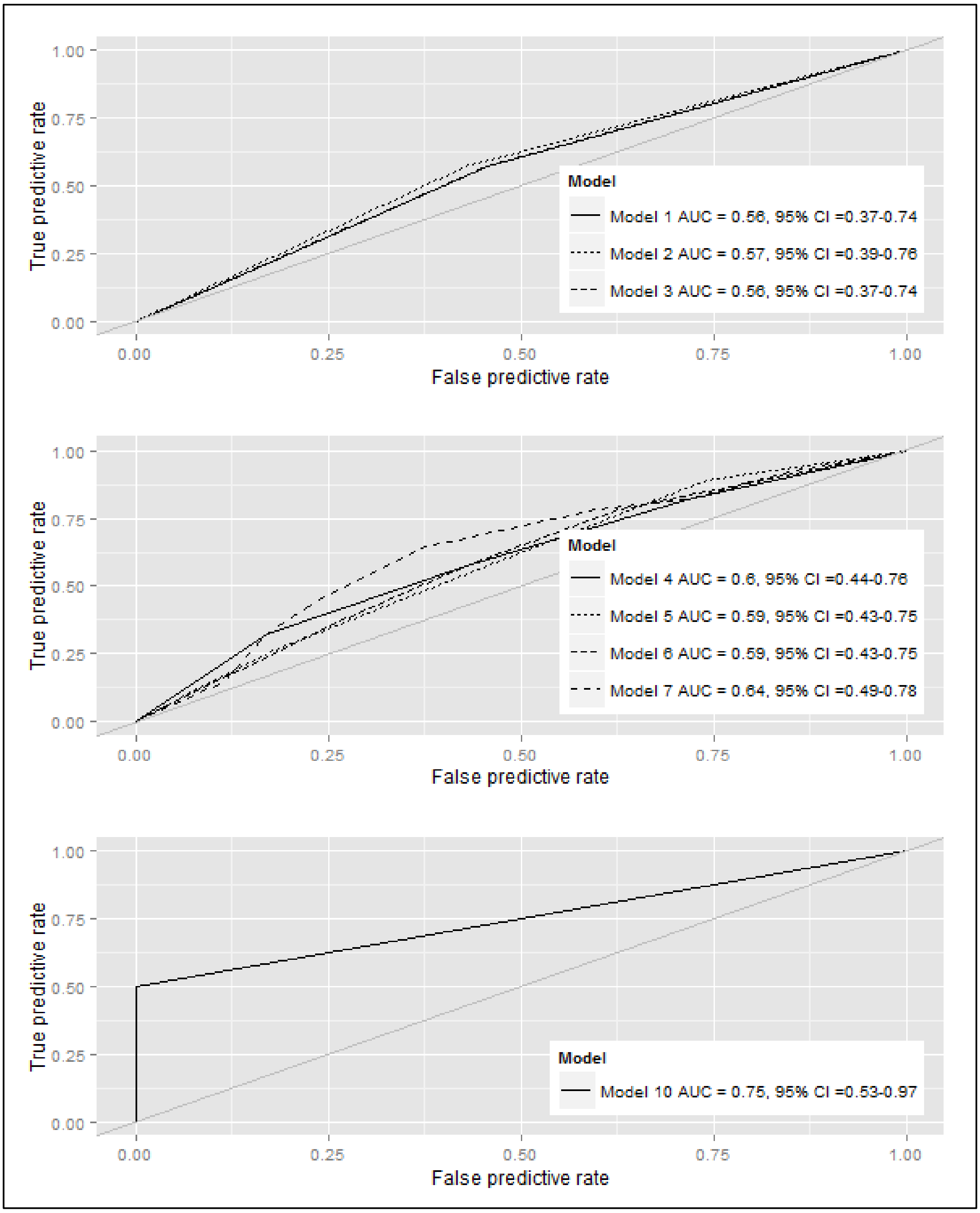


Figure 132: ROC curves and cross validated AUC (with confidence intervals) of models applied to the housing principle welfare measure from the Belgian WQ® data with outcome defined by the median. N = 63.

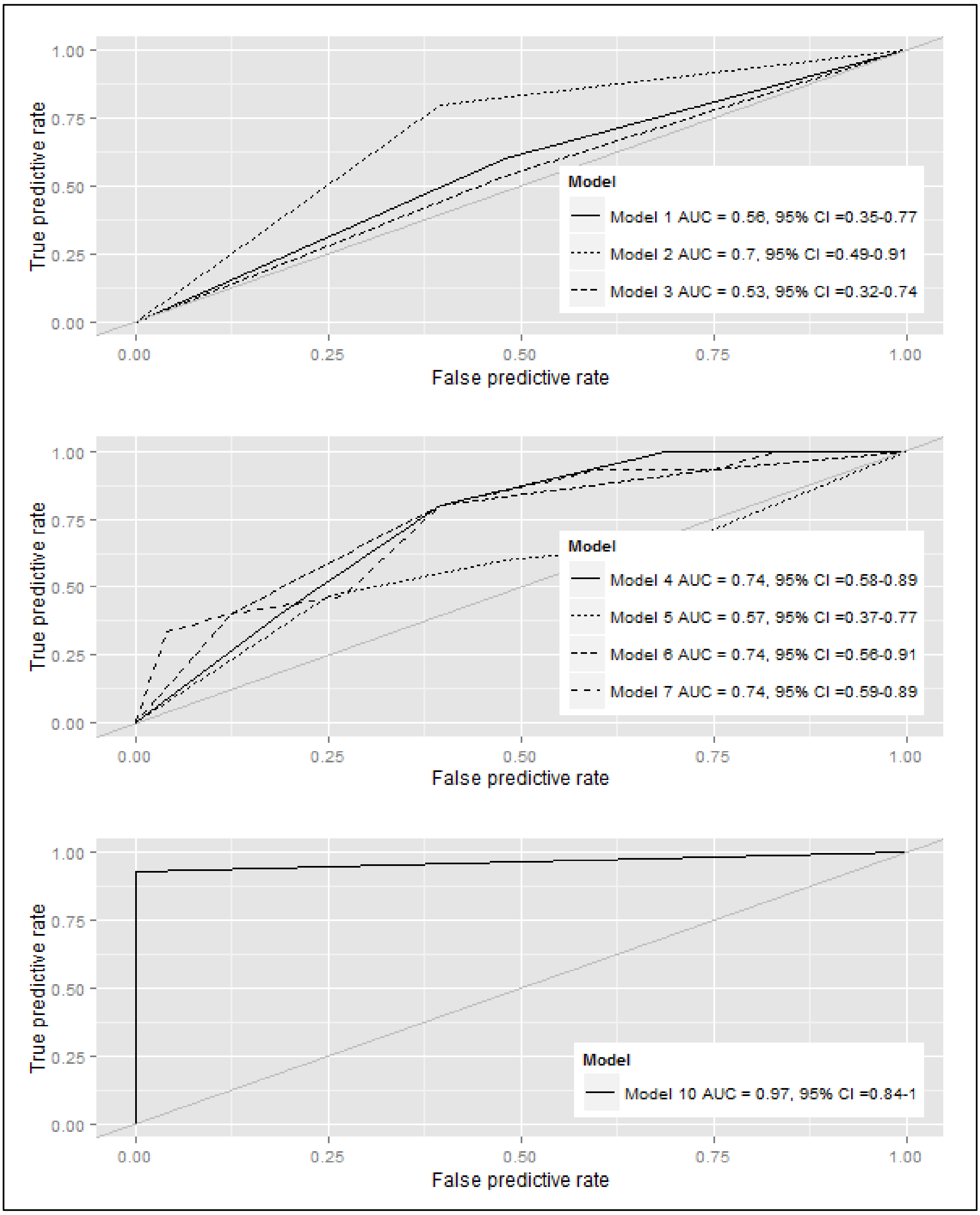


Figure 133: ROC curves and cross validated AUC (with confidence intervals) of models applied to the housing principle welfare measure from the Belgian WQ® data with outcome defined by the 25th percentile. N = 63.

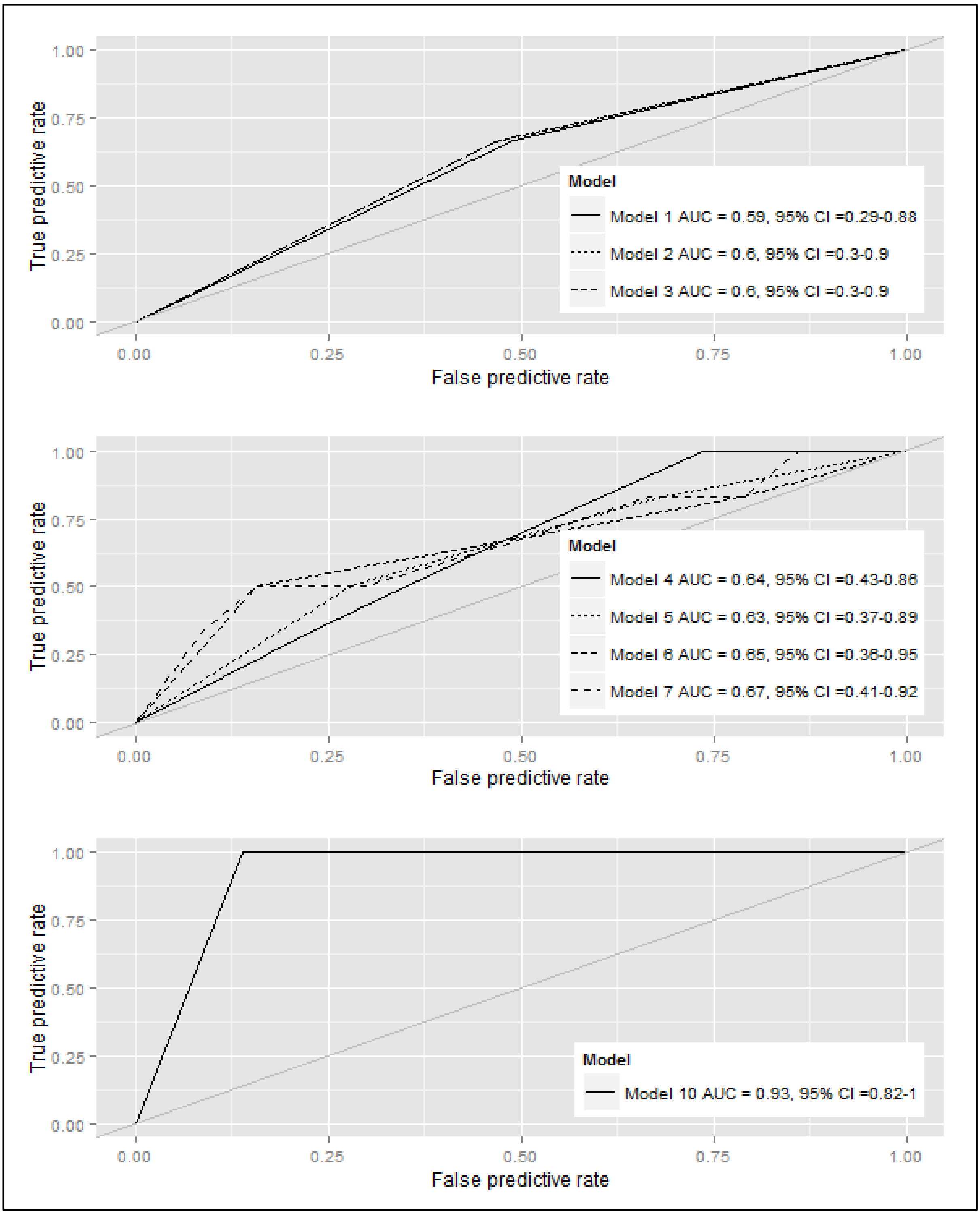


Figure 134: ROC curves and cross validated AUC (with confidence intervals) of models applied to the housing principle welfare measure from the Belgian WQ® data with outcome defined by the 10th percentile. N = 63.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the

Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 92: .

Table 92: Comparing model results from ten different models applied to the WQ® housing principle welfare measure using three different outcomes. Belgian WQ® data, N = 63. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	94.0	0.2598	0.5571	76.8	0.1913	0.5604	47.231	0.0909	0.5877
2	SCC	93.6	0.2579	0.5714	69.6	0.1706	0.7021	47.091	0.0907	0.5965
3	Lameness	94.0	0.2598	0.5571	77.3	0.1930	0.5271	47.091	0.0907	0.5965
4	Mortality + SCC	96.8	0.2625	0.6015	72.7	0.1780	**0.7375**	50.469	0.0938	0.6433
5	Mortality + Lameness	96.9	0.2632	0.5913	80.9	0.1965	0.5688	50.802	0.0926	0.6301
6	SCC + Lameness	97.2	0.2648	0.5929	72.7	0.1736	0.7361	49.972	0.0913	0.6535
7	Mortality + SCC + Lameness	100.0	0.2676	0.6372	76.1	0.1804	**0.7438**	53.643	0.0943	0.6652
8	All factors, additive	NA	NA	NA	NA	NA	NA	NA	NA	NA
9	All factors, incl. interactions	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	Conditional Inference Tree	NA	NA	0.7500	NA	NA	**0.9667**	NA	NA	**0.9298**

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

In the logistic regressions, the mortality and SCC seemed to some extend to be associated with the risk of being in the poorest welfare group based on the housing principle with the threshold at the P25. In the conditional inference tree models, housing was deemed significant. However, caution should be taken due to small number of observations.

11.2.5.3. Feeding principle

When the outcome was defined by the median neither any ABMs nor factors of variation were significantly associated with the feeding principle score in the Belgian WQ® data. Likewise, none of the ROC AUC significantly differed from 0.5.

In Table 93: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 94: and Table 95: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed. Only when using the P25 for the definition of the outcome, any significant nodes were detected; please see Figure 135: for a detailed presentation of the structure and the nodes in this model.

Table 93: Model formulas for models with outcome defined by the median. Belgian WQ® data, feeding principle, N = 63.

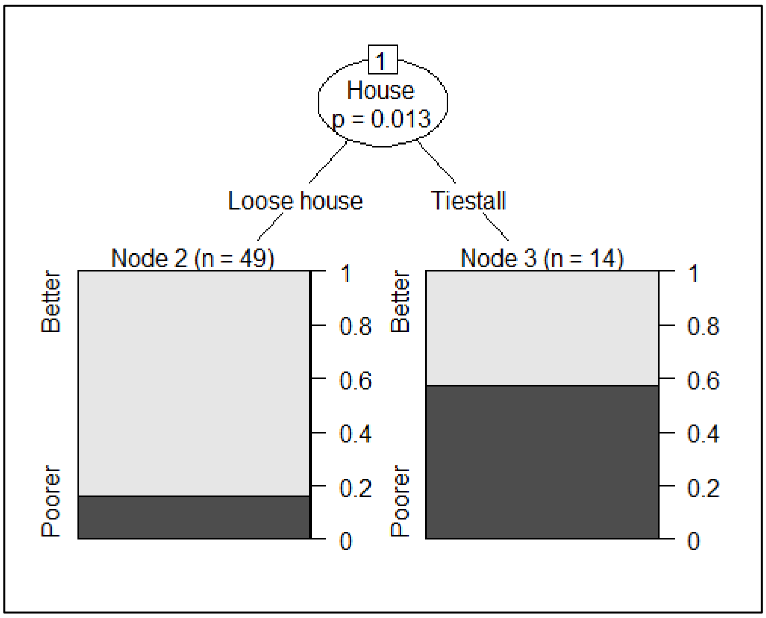
Model	Formula	F-test, P value
Model 1	Mortality	0.3783
Model 2	SCC	0.5271
Model 3	Lameness	0.2551
Model 4	Mortality + SCC	0.5385
Model 5	Mortality + Lameness	0.4192
Model 6	SCC + Lameness	0.3494
Model 7	Mortality + SCC + Lameness	0.4650
Model 8	No significant variables	NA
Model 9	No significant variables	NA
Model 10	No inner nodes	NA

Table 94: Model formulas for models with outcome defined by the 75th percentile. Belgian WQ® data, feeding principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.5138
Model 2	SCC	0.5138
Model 3	Lameness	0.5138
Model 4	Mortality + SCC	0.6647
Model 5	Mortality + Lameness	0.5824
Model 6	SCC + Lameness	0.5824
Model 7	Mortality + SCC + Lameness	0.6305
Model 8	Housing	0.0032
Model 9	Herd size*Housing	0.0002
Model 10	Housing	NA

Table 95: Model formulas for models with outcome defined by the 90th percentile. Belgian WQ® data, feeding principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.65560
Model 2	SCC	0.03215
Model 3	Lameness	0.72110
Model 4	Mortality + SCC	0.09436
Model 5	Mortality + Lameness	0.87260
Model 6	SCC + Lameness	0.10050
Model 7	Mortality + SCC + Lameness	0.19190
Model 8	No significant variables	NA
Model 9	Did not converge	NA
Model 10	No inner nodes	NA



House: housing system

Figure 135: Conditional inference tree model with outcome defined by the median. Belgian WQ® Feeding principle score, N =63.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 136: , Figure 137: and Figure 138: . No significant variables were detected in model 8, 9 and 10, when defining the outcome using the median or in model 8 and 10when using the P10. Model 10 with the P10 outcome did not converge. Thus, for these models no ROC curves are presented. When P75 was defining the outcome, the reduced model 8 and the conditional inference tree model (model 10) contained the same variables and therefore the ROC curves were identical.

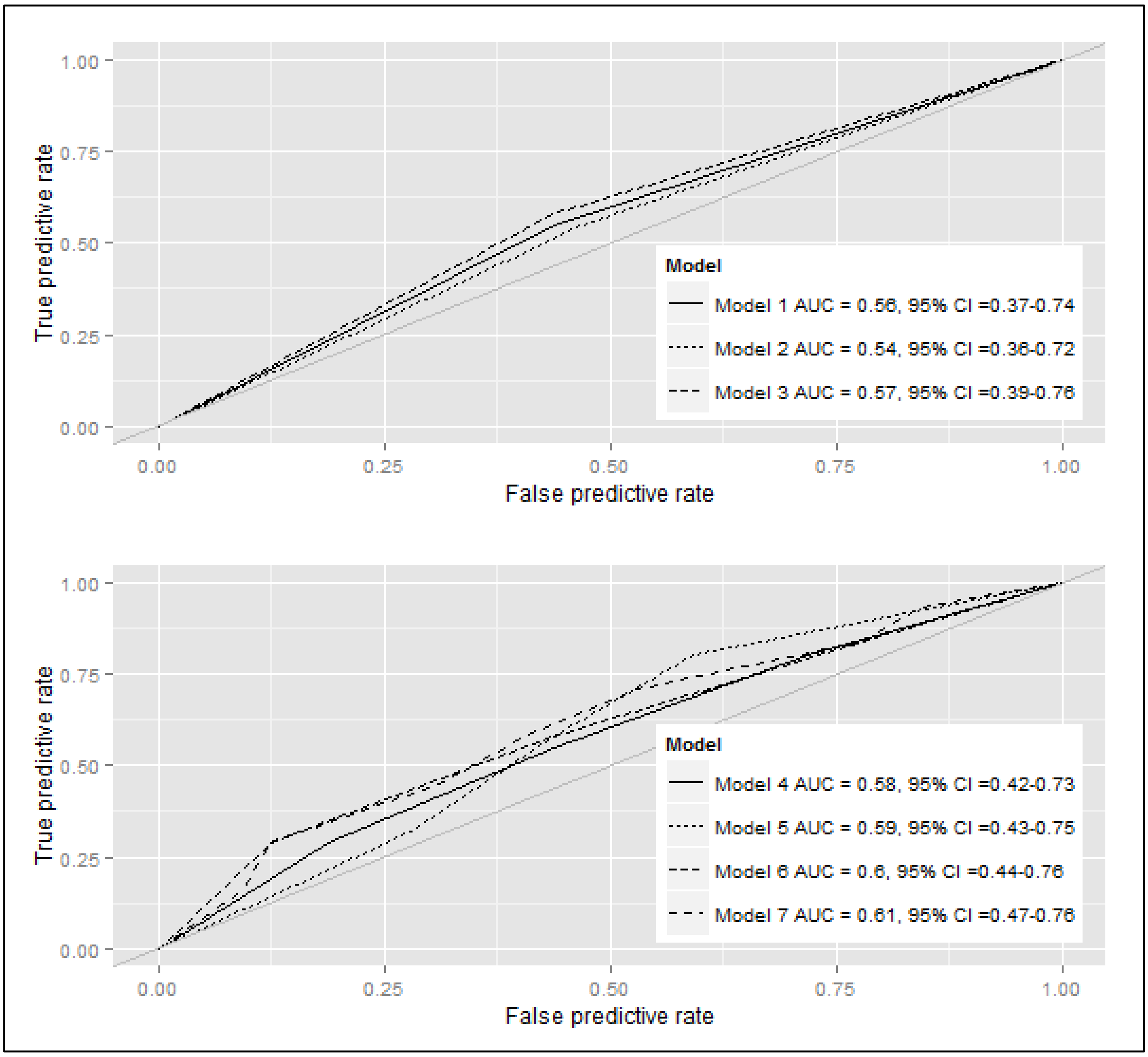


Figure 136: ROC curves and cross validated AUC (with confidence intervals) of models applied to the feeding principle welfare measure from the Belgian WQ® data with outcome defined by the median. N = 63.

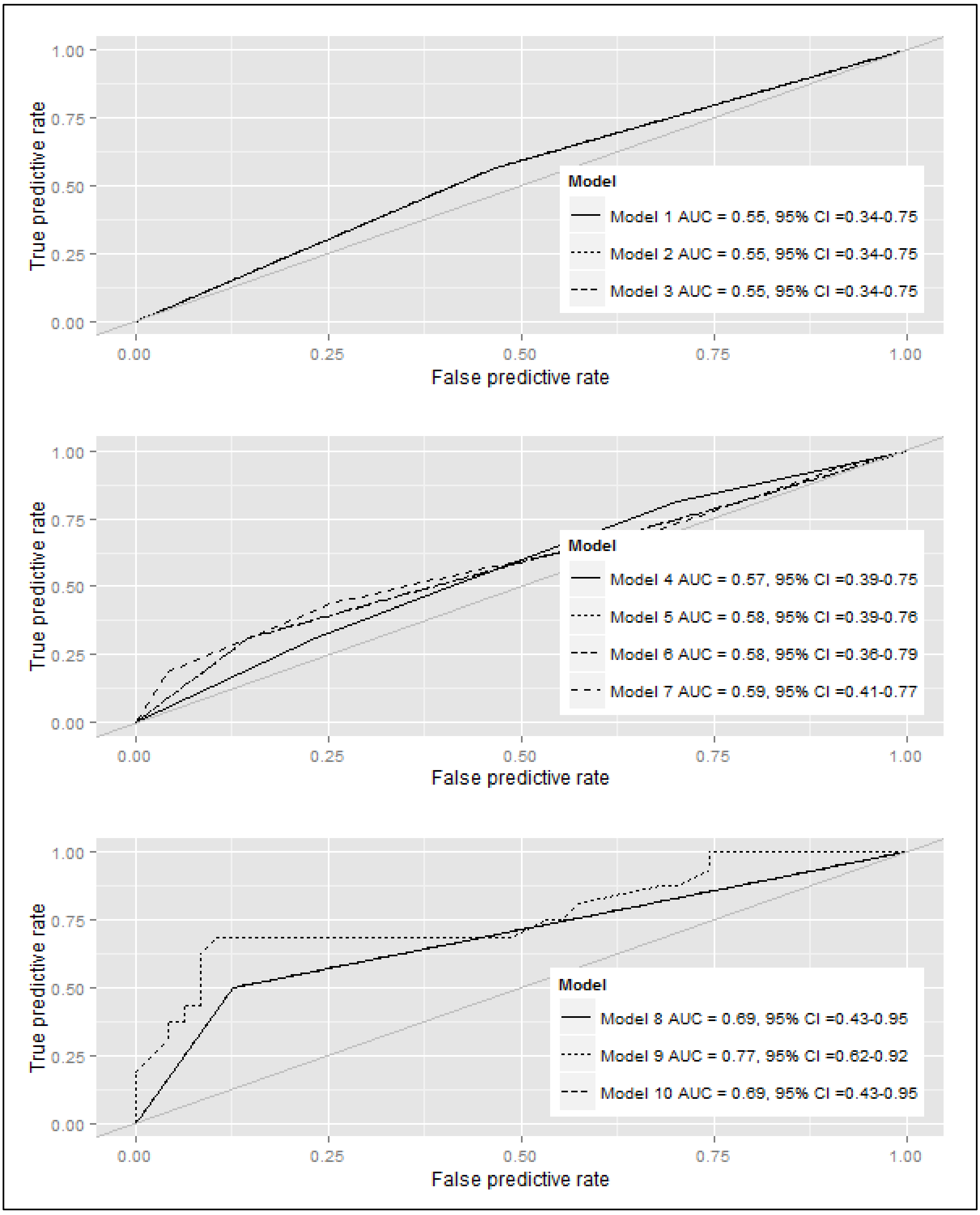


Figure 137: ROC curves and cross validated AUC (with confidence intervals) of models applied to the feeding principle welfare measure from the Belgian WQ® data with outcome defined by the 75th percentile. N = 63.

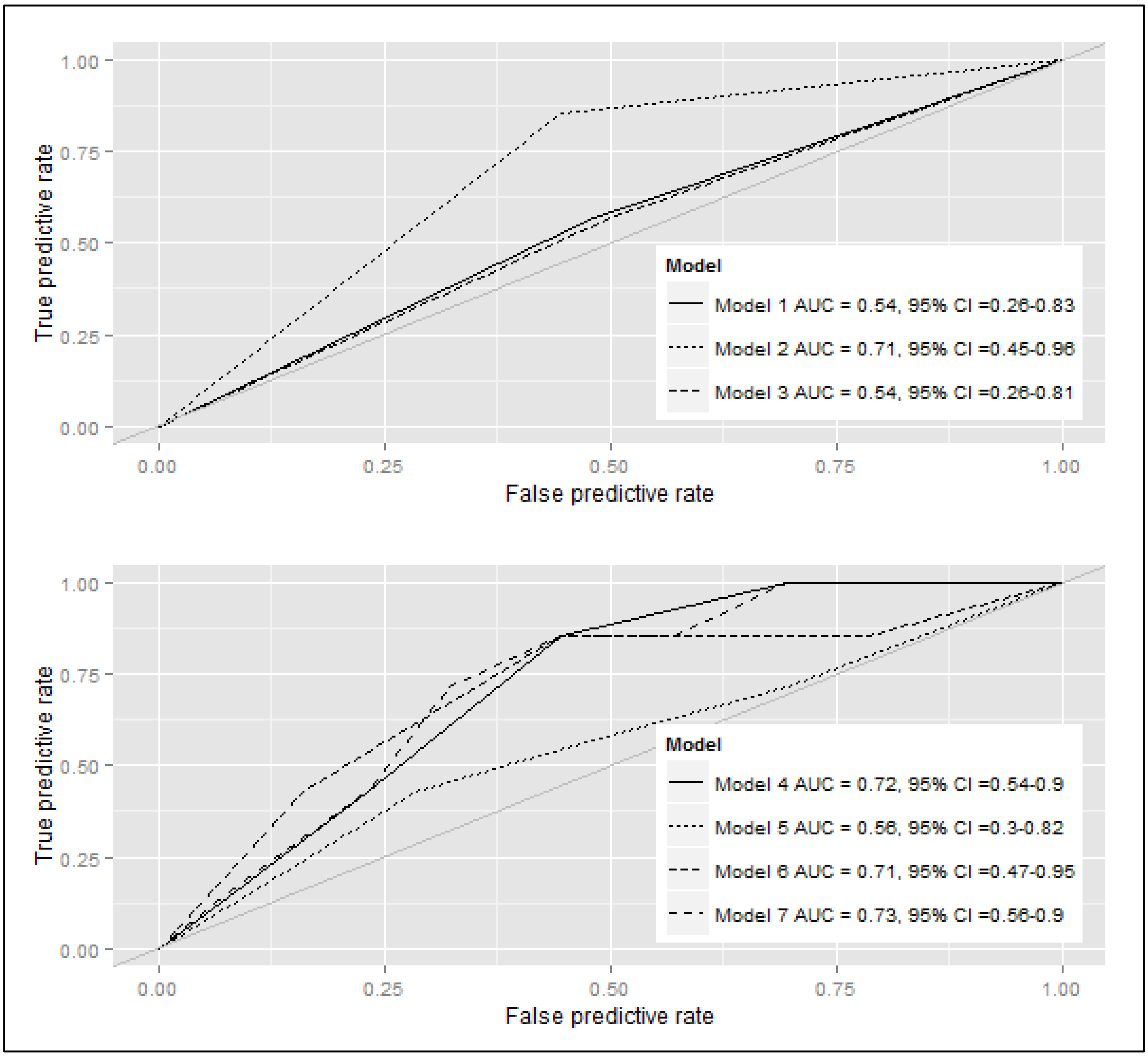


Figure 138: ROC curves and cross validated AUC (with confidence intervals) of models applied to the feeding principle welfare measure from the Belgian WQ® data with outcome defined by the 90th percentile. N = 63.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 96:

Table 96: Comparing model results from ten different models applied to the WQ® feeding principle welfare measure using three different outcomes. Belgian WQ® data, N = 63. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	94.8	0.2632	0.5554	79.3	0.2007	0.5472	52.0	0.1050	0.5446
2	SCC	95.2	0.2648	0.5398	79.3	0.2007	0.5472	47.6	0.0983	0.7054
3	Lameness	94.3	0.2610	0.5716	79.3	0.2007	0.5472	52.1	0.1051	0.5357
4	Mortality + SCC	98.5	0.2700	0.5766	83.0	0.2062	0.5698	51.7	0.1035	**0.7232**
5	Mortality + Lameness	98.0	0.2680	0.5912	82.7	0.2049	0.5751	56.1	0.1076	0.5574
6	SCC + Lameness	97.6	0.2661	0.5993	82.7	0.2049	0.5751	51.8	0.1032	0.7079
7	Mortality + SCC + Lameness	101.3	0.2733	0.6134	86.2	0.2093	0.5891	55.8	0.1079	**0.7270**
8	All factors, additive	NA	NA	NA	71.0	0.1737	0.6862	NA	NA	NA
9	All factors, incl. interactions	NA	NA	NA	67.2	0.1542	**0.7706**	NA	NA	NA
10	Conditional Inference Tree	NA	NA	NA	NA	NA	0.6862	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

11.2.5.4. Behaviour principle

In Table 97: the formulas and the F-test p-values of the different models applied to the data using the median as the threshold for the dichotomisation of the outcome are shown. The formulas and the F-test p-values for the models with the outcome dichotomised using the P75 and the P90 are shown in Table 98: and Table 99: respectively. For the additive model (model 8) and the interaction model (model 9) the final model found by stepwise, backwards elimination are shown. For the conditional inference tree model (model 10), the significant variables are listed; please see Figure 139: , Figure 140: and Figure 141: for a detailed presentation of the structure and the nodes in these models.

Table 97: Model formulas for models with outcome defined by the median. Belgian WQ® data, behaviour principle, N = 63.

Model	**Formula**	**F-test, P value**
Model 1	Mortality	0.8981
Model 2	SCC	0.8981
Model 3	Lameness	0.8981
Model 4	Mortality + SCC	0.9845
Model 5	Mortality + Lameness	0.9795
Model 6	SCC + Lameness	0.9795
Model 7	Mortality + SCC + Lameness	0.9956
Model 8	Unreliable due to rank deficiency	NA
Model 9	Unreliable due to rank deficiency	NA
Model 10	Pasture	NA

Table 98: Model formulas for models with outcome defined by the 75th percentile. Belgian WQ® data, behaviour principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.6128
Model 2	SCC	0.6128
Model 3	Lameness	0.0929
Model 4	Mortality + SCC	0.7828
Model 5	Mortality + Lameness	0.1640
Model 6	SCC + Lameness	0.1640
Model 7	Mortality + SCC + Lameness	0.2184
Model 8	Herd size + Pasture	0.0002
Model 9	Pasture + Herd size*Lameness	<0.0001
Model 10	Pasture	NA

Table 99: Model formulas for models with outcome defined by the 90th percentile. Belgian WQ® data, behaviour principle, N = 63.

Model	Formula	F-test, P value
Model 1	Mortality	0.6556
Model 2	SCC	0.7211
Model 3	Unreliable due to rank deficiency	NA
Model 4	Mortality + SCC	0.8425
Model 5	Unreliable due to rank deficiency	NA
Model 6	Unreliable due to rank deficiency	NA
Model 7	Unreliable due to rank deficiency	NA
Model 8	Unreliable due to rank deficiency	NA
Model 9	Did not converge	NA
Model 10	Pasture	NA

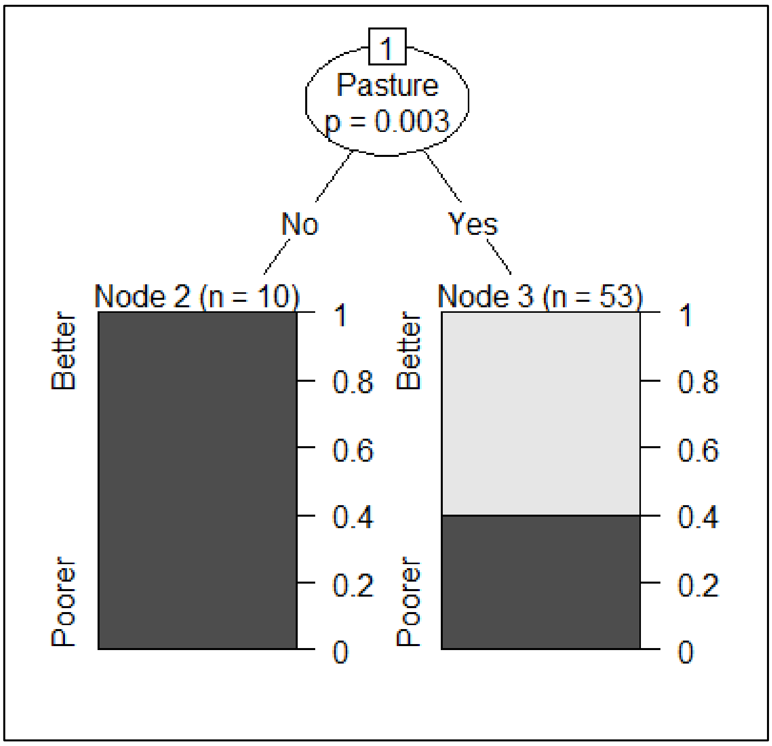


Figure 139: Conditional inference tree model with outcome defined by the median. Belgian WQ® Behaviour principle score, N =63.

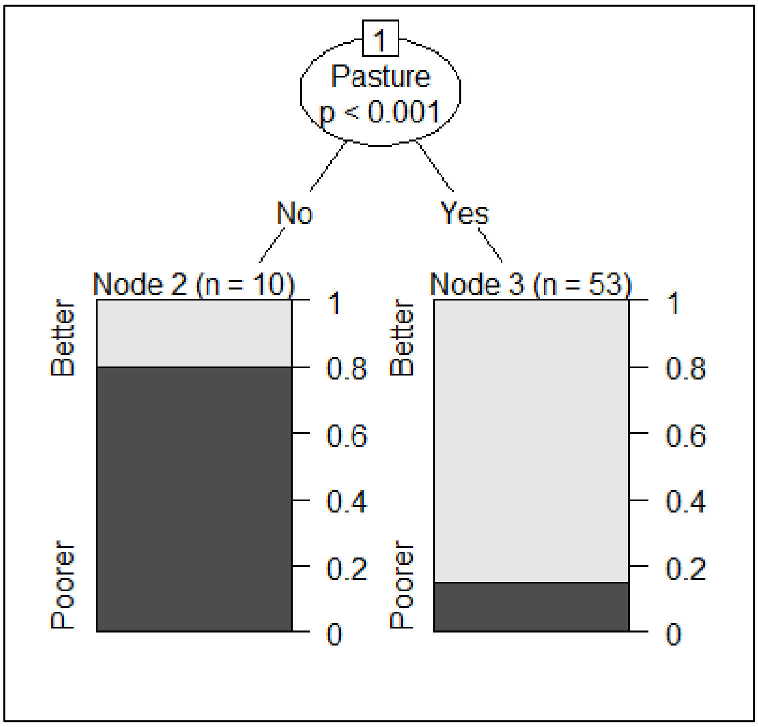


Figure 140: Conditional inference tree model with outcome defined by the 25th percentile. Belgian WQ® Behaviour principle score, N =63.

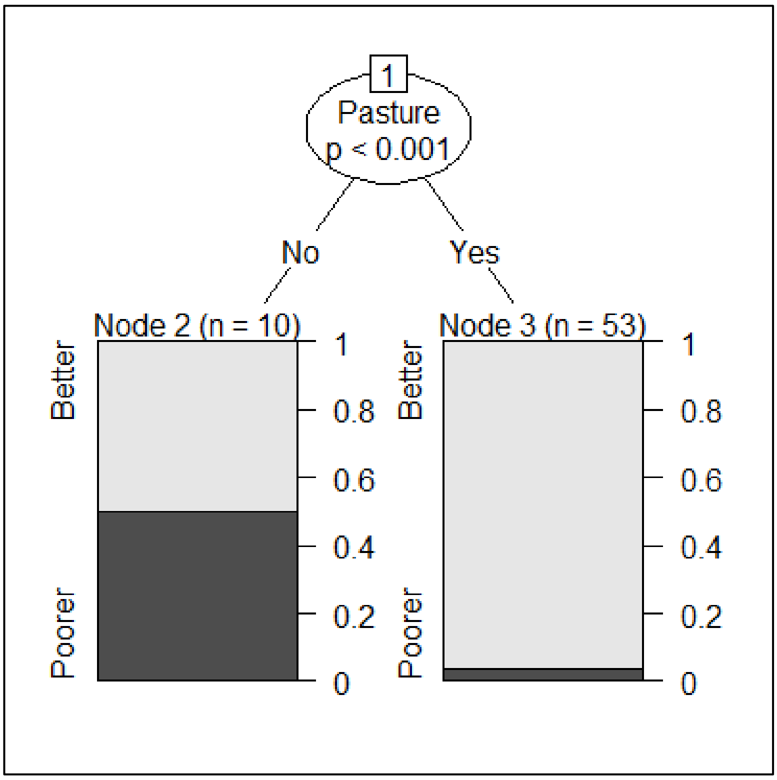


Figure 141: Conditional inference tree model with outcome defined by the 10th percentile. Belgian WQ® Behaviour principle score, N =63.

The ROC curves and the cross validated AUC (with confidence intervals) are presented in Figure 142: , Figure 143: and Figure 144: .

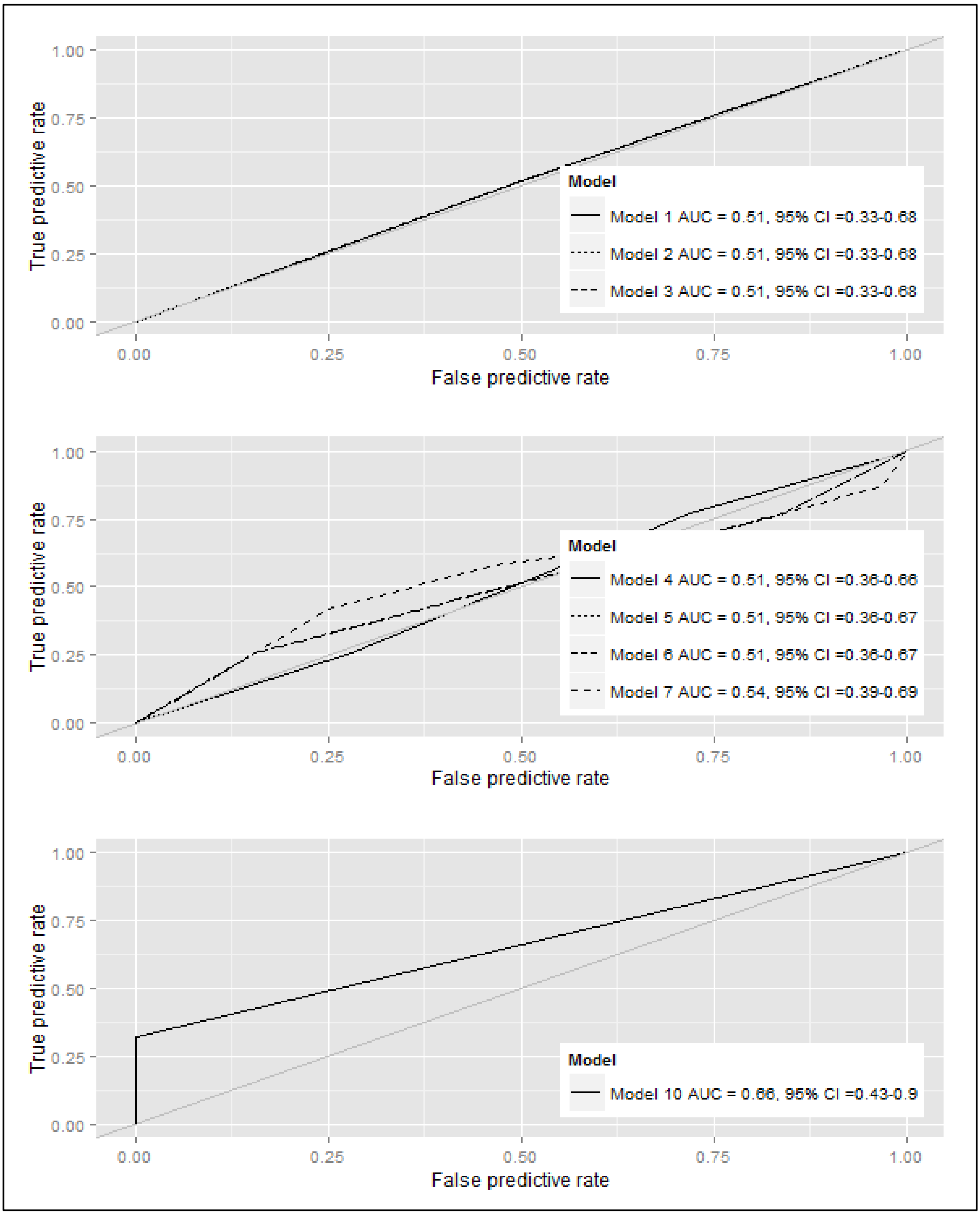


Figure 142: ROC curves and cross validated AUC (with confidence intervals) of models applied to the behaviour principle welfare measure from the Belgian WQ® data with outcome defined by the median. N = 63.

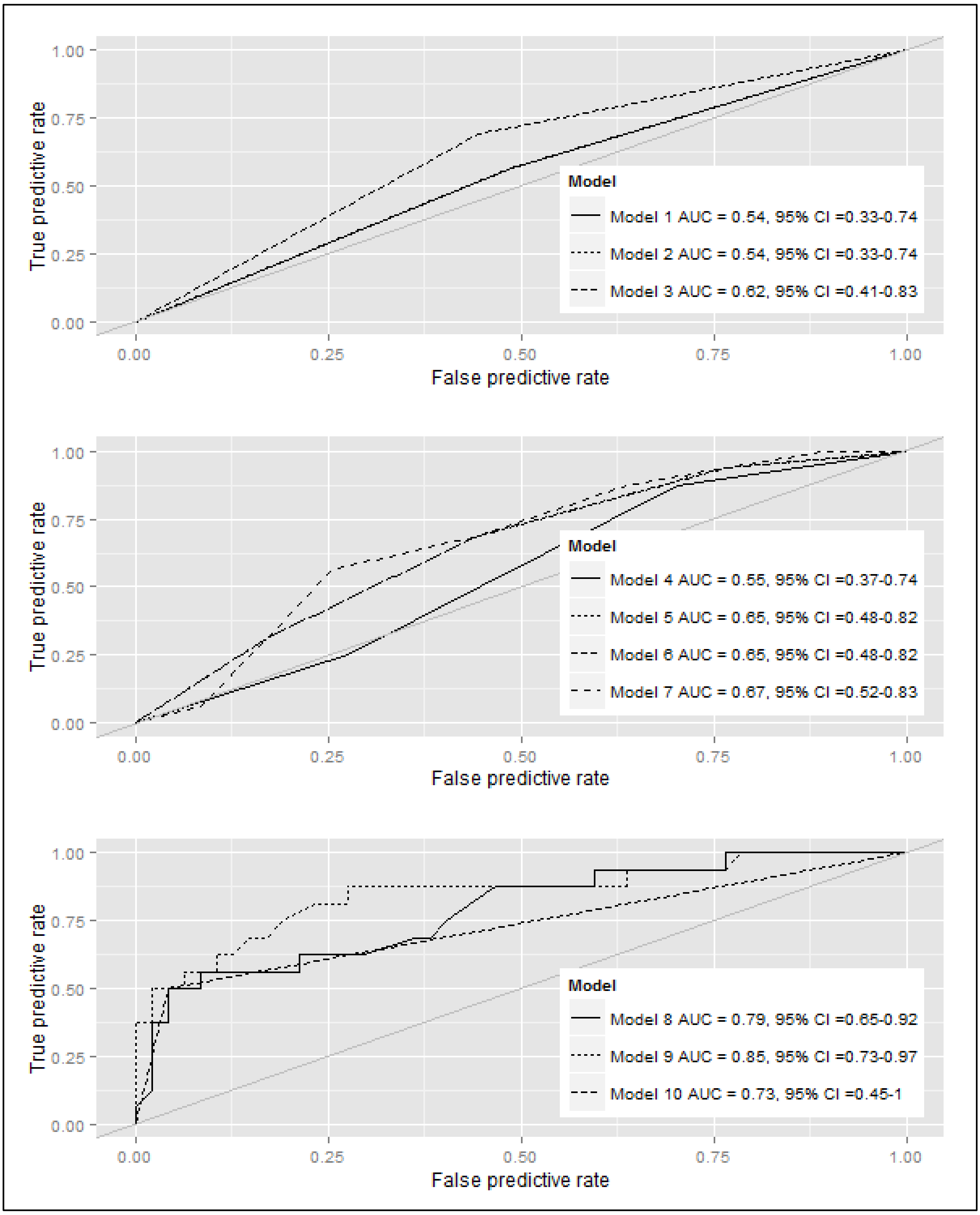


Figure 143: ROC curves and cross validated AUC (with confidence intervals) of models applied to the behaviour principle welfare measure from the Belgian WQ® data with outcome defined by the 75th percentile. N = 63.

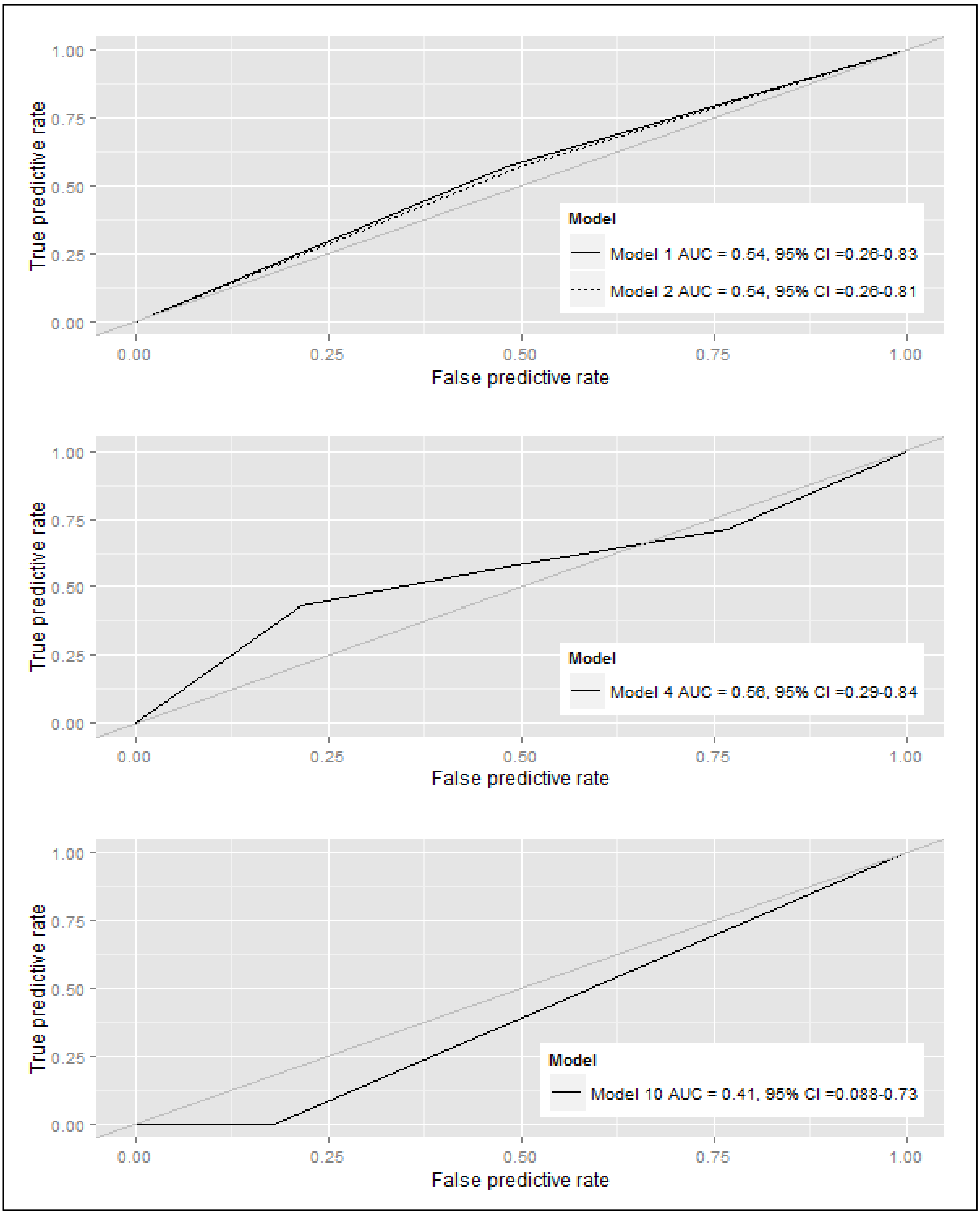


Figure 144: ROC curves and cross validated AUC (with confidence intervals) of models applied to the behaviour principle welfare measure from the Belgian WQ® data with outcome defined by the 90th percentile. N = 63.

In order to evaluate whether the ABMs alone, ABMs in combination or ABMs combined with the factors of variation are the best approach for benchmarking herds with poor welfare the models are compared using the

Bayes Information Criteria (BIC), the mean squared prediction error (MSPE) found by cross validation of the models and the AUC of ROC curves. These parameters are shown in Table 100: .

Table 100: Comparing model results from ten different models applied to the WQ® behaviour principle welfare measure using three different outcomes. Belgian WQ® data, N = 63. AUCs significantly different from 0.5 are marked in bold.

		Outcome median			Outcome P25			Outcome P10		
Model	**Explanatory variables**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**	**BIC**	**MSPE**	**ROC AUC**
1	Mortality	95.6	0.5606	0.5081	79.428	0.2012	0.5366	52.04	0.1050	0.5446
2	SCC	95.6	0.2664	0.5081	79.428	0.2012	0.5366	52.112	0.1051	0.5357
3	Lameness	95.6	0.2664	0.5081	76.862	0.1931	0.6203	NA	NA	NA
4	Mortality + SCC	99.7	0.2752	0.5116	83.338	0.2075	0.5532	56.04	0.1077	0.5625
5	Mortality + Lameness	99.7	0.2752	0.5136	80.212	0.1981	0.6509	NA	NA	NA
6	SCC + Lameness	99.7	0.2752	0.5136	80.212	0.1981	0.6509	NA	NA	NA
7	Mortality + SCC + Lameness	103.8	0.2846	0.5408	83.538	0.2049	**0.6709**	NA	NA	NA
8	All factors, additive	NA	NA	NA	67.163	0.1476	**0.7852**	NA	NA	NA
9	All factors, incl. interactions	NA	NA	NA	64.292	0.1436	**0.8477**	NA	NA	NA
10	Conditional Inference Tree	NA	NA	0.6613	NA	NA	0.7287	NA	NA	NA

BIC = Bayes information criteria, MSEP = Mean squared prediction error, ROC AUC: Receiver operating characteristic, Area under curve. ROC AUC

None of the ABMs were on their own significantly associated with the behavioural principle. Lameness interacted with herd size meaning that a herd with a high prevalence of lameness was more likely to classify as a poorer welfare herd the bigger the herd was.

Access to pasture seemed to be of some importance and was deemed significant in the logistic regressions with the P25 outcome and in all the conditional inference tree models. However, in the health principle, access to pasture was associated with an increased risk of classifying in the poorer welfare group which was opposite previous findings in the other datasets.

11.2.5.5. Summarising results of Belgian Welfare Quality® data analyses

The ABMs seemed to be most closely associated with the health principle. Mortality and SCC in combination were also associated with the housing principle, whereas lameness interacted with herd size in the P25 model of the behavioural principle.

Surprisingly, herd having access to pasture were more likely to classify as poorer welfare herds based on the health principle, whereas no access to pasture was associated with poorer welfare when defined by the behavioural principle. Tiestall and increasing herd size seemed to associated with a higher risk of classifying as a poorer welfare herd

11.2.6. **Summary of results from Objective 5**

In Table 101: the results of the statistical analyses from Objective 5 are summarised.

In Objective 5, nine different welfare assessment outcomes (IZSLER/CReNBA, Otten animal base, Otten register based, Otten system based, Burow, Welfare Quality® health principle, Welfare Quality® housing, Welfare Quality® feeding principle and Welfare Quality® behaviour principle) were used for analysing the ability of the ABMs and the factors of variation to predict the welfare status of the herds. Except from the IZSLER/CReNBA data, the association between the ABMs and the overall welfare outcomes were generally low and the ROC AUCs were rarely significantly larger than 0.50. Equally, the combination of the ABMs seemed beneficial in the IZSLER/CReNBA data but results from the other datasets could not confirm this finding.

When significant associations between the ABMs and the risk of classifying as a poorer welfare herd were found, results confirmed the expectation from the expert opinions in previous objectives: Increased levels of mortality, SCC and lameness were associated with poorer welfare. Where mortality and SCC were both associated with the poorer welfare outcome in five out of the thirteen different welfare measurements, lameness seemed to be slightly more sensitive as it was found significant in seven out of the thirteen measures.

In the IZSLER/CReNBA data, models containing the combined effects of the ABMs performed better than models including the effect of the only one ABM when judged by lower BIC and MSPE values produced by these models and the increased ROC AUC. Yet, this was not the case in all datasets. Often, only one ABM was significantly associated with the overall welfare and combining this with other ABMs by forcing these into the model resulted in increasing BIC and MSPE values and decreasing ROC AUCs. However, the relatively small number of observations in some of the datasets resulted in rank deficiency problems and thus unreliable model estimates and thus made it difficult to investigate interactions between the variables. The conditional tree models were used in an attempt to identify structures in the data that were not revealed by the logistic regression models. Here, only the tree models build on the IZSLER/CReNBA data resulted in more complex models and again, the relatively small sample size in the other datasets may have influenced this as the tree modelling process haltered if nodes contained less than 20 observations or if terminal nodes contained less than 7 observations. Generally, the conditional inference tree models did not increase the predictive value of the models compared to the logistic regression models.

Seven out of the ten factors of variation were identified in at least one of the datasets. Days in milk and parity were not identified in any of the data. Geographical region was identified in the Danish data (Otten and Burow) and in the IZSLER/CReNBA data. However, in Denmark the regional differences are minimal and therefore it did not make any biological sense to consider this variable in the Danish data. Also, the data was skewed and there were very few observations from some regions. In the IZSLER/CReNBA data, the definition of the geographical region was not well described and there were many levels. Thus, the biological meaning of using region as an explanatory variable was considered too low.

For the remaining factors of variation (production type, housing system, floor condition, pasture access, herd size, breed and milk yield), all of them were found significantly associated with the welfare status in at least one out of the thirteen models. However, decreasing herd size and milk yield were both associated with increased risk of poorer welfare in the IZSLER/CReNBA data, whereas in the other data where herd size and milk yield were deemed important, increasing herd size and milk yield were associated with the risk of classifying as a poorer welfare herd.

Table 101: : Summary of results of the statistical analyses performed on all five datasets and thirteen different welfare measures. Arrows indicate which level of the factor that was associated with an increased risk of being classified as poorer welfare herd.

	IZSLER/ CReNBA	Otten			Burow	French Welfare Quality®				Belgian Welfare Quality®			
		Animal based	System	Register		Health	Housing	Feeding	Behav.	Health	Housing	Feeding	Behav.
Mortality	↑	NS	↑	↑	NS	NS	NS	NS	NS	↑	(↑)	NS	NS
SCC	↑	NS	NS	↑	NS	NS	NS	NS	NS	↑	↑	(↑)	NS
Lameness	↑	↑	↑	NS	↑	NS	NS	(↑)	NS	↑	NS	NS	(↑)
Prod. type		(Organic)	Organic	NS									
Housing	(Tethering)					Cubicles	Cubicles	NS	(Cubicles)	NS	Tie stall	Tie stall	NS
Floor	Unsuitable												
Pasture						(No)	(No)	(No)	(No)	Yes	NS	NS	No
Herd size	↓	NS	NS	NS	NS	↑	↑	NS	↑	NS	NS	↑	↑
Breed						Milk	Milk	NS	Milk				
Milk yield	↓	NS	NS	↑	NS								

↑= Increased level of factor associated with an increased risk of classifying as poorer welfare herd; ↓ = Decreased level of factor associated with an increased risk of classifying as poorer welfare herd; NS = Non-significant. ( ) = Association as part of interaction or significant in logistic regression but with low predictive value. Empty cell means that the factor was not present in the data.

11.2.7. Suggestions for fine tunings of the approach

For a future use of ABMs for evaluation of dairy cow welfare to be successful, results from Objectives 4 and 5 indicate that the definition of the ABMs and the way they are recorded need to be subjected to careful investigation in order to standardise these measures between countries.

Lameness was deemed as an important ABM in the expert elicitations in Objective 2 and 4 and was also associated with different overall welfare measures in the analyses performed in Objective 5. Unfortunately, lameness is at the moment not routinely collected in any of the member states participating in this project. Suggestions regarding the future collection of the ABM lameness will be given in Objective 6.

For the ABMs mortality, results indicate that it may be interesting as a future indicator regarding the overall welfare status at the herd level. Within the EU it is mandatory to register cow mortality and it can thus be expected that this indicator would be present in most – if not all - countries. However, some issues are still to be considered. It is suggested that mortality should be recorded as a rate, for example by using the annual mortality rate (as suggested in Objective 2 (Table 23:). Using the annual mortality rate will ensure comparability across countries regardless of differences in herd size and production systems etc. and it is less sensitive to sudden changes in the level when compared to a point estimate. Still, it needs to be ensured that data found in different registers defines which animals that count as dead animals – e.g. whether emergency slaughtered cows are included or not.

Furthermore, from the descriptive analyses of the research datasets used in Objective 5 as well as the data collected in Objective 4 it could be seen that the annual mortality rate differed substantially between the countries providing these data. Therefore, the definition of one common threshold defining acceptable versus unacceptable levels of mortality cannot be made unless these differences between countries are taken into account. For example, this could be done by defining thresholds specific for each country or alternatively, by using a data driven threshold like for example a given percentile.

Regarding SCC used as an ABM in the future, the level of the BMSCC was comparable between Denmark, Belgium and Italy – the countries that provided SCC data for Objective 4. This could indicate that more generally defined thresholds may be used for this parameter. In the three datasets that were provided for Objective 4, three different types of SCC were present. These were bulk milk SCC (BMSCC), percentage of animals above 400,000 cells/ml (HSCC) and individual cow SCC (ISCC). While ISCC could be transformed in such a way that it can be merged with HSCC, it can't be transformed into BMSCC. Thus, Objective 4 concluded that the type of SCC most feasible to collect at the moment is the BMSCC. In Objective 5, the BMSCC at the day of visit was associated with the overall welfare in the IZSLER/CReNBA data and in the Otten data an association between the BMSCC averaged over one year and the register based welfare measure was identified. In order to reduce the effect of single day variation in the BMSCC, it is suggested that BMSCC should be aggregated over time. In Objective 2, it was suggested to aggregate the BMSCC as the proportion of BMSCC measures over a given threshold within a period of three months. Alternatively, the mean BMSCC for a given period of time could be used. A comparable aggregation could be used for ISCC if they are more commonly collected at some time in the future.

12. Objective 6

As part of the other activities in the project, ABMs and risk factors have been identified where no routine recordings are made. The aim of objective 6 was to discuss possibilities for a future collection of a few selected ABMs/factors of variation in dairy herds.

Based on the results from the other objectives of this project, this task focused on two ABMs/factors of variation which are not already collected on a routine basis but are regarded as important for an overall welfare assessment. For each of these two variables suggestions on how to measure them will be given and potentials and limitations regarding their possible future use will be discussed. This task was based on the existing literature and expert interaction within the project.

Based on the results from Objective 2 stating that annual mortality rates, somatic cell count and measures of lameness are suitable indicators of dairy cows welfare an additional list of relevant factors of variation was generated in Objective 3. Table 102: presents these factors of variation and the scores from Objective 3 indicating the mean relevance and feasibility.

Table 102: List of factors of variation of relevance for an overall welfare assessment on farm level from Objective 3. Mean relevance scores on a scale from 1 (low relevance) to 5 (high relevance). Feasibility is number of experts out of the panel of 11 experts judging that the variable would be easy to collect in the field and keep updated. Sorted after decreasing mean relevance.

Factors	Mean relevance	Feasibility
Density	4.4	6
Parity	4.2	11
Housing system	4.2	10
Empathy, threshold to recognise disease and pain, call veterinarian	4.2	0
Ability to detect/report diseases and welfare issues	4.2	0
Hoof status	4.0	2
Veterinary treatments	3.9	6
Age	3.8	9
Animal care/hygiene	3.8	1
Injuries, accidents, pain	3.8	2
Floor type	3.7	10
Body condition score	3.6	2
Days in milk/after calving	3.6	10
Reproductive problems	3.6	3
Floor quality	3.6	4
Management of dry cows	3.6	3
Culling rules	3.6	1
Outdoor conditions	3.6	4
Access to pasture	3.5	10
Water quality and availability	3.5	1

Factors	Mean relevance	Feasibility
Milk production/yield	3.5	11
Udder status	3.5	0
Bedding material	3.5	7
Negative energy balance	3.5	0

Based on the knowledge generated during Objectives 1-5 of this project it was decided to include lameness in Objective 6. Lameness has been demonstrated to have clear associations to cow welfare and is not routinely collected in many countries. From Table 1 it can be seen that density (stocking density) has a high score for relevance and a low score for feasibility. Even though the feasibility of future recordings and the difficulties in the interpretation of the relevance for animal welfare of different levels of stocking density measures was questioned by the project consortium, EFSA representatives still wished density to be discussed in Objective 6. Hence, the two variables to be further investigated in Objective 6 were lameness and density.

12.1. Possible future collection of data on lameness and density

12.1.1. **Lameness**

Lameness is considered a major welfare issue in modern intensive dairy production (Bruijnis et al., 2012) and causes significant economic losses to the farmers (Ettema et al., 2010). The prevalence of lameness in dairy farms varies between 25-70% in UK farms (Bennett et al., 2014) with similar results from the rest of Europe (Van Hertem et al., 2014).

12.1.1.1. How to assess lameness

A wide selection of methods for assessing lameness by observing the locomotion/gait of cattle has been developed in recent years (Whay, 2002). These are often referred to as 'locomotion scoring' or 'gait scoring' systems. The Welfare Quality® project combined aspects of several of these methods into one assessment that in the end calculates a score from 0-100 (where 0 is worst situation and 100 is best situation) for the lameness related welfare status in a dairy herd (Welfare Quality®, 2009).

A procedure which may be used to assess lameness on a regular basis is described in the two tables below. Table 103: describes the procedure when assessing lameness in a loose housing system either with concrete and/or rubber matted floor or in a loose housing system with an alternative floor type such as deep straw bedding. 0describes the procedure for assessing lameness in a tie-stall barn.

Table 103: Assessments of lameness of dairy cows in loose housing systems (Welfare Quality®, 2009).

Sample size	Sample size according to Table 106:
Method description	This measure applies to lactating cows, dry cows and pregnant heifers if kept with lactating animals as well as all dry cows if kept separately, able to move freely and individually controlled, i.e. loose housed animals as well as animals which are kept in tie stalls but are released at least twice a week. Lameness describes an abnormality of movement and is most evident when the legs are in motion. It is caused by reduced ability to use one or more limbs in a normal manner. Lameness can vary in severity from reduced ability to inability to bear weight. Indicators of lameness are: • irregular foot fall • uneven temporal rhythm between hoof beats • weight not borne for equal time on each of the four feet The following gait attributes are taken into account: • timing of steps • temporal rhythm • weight-bearing on feet. Assess the gait score of the animal. All animals should be walked in a straight line on a hard, level, non-slippery surface on which they would normally walk. The assessor should view them from the side and/or behind. Animals must not be assessed when they are turning.
Classification	Individual level: 0 – Not lame: timing of steps and weight-bearing equal on all four feet. 1 – Lame: imperfect temporal rhythm in stride creating a limp 2 – Severely lame: strong reluctance to bear weight on one limb, or more than one limb affected Herd level: Percentage of not lame animals (score 0) Percentage of moderately lame animals (score 1) Percentage of severely lame animals (score 2)

Table 104: Assessments of lameness of dairy cows in tie-stall barns (Welfare Quality®, 2009).

Sample size	Sample size according to Table 106:
Method description	This measure applies to all lactating cows, dry cows and pregnant heifers if kept with the lactating animals kept in tie stalls and which are not released at least twice a week. Lameness describes an abnormality of movement and is most evident when the legs are in motion. It is caused by reduced ability to use one or more limbs in a normal manner. However, in some tie stall systems it will not be practical to release the cows to carry out gait scoring. A method for detecting lame cows in tie stalls has been developed and validated against gait scoring. The 'stall lameness score' is based upon the following indicators: Resting: • Resting a foot (one more than another). Standing: • Standing on the edge of a step (to avoid bearing weight on one foot/part of foot). Stepping: • Frequent weight shifting between feet ("stepping"), or repeated movements of the same foot (this could also be due to nervousness, flies, or anticipation of feeding.) Reluctance: • Reluctance to bear weight on a foot when moving. Assess the score of the animal. Firstly observe how the cow stands when undisturbed. Then move the cow to the left and to the right, observing how she shifts weight from foot to foot. Then observe the position the cow returns to after movement. If the cow has been lying down, get it up and wait 3 - 4 minutes before assessing. Individual level: 0 – Not lame: cow showing none of the indicators listed above 2 – Lame: cow showing at least one of the four indicators listed above
Classification	Herd level: Percentage of not lame animals (i.e. score 0) Percentage of severely lame animals (i.e. score 2)

12.1.1.2. Calculating the score for lameness

In order to calculate a final welfare score for lameness the index generated within the Welfare Quality® project can be used (Welfare Quality®, 2009).

First the index ($\mathbf{I_l}$) for lameness is calculated using the % of lame animals: $\mathbf{I_l} = 100 -$ (% lame cows)

This index is computed into a score using the I-spline functions below Figure 145: .

When $\mathbf{I_l} \leq 78$ then Score = $(0.0988 \times \mathbf{I_l}) - (0.000955 \times \mathbf{I_l}^2) + (5.34E^{-05} \times \mathbf{I_l}^3)$

When $\mathbf{I_l} \geq 78$ then Score = $-2060 + (79.3 \times \mathbf{I_l}) - (1.02 \times \mathbf{I_l}^2) + (0.00439 \times \mathbf{I_l}^3)$

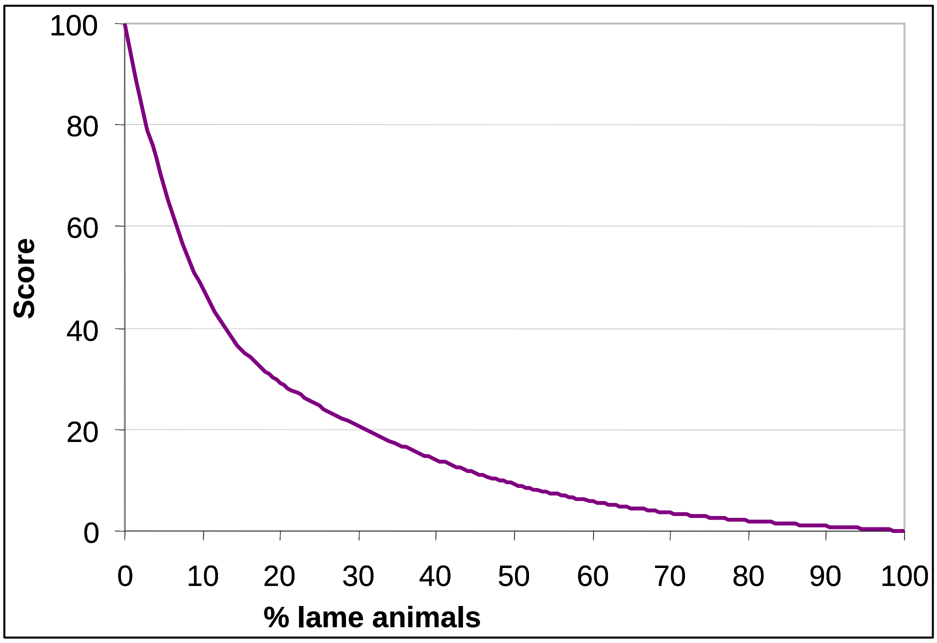


Figure 145: Calculation of the partial score for lameness according to the % lame animals (Welfare Quality®, 2009).

12.1.1.3. Number of farms to be scored per year

Obviously, in cases where lameness status needs to be monitored at the individual herd level, all herds in a given population needs to be scored. If – on the other hand – a measure of the 'general level of lameness' in a region, e.g. a country, is desired, this section will give some indications as to how many herds need to be scored. The number of herds to be selected for lameness assessment in each country in order to estimate the level of lameness in the country is dependent on the estimated average prevalence of cows scored as lame. For the calculated examples in Table 105: a formula for the calculation of a sample size in a population from Houe et al. (2004) was used. For example, if we assume that the average proportion of lame cows is 20% and that we can accept a maximum allowable error of +/- 5% then we will need to assess all cows in 230 herds in order to be able to draw conclusions on the population in that specific country. The number of herds scored only depends to a very small extend on the population size. If the population e.g. increases from 3500 to 50000 herds, the number of herds needed in the above example would increase only marginally from 230 to 250.

Table 105: Calculations of number of herds needed to estimate the overall lameness prevalence in a country under the following assumptions: Confidence level 95%, maximum allowable error 5% and population size (number of herds in the country) 3500 (Houe et al., 2004).

Prevalence of lameness	Estimated sample size
5%	72
10%	134
20%	230
40%	334

Since the prevalence of lameness varies over the seasons it is important that the time for assessments is chosen on a random basis over the year.

12.1.1.4. How to select the animals for assessing lameness

The number of animals selected for measuring lameness is depending on the outline of the barn. If the dairy cows are kept in one group the number of animals to score is presented in Table 106: . If, however, the cows are kept in different groups the sample size of each group should be seen as if the groups were individual herds (Welfare Quality®, 2009).

Table 106: Sample size for clinical scoring depending on the herd size (Welfare Quality®, 2009).

Herd size	Number of animals to score (suggestion A)	If A is not feasible
0	30	30
40	30	30
50	33	30
60	37	32
70	41	35
80	44	37
90	47	39
100	49	40
110	52	42
120	54	43
130	55	45
140	57	46
150	59	47
160	60	48
170	62	48
180	63	49
190	64	50
200	65	51
210	66	51
220	67	52
230	68	52
240	69	53
250	70	53
260	70	54
270	71	54
280	72	54
290	72	55
300	73	55

The following should be seen as guidelines for making the assessment more efficient when assessing lameness (modified from Welfare Quality®, 2009):

1. A random sample can be obtained by selecting every n^{th} animal in the milking parlour. These animals are marked, to enable re-identification afterwards for data-collection.

2. If animals can be locked in a feeding rack, they can be selected by choosing every n^{th} animal in the row(s). Data collection can be carried out immediately.

3. To simplify the assessment, animals can be marked with a stock marking device after assessing them.

4. If animals are kept in different groups, proportionate sampling according to group size should be carried out.

12.1.1.5. Automatic measures of lameness

The worldwide tendency of more and more automated farms, as for example the introduction of automatic milking system and activity meters, has resulted in a substantial amount of data automatically collected every day on an individual cow level. Automatic measures have the benefit of not requiring the presence of human observers. This is labour saving and additionally the need for human presence may bias recordings since the presence of a human observer may increase the likelihood of the animals hiding any vulnerability (Weary et al., 2009). Some of these data can be used for an automatic assessment of lameness such as visit to feeders and the automatic milking system (Borderas et al., 2008, Miguel-Pacheco et al., 2014), automatic recordings of lying time and lying bouts (Ito et al., 2010, Thomsen et al., 2012), weight distribution between the legs while walking and standing (Pastell and Kujala, 2007), automated image analysis (Song et al., 2008), image analysis of 3D video images of cow gait (Van Hertem et al., 2014) and activity meters (Chapinal et al., 2011).

12.1.2. **Density**

12.1.2.1. How to assess density

Density was judged highly relevant in Objective 3 (Table 102:) and a good and easily interpretable measure of density would be relevant to have. This is however not as straightforward as it may seem. First of all there are several measures of ‘density’ based on different resource measures. In Table 107: a number of resource based measures of relevance for assessing density have been identified for three different housing systems. Secondly, the effects different density parameters may have on animal welfare are very much depending on other (resource and management) factors and generally not clearly defined (see 0).

Table 107: A list of resource based measures relevant for assessing density in three different housing systems for dairy cows. For each of the housing system the resource based measure of importance are marked with an 'x'.

Resource based measure	Tie-stall barn	Loose housing	
		With cubicles	With deep bedding
Number of drinkers per animal	x	x	x
Number of feeding places per animal		x	x
Number of concentrate feeders per animal		x	x
Number of cubicles per animal		x	
Area of barn in m^2	x	x	x
Area of alleys in m^2		x	x
Area of straw bed in m^2			x
Size of cubicles/stalls	x	x	
Delivery of fresh feed (times per day)	x	x	x
Delivery of fresh concentrate if no automatic feeders (times per day)	x	x	x

12.1.2.2. Problems and limitations when assessing density as a measure for animal welfare

In order to be able to use density as a reliable risk factor for animal welfare there are several problems and limitations that need to be addressed. As can be seen from Table 107: the measurement of density in itself is not at all straightforward. A lot of different measures have to be taken into account when quantifying the 'overall density' in a system. Additionally, the most crucial factor that can have an immense impact on the welfare of dairy cows under the same and different densities is the variability in management routines. The most crucial limitation of using density as an indicator of animal welfare it that a well managed farm with a high density could, in principle, have a better animal welfare than a not so well managed farm with a low density. Another limitation is that there is no solid scientific information on the relation between density parameters and welfare outcomes under different circumstances. This makes the interpretation (i.e. relation to animal welfare) of information collected on density parameters questionable. In Table 108: the current identified limitations and problems for the different density parameters from Table 107: is presented.

Table 108: Identified problems and limitations for the resource based measures needed for evaluating density as an indicator for animal welfare

Resource based measure	Identified problems and limitations for the assessment
Number of drinkers per animal	In some farms it might be okay for many cows to share a few drinkers if these drinkers are functioning well and are clean.
Number of feeding places per animal	In some farms it might be okay to have e.g. 3 cows per feeding place since they are fed new silage (or TMR) every hour (feed available ad libitum), whereas in another farm where cows are fed e.g. 2 or 4 times per day it might be a problem.
Number of concentrate feeders per animal	In a homogeneous group of cows with low occurrence of agonistic behaviour it might be okay to have less concentrate feeders where in a group of cows with changes in group composition and a lot of agonistic behaviour this is a problem.

Resource based measure	Identified problems and limitations for the assessment
Number of cubicles per animal	In a homogeneous group of cows with low occurrence of agonistic behaviour it might be okay to have less cubicles than cows whereas in a group of cows with a lot of agonistic behaviour and changes in the group dynamics it would need more cubicles than cows in order to ensure that each cow can lie down whenever she needs to without having to fight over the cubicle.
Area of barn in m2	The design and area of the barn might have an impact on the welfare of the cows but a small area that is well designed might be as sufficient as a not so well designed larger area.
Area of alleys in m2	The cows' ability to walk on any given surface is not only determined by the space around them but also on the slipperiness of the surface and if it is solid concrete or rubber mats. Additionally, the design of the barn (blind ends etc.) will also have an impact.
Area of straw bed in m2	The area of the straw bed does not give a sufficient measure of the comfort the cows is experiencing when lying, or while trying to lie down or get up again. This is also depending on the thickness of the bed and how wet it is.
Size of cubicles/stalls	The area of the cubicle/stall does not alone give a sufficient measure of the comfort the cow is experiencing while lying down, or while trying to lie down or get up again. The comfort of the lying area/cubicle is also dependent on e.g. the cubicle dividers, the structure of the surface and if the cubicle is fitted with a mattress.
Delivery of fresh feed (times per day) (concentrate/roughage/TMR)	This measure is in close relation to number of feeding places per cow. Few feeding places per cow will be less of a problem if fresh feed is delivered many times per day (ad libitum).

DISCUSSION

The overall aim of the current project was to investigate the potential of using routinely collected ABMs for the evaluation of the welfare status in dairy herds in the European countries. To fulfil this, valid and robust ABMs closely related to adverse effects in the individuals' welfare and to the overall herd welfare are needed. Also, it should be possible to collect data on the ABMs routinely in a uniform manner across the European countries.

The value of an ABM as a welfare indicator first of all relies on a strong association between the ABM and the adverse effect. The results of the data analysis on Objective 2 revealed only few ABMs with relevant distributions and strong association with the WAEs. In epidemiological terms, sensitivity (Se) and specificity (Sp) inform on the performance of a binary classification test. The Se and Sp report how well a test performs in detecting a given condition compared to a reference standard. The reference standard test is the best available test under reasonable circumstances. Technically, it is feasible to assess multiple thresholds (cut-offs) used to deem a test positive or negative and then estimate the sensitivity and specificity at each cut-off, depending on whether a high sensitivity or a high specificity is more desirable. There is a consensus that tests should be fit-for-purpose (OIE, 2003) and decisions on specific purposes should be made prior to embarking on a study (Nielsen et al., 2011). In principle, multiple cut-offs can be selected even if no specific purpose has been decided. However, the interpretation of the resulting estimates is not necessarily meaningful in a different context. In the context of this part of the project, the reference standard was the WAE as a proxy of welfare problems. The aim was to evaluate the performance of selected ABMs in identifying WAEs. Thus, the ABMs were not compared to an existing well-defined reference standard. Therefore, we needed to define cut-offs for the WAE as well as for the ABMs before estimation of Se and Sp. The definition of thresholds for good or poor welfare can vary among stakeholders and in the end it is a political/ethical decision on what is unacceptable. Therefore, the approach taken here was a more pragmatic suggestion based on expert opinions and the available data. The chosen cut-offs used for the calculation might have influenced this but due to the risk of committing statistical type I errors, a decision-based approach was taken, which can be considered scientifically sound. The definition of a cut-off for a diagnostic test depends strongly on the purpose of the implementation of the test and also on the prevalence of the target condition. The cut-offs relevant for surveillance of the overall animal welfare in member states could very well differ from those relevant for pointing out specific herds with welfare problems. Therefore, in the context of this project, weight has been put on feasibility in terms of data availability in the field. Concerning the prevalence of the target condition, the descriptive analyses in Objective 2 revealed rather large differences in data distribution between the available datasets. This difference between data from different countries was also confirmed by the descriptive analysis of the data collated in Objective 4. In Objective 2, the consortium discussed whether the definition of cut-offs for the analyses of sensitivity and specificity could be data driven. This could for example have been obtained by using the 25% - or another - percentile. However, this would have resulted in very different cut-offs in the three datasets. Alternatively, cut-offs could be defined with percentiles found in larger data bases and thereby depicting the levels found in the dairy cow population (in a given region or country). Also, the definitions of the ABMs as well as the WAEs are important. Again, in this project practicability and feasibility have been important in the choice of ABM definitions.

In Objective 3, a literature search identified a large number of parameters (factors of variation) associated with at least one of the categories investigated, i.e. ABMs/WAEs associated with mortality, mastitis/elevated SCC and lameness. The list of articles is not exhaustive, given the scope of the subject and the project allocated time, but we believe the literature search permitted the identification of the main scientifically investigated parameters. By regrouping similar parameters and excluding anecdotal or irrelevant ones, 63 relevant factors of variation were finally allocated to five entity classes. The largest class is the management class which includes parameters dealing with stockmanship, hygiene, herd health, milking, dry cow management, feeding and culling. The housing class includes parameters dealing with housing systems, space organisation, material, floor and climate. The animal class includes parameters related to genetics, physiology and health. The last class deals with general factors such as climatic conditions.

In the context of this work, 63 parameters were not realistically implementable for field surveys: in order to be effectively used, a lower number of important parameters were needed. A systematic selection method was discussed and adopted within the consortium. The expert consultation for the selection of the final factors of variation was restricted to partners from the consortium due to time constraints limiting our capacity to mount a wider evaluation by external experts. Furthermore, the consortium gathers well-versed individuals in the domains of animal/veterinary science, welfare and with experience with dairy cattle production, health and welfare in the field and we are thus confident in the process. First, parameters were judged for their relevance as regards the three ABMs, in conjunction with their feasibility of collection in the field. This selection method was used to exclude parameters that could only be collected with great difficulty or not at all, at least for the time being. For example, although a number of studies show the positive impact of empathy on several welfare indicators, we do not have the means to measure this parameter in the field routinely. The different scores of relevance given by the different experts show a large degree of agreement. For example, the parameter 'parity' was the factor with the highest mean relevance score and was scored 4 or 5 by 10 of the experts. Secondly, experts were asked to select parameters needed to characterise a given population.

In order to evaluate the ability of the selected ABMs to predict the overall welfare status at herd level, the ABMs were tested using different data comprising welfare measures. One gold standard describing the overall welfare status in the herds would have been the optimal outcome for any statistical analysis regarding the importance of each ABM and factor of variation. Yet, one of the major challenges in animal welfare assessment is exactly the lack of a gold standard. As outlined in the introduction (see Secion 1.1) research in animal welfare assessment have used different approaches and the decision of which type of indicators to collect and how to aggregate them very much depends on the purpose of the welfare assessment.

As no gold standard for overall welfare assessment in dairy herds exists, testing was done using all the available welfare definitions in the datasets instead of choosing one as a gold standard. Thereby, the reliability of any given parameter (ABM or factor of variation) could increase based on the reflection that if any parameter was capable of correctly predicting the welfare status in multiple different welfare assessment systems, the association of that parameter with the welfare status would be more robust. However, as the ABMs and also the factors of variation were typically used in the calculation of the overall welfare status there was always a danger of creating a self-fulfilling prophecy when they were also used for predicting the welfare status. Also, the number of indicators and exact procedure of aggregating them in the different systems are of importance. If many indicators are used to calculate the welfare status, the weight of each indicator will be smaller than if the calculation is based on a few indicators. Also, different welfare indicators could be weighted differently in different aggregation procedures also depending on whether compensation is allowed in the aggregation or not. This has to be kept in mind and caution should be taken in the interpretation of the results of the statistical analyses. Still, the chosen procedure can give indications of whether or not a given parameter is at all associated with the overall welfare status of the herd regardless of how it is measured or whether the parameter is more closely related to one specific way of measuring dairy cow welfare.

The welfare assessment systems used in Objective 5 to evaluate the predictive ability of the ABMs and associated factors of variation represented different approaches to the evaluation of the overall welfare: resource based (the Otten resource based measure), animal based (WQ® health principle, the

Otten animal based and the Otten register based system (indirect ABMs)) and combinations hereof (the IZSLER/CReNBA system, WQ® housing, feeding and behaviour principles and the Burow system). Given that no gold standard for the overall welfare assessment exists, the chosen approach covered different aspects and principles within the concept of welfare assessment.

The five datasets were all originally collected for research purposes. In the Danish data (Otten + Burow), inclusion criteria were defined as herds larger than 100 cows having loose-house systems with cubicles in the resting area. These criteria were chosen in order to create a sample that was representative of the dairy herd population in Denmark. From these herds, herds with access to pasture grazing were selected for the Burow data, as this was the focus of that study. For the IZSLER/CReNBA data herds were randomly selected from the dairy cow population and used for the development of the welfare assessment protocol. The WQ® data from France as well as Belgium were used to compare traditional versus modern production systems and herds were selected in order to represent these production types. When comparing to the descriptive analyses performed on the data collected in Objective 4, the levels of annual mortality rate and SCC seem comparable to the general dairy poppulation and based on this it seems reasonable to consider the data used in this project as resonably representative of the dairy herd population in the respective countries.

The data from the IZSLER/CReNBA data base contained 608 herds and thus models logically were more stable and confidence intervals more narrow than in the other datasets. Compared to the Welfare Quality® principles, the aggregation procedure in the IZSLER/CReNBA score was more simplistic as it is based on sums of weighted score whereas the aggregation of the Welfare Quality® principles was more complex. This might also partly explain why the ABMs seemed to be more closely related to the welfare outcome in the IZSLER/CReNBA data. In the literature, it has previously been suggested that due to different numbers of indicators in the different principles and the complex aggregation procedure the overall classification in the Welfare Quality® protocol is rather insensitive to changes in the animal based indicators lameness and SCC (De Vries et al., 2013). Whether this is true for the principles are not known. In the current project, the ABMs were not significantly associated with any of the principles in the French data. Oppositely, all three ABMs were significantly associated with the health principle in the Belgian data whereas results from the three other principles varied in which ABMs showed any association with the outcome.

All the welfare assessment scores used in Objective 5 were provided on a continuous scale. In order to investigate the potential in using the ABMs and the factors of variation for benchmarking the welfare status of the herds, the welfare outcome were dichotomised using three different thresholds: the median, the 25th or the 75th percentile and the 10th or the 90th percentile. In the end, the level at which herds should be benchmarked is a political decision of what is acceptable and unacceptable dairy cow welfare. For the purpose of Objective 5, these three thresholds were chosen in order to investigate whether the different variables were able to predict the welfare status at different detection levels. The more sensitive a predictor is; the smaller a difference can be detected. The difference between the mean welfare scores in the two groups defined by the median is expected to be smaller than the difference between two groups defined by the 90th percentile. Thus, if a parameter was able to predict the welfare status correctly at all three levels it could indicate that this parameter might be more robust, than a parameter only capable of correctly predicting the welfare status of the more extreme herds. However, due to the small sample size in some of the datasets, the dichotomisation in some cases led to very low number of observations with the poorer welfare status and resulted in problems with rank deficiency and unstable models.

In the IZSLER/CReNBA data, all variables including the ABMs were given as categorised variables. In the other datasets, the ABMs were dichotomised before the statistical analyses. In Objective 2, the ABMs were defined and thresholds for the dichotomisation were suggested based on literature and discussions within the consortium (see Table 23:). However, these thresholds were not always meaningful with the data available for analyses in Objective 5. Instead, thresholds for the dichotomisation of the ABMs for the analyses were data driven by using the value of the median of the ABM. This resulted in rather different thresholds when comparing the different datasets. Interestingly, the data driven thresholds found in Objective 5, were generally higher than the thresholds suggested in Objective 2. As mentioned, the reason for using the data driven thresholds was to obtain a meaningful distribution of data for the analyses of the association between the ABMs and the overall welfare status. Thus, the thresholds are not depicting an acceptable/unacceptable level of a given ABM regarding the welfare status but were rather used to evaluate whether a higher level of a given ABM was associated with a higher risk of a decreased overall welfare. Nevertheless, the variation in the levels of the ABMs in the different datasets used for the analyses in Objective 5 suggests that these differences needs to be investigated firmly in more routinely collected data before general thresholds of acceptable and unacceptable levels of the ABMs can be suggested.

One of the aims of Objective 4 was to run a pilot test to evaluate whether routinely collected data from the countries represented in the consortium could be collated in a common data platform. The data model developed in Objective 4 was able to perform this task. Nevertheless, it was concluded that data were not collected in a uniform manner in the partner countries that provided data (Belgium, Denmark and Italy). Optimally, data from Objective 4 could have been used to test models developed during Objective 5. One approach could have been to use models from Objective 5 to predict the welfare status of herds in the data from Objective 4. However, the biological meaningfulness of doing this was deemed low because not all of the parameters used for the models in Objective 5 were present in the Objective 4 data and further, the ABMs in the Objective 4 data was not uniformly defined. Finally, a measure of the overall welfare was not available in the data from Objective 4.

13. CONCLUSIONS

In the following sections, the main conclusions from each of the six objectives are presented. Based on these conclusions and the knowledge generated during the project a number of recommendations are given.

13.1. Objective 1

The WAEs identified in Objective 1 were: ‘Mortality – unassisted’, ‘Mortality – euthanised’, ‘Leg injuries’, ‘Foot disorders’, ‘Behavioural disruption – flooring’, ‘Behavioural disruption – Rest’, ‘Behavioural disruption – Feeding’ (in cubicles) and ‘Exhaustion’.

The ABMs already collected in the field that allow detection of these WAEs are:

- ABMs routinely collected on most farms:
    - Numbers of deaths –unassisted and euthanised
    - Evidence of mastitis
- ABMs routinely collected in only a sample of herds:
    - Numbers of foot lesions
    - Measures of lameness
    - Numbers of leg lesions/swellings

‘Numbers of deaths (unassisted and euthanised)’ should be investigated further as they are very commonly-recorded in most member states and associated with high-ranking WAEs (‘Foot disorders’, ‘Leg injuries’). ‘Evidence of mastitis’ should be investigated further as it is commonly-recorded on most farms and reported to be associated with a high-ranking WAE (‘Leg disorders’). ‘Numbers of foot lesions’, ‘Measures of lameness’, ‘Numbers of leg lesions/swellings’ should be investigated further as they are commonly-recorded on representative samples of farms and associated with high ranking or commonly-regarded WAEs (‘Foot disorders’, ‘Leg disorders’).

‘Behavioural disruption – Rest’ was identified as one of the WAEs. ‘Cleanliness score’ has been reported as indicating disruption to rest and lying. Disruption to rest may also be indicated by ‘Number of foot lesions’ and ‘Number of leg lesions/swellings’. Thus, if it is desirable to limit the number of ABMs considered, ‘Cleanliness score’ would not be recommended. Further, there was no standardised methodology used to assess cleanliness in the field.

‘Behavioural disruption – Feeding’ and ‘Exhaustion’ were also identified as WAEs. ‘Measures of nutritional status’ have been associated with these two WAEs. Consortium partners have indicated that other ABMs which have already been recommended for detecting other more highly-rated WAEs also detect ‘Behavioural disruption – Feeding’ and ‘Exhaustion’. Thus, if it is desirable to limit the number of ABMs considered, measures of nutritional status would not be recommended. Further, like other ABMs, there was no standardised methodology used to assess nutritional status in the field.

Thus, a suite of six ABMs (‘Numbers of deaths – unassisted’ and ‘Numbers of deaths – euthanised’, ‘Evidence of mastitis’, ‘Numbers of foot lesions’, ‘Measures of lameness’, ‘Numbers of leg lesions/swellings’ (hereafter called ‘Number of leg lesions’) could contribute to detecting all eight of the WAEs identified. It is proposed that these six ABMs comprise the indicators/measures to be considered in the remaining objectives of the project.

13.2. Objective 2

In Objective 2, the validity and robustness of the ABMs selected in Objective 1 was evaluated. Based on these results, three ABMs were recommended for further investigations: 1) ‘Number of deaths’ as a direct measure of ‘Mortality’. For the data analysis, it was defined as an annual mortality rate with a cut-off of 5%. It was not significantly correlated to any of the other WAEs in the data analysis. It is recommended as a useful ABM because data on mortality is deemed as both valid and robust in the literature review and finally, mortality is a routinely collected parameter in most member states. 2) SCC is recommended as a measure as it is associated with important welfare measures such as overall welfare, mortality and lameness. For the data analysis, ‘Evidence of mastitis’ was defined in two ways: Either as the geometric mean of the bulk milk SCC during the last three months (cut-off = 300000 or 400000) or as a cow level SCC calculated as proportion of SCC measures above 400000 during the last three months (cut-off = 10%). In the data analysis, an association between SCC and mortality was found. 3) ‘Measures of lameness’. In the literature, lameness scoring is widely used as a measure of painful foot lesions in cows. The validation of lameness scoring is difficult because of the lack of a gold standard measuring pain in cows. However, it is deemed as sensitive towards the most painful foot disorders such as sole ulcers, white line disease and phlegmon. In the data analysis, ‘Measures of lameness’ was defined as the proportion of moderately or severely lame cows (cut-off = 8%). In the available data, ‘Measures of lameness’ was correlated to ‘Mortality, overall’ and ‘Leg lesions’. Data on lameness might not currently be readily available in the field. Nonetheless, it was deemed an important ABM as it was correlated to two of the WAEs in the data analysis. Furthermore, future development of automatic lameness detection could make it easier to obtain routinely collected lameness data.

The remaining ABMs identified in Objective 1 still have problems with robustness and are not recommended.

13.3. Objective 3

As a result of the literature search and the consultation of members from the consortium, 10 factors of variation were identified as being the most important to collect. They were chosen for their relevance as risk factors/factors of variation for mortality, mastitis/elevated SCC and lameness, their feasibility for routine field collection and their capacity to characterise populations. These parameters, ranked by mean relevance score, were:

- Parity
- Housing system
- Floor type
- Days in milk/after calving
- Access to pasture
- Milk production/yield
- Herd size
- Breed
- Geographical region
- Organic dairy production

Of these parameters, ‘housing system’, ‘floor type’, ‘access to pasture’ and the ABM ‘lameness’ were not collected routinely in any of the countries included in this study. Collection of these parameters in the future may have the potential to improve the possibilities of an epidemiological surveillance of the welfare of dairy cows.

13.4. Objective 4

The aims of Objective 4 were to:

- Develop a model to collate routinely collected data (RCD) from different countries represented in the ANIBAM consortium, based on the elements selected in objective 3

- Perform a pilot test to collate this data

As regard to the first aim, it can be concluded that it is possible to make a model to collate routinely collected data on the ABMs and factors country, region, animalID, herdID, holdingID, year, month, day, breed, production type, housing, flooring, bedding, herd size, access to pasture, milk yield, parity, days in milk, SCC, density, annual mortality rate and lameness. This model can be found in Appendix U.

As regard to the pilot test, it can be concluded that data are not collected in a uniform manner in the partner countries that provided data (Belgium, Denmark and Italy). In Denmark mortality and SCC data are collected at the same institute, in Italy and Belgium this is not the case. Collating the data immediately at the source would make collating the data at a European level easier.

Elements that were not collected mainly related to management (e.g. access to pasture and stocking density), housing, flooring and bedding and the ABM lameness. Expert opinion indicated that these elements are highly relevant and/or related to the welfare of dairy cattle (Objective 3).

Data that were collected in all countries sometimes differed substantially in form. Consequently, several transformations were needed in order to merge the datasets. Transformations were needed on the region-element (sampArea) for Italy and Denmark, on breed for Italy and Denmark, production type for Italy, BMSCC and SCC for Italy and Denmark and on AMR for Italy.

Annual mortality rate was the easiest of the three ABMs to merge: only a minor transformation was needed for the Italian database. SCC was harder to put into one dataset. In the three datasets that were provided, three different types of SCC were present. These were bulk milk SCC (BMSCC), percentage of animals above 400,000 cells/ml (HSCC) and individual SCC (ISCC). While ISCC could be transformed in such a way that it can be merged with HSCC, BMSCC cannot. The best data to collect for SCC should be based on their relation with animal welfare, but BMSCC seems most feasible at this moment to collect and collate.

Three factors of variation needed transformation: region, breed and production type. Therefore coding was used: the NUTS coding for region, SSD coding for production type and a coding for breed seems advisable for European-wide use to be able to easily collate the data. The SSD coding for production type was not sufficient for the ANIBAM project since transhumance did not fit into any of the categories. Therefore it is advisable that a transhumance category is added to the SSD.

The overall conclusion that can be drawn from Objective 4 is that a Europe-wide database representing all partner countries will not be possible at this point in time without having to perform many transformations and without accepting a lot of missing data.

13.5. Objective 5

In Objective 5, nine different welfare assessment outcomes (IZSLER/CReNBA, Otten animal based, Otten register based, Otten system based, Burow, Welfare Quality® health principle, Welfare Quality® housing, Welfare Quality® feeding principle and Welfare Quality® behaviour principle) were used for analysing the ability of the ABMs and the factors of variation to predict the overall welfare status of the herds. Except from the IZSLER/CReNBA data, the association between the ABMs and the overall welfare outcomes were generally low and the ROC AUCs were rarely significantly larger than 0.50. Equally, the combination of the ABMs seemed beneficial in the IZSLER/CReNBA data but results from the other datasets could not confirm this finding.

When significant associations between the ABMs and the risk of classifying as a poorer welfare herd were found, results confirmed the expectation from the expert opinions in previous objectives: Increased levels of mortality, SCC and lameness were associated with poorer welfare. Where mortality and SCC were both associated with the poorer welfare outcome in five out of the thirteen different welfare measurements, lameness seemed to be slightly more sensitive as it was found significant in seven out of the thirteen scenarios.

In the IZSLER/CReNBA data, models containing the combined effects of the ABMs performed better than models including the effect of the only one ABM. Yet, this was not the case in all datasets. Generally, the conditional inference tree models did not increase the predictive value of the models compared to the logistic regression models.

Seven out of the ten factors of variation were found in at least one of the datasets. Days in milk and parity were not found in any of the datasets. Geographical region was found in the Danish data (Otten + Burow) and in the IZSLER/CReNBA data. For the remaining factors of variation (production type, housing system, floor condition, pasture access, herd size, breed and milk yield), all of them were found significantly associated with the welfare status in at least one out of the thirteen models. However, decreasing herd size and milk yield were both associated with increased risk of poorer welfare in the IZSLER/CReNBA data, whereas in the other data where herd size and milk yield were deemed important, increasing herd size and milk yield were associated with the risk of classifying as a poorer welfare herd.

13.6. Objective 6

Objective 6 discussed two factors which may be relevant to record routinely in the future: lameness and stocking density. Lameness is a serious welfare problem. Routine recordings of lameness are not widely available. An evaluation of lameness can be done in many different ways and the methodology described in the present report is just an example of how it could be done. Routine recordings of lameness may have the potential to improve the precision of assessments of the overall welfare status in dairy herds in the future. Further research in needed regarding the relationship between lameness and overall welfare.

There are different measures of density of which the exact impact on animal welfare is depending on other factors. It is not possible to define one or a few most relevant parameters of density. Stocking density in dairy cow herds is neither straightforward to record in a meaningful and comprehensive way nor easy to interpret in relation to animal welfare. The use of density as a measure of overall welfare is therefore not recommended.

13.7 RECOMMENDATIONS

The relationship between selected ABMs and overall welfare at the herd level is complex and still not sufficiently studied. In this project, the association between ABMs and overall welfare differed between different datasets. Therefore, a system using routinely collected ABMs to predict the overall welfare at herd level in dairy herds does not seem realistic based on the results from this project.

The lack of a gold standard for overall animal welfare at the herd level in dairy herds makes it very difficult to quantify the ’true status’ regarding animal welfare at herd level.

As ABMs in many cases are used to ’calculate’ the overall welfare at herd level, there is a risk that a ’self-fulfilling prophecy’ may occur: A high prevalence of a certain ABM (or combination of ABMs) may be associated with the overall welfare partly because the ABM(s) are ’a mathematical part’ of the overall welfare. This should always be kept in mind when such associations are evaluated.

The use of a data model like the one presented in this project seems like a feasible way to collate routinely collected data on ABMs as well as factors of variation. However, routine recordings of ABMs should be standardised as much as possible between countries. If routinely collected data on ABMs are to be used to evaluate overall welfare – and perhaps compare overall welfare between e.g. countries – the routinely collected data should preferably be in a standardized format allowing them to be inserted into a database without the need for transformations etc. as this will both decrease the workload and minimize the risk of errors.

REFERENCES

Alvåsen, K., M. J. Mörk, C. H. Sandgren, P. T. Thomsen, and U. Emanuelson. 2012. Herd-level risk factors associated with cow mortality in Swedish dairy herds. Journal of Dairy Science 95, 4352-4362.

Anonymous. 2001. Interdisciplinary report "housing design for cattle – Danish recommendations. Third edition 2001". The Danish agricultural advisory centre. Translated into English and issued in 2002. 122 pp.

Bartussek, H. 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. Livestock Production Science 61(2-3), 179-192.

Bartussek, H., C. Leeb and S. Held. 2000. Animal Needs Index for Cattle. ANI 35, L/2000. http://www.bartussek.at/veroeffentlichungen/511134991b0db8204/index.html.

Bennett, R. M., Z. E. Barker, D. C. J. Main, H. R. Whay and K. A. Leach. 2014. Investigating the value dairy farmers place on a reduction of lameness in their herds using a willingness to pay approach. Veterinary Journal 199 (1), 72-75.

Bicalho, R. C. and G. Oikonomou. 2013. Control and prevention of lameness associated with claw lesions in dairy cows. Livestock Science, Article in press.

Bicalho, R. C., S. H. Cheong, G. Cramer and C. L. Guard. 2007. Association between a visual and an automated locomotion score in lactating Holstein cows. Journal of Dairy Science 90, 3294-3300.

Borderas, T. F., A. Fournier, J. Rushen and A. M. B. De Passillé. 2008. Effect of lameness on dairy cows' visits to automatic milking systems. Canadian Journal of Animal Science 88, 1-8.

Botreau, R., M. Bonde, A. Butterworth, P. Perny, M. B. M. Bracke, J. Capdeville, and I. Veissier. 2007. Aggregation of measures to produce an overall assessment of animal welfare. Part 1: a review of existing methods. Animal 1 (8), 1179-1187.

Brenninkmeyer, C. and C. Winckler. 2012. Scientific report submitted to EFSA: Relationships between animal welfare hazards and animal based welfare indicators. Available online: http://www.efsa.europa.eu/en/supporting/doc/253e.pdf

Brenninkmeyer, C., S. Dippel, S. March, J. Brinkmann, C. Winckler and U. Knierim, 2007. Reliability of a subjective lameness scoring system for dairy cows. Animal Welfare 16, 127-129.

Broom, D. M. 1986. Indicators of poor welfare. British Veterinary Journal 142 (6), 524-526.

Broom, D. M. 1996. Animal Welfare Defined in Terms of Attempts to Cope with the Environment. Acta Agriculturae Scandinavica, Section A - Animal Science Supplementum 27, 22-28.

Broom, D. M. 2010. Animal Welfare. An Aspect of Care, Sustainability, and Food Quality Required by the Public. Journal of Veterinary Medical Education 37 (1), 83-88.

Bruijnis, M. R. N., B. Beerda, H. Hogeveen and E. N. Stassen. 2012. Assessing the welfare impact of foot disorders in dairy cattle by a modeling approach. Animal 6, 962-970.

Burnside, E. B., S. B. Kowalchuk, D. B. Lambroughton and N. M. MacLeod. 1971. Canadian dairy cow disposals. I. Differences between breeds, lactation numbers and seasons. Canadian Journal of Animal Science 1971, 75-83.

Burow, E., T. Rousing, P. T. Thomsen, N. D. Otten, and J. T. Sørensen. 2013. Effect of grazing on the cow welfare of dairy herds evaluated by a multidimensional welfare index. Animal 7 (5), 834-842.

Calderon, D. F. and N. B. Cook. 2011. The effect of lameness on the resting behavior and metabolic status of dairy cattle during the transition period in a freestall-housed dairy herd. Journal of Dairy Science 94, 2883-2894.

Capdeville, J. and I. Veissier. 2001. A Method of Assessing Welfare in Loose Housed Dairy Cows at Farm Level, Focusing on Animal Observations. Acta Agriculturae Scandinavica, Section A - Animal Science 51 (1 supp. 30), 62-68.

Capion, N., S. M. Thamsborg and C. Enevoldsen. 2008. Prevalence of foot lesions in Danish Holstein cows. The Veterinary Record 163, 80-85.

Chapinal, N., A. M. de Passillé and J. Rushen. 2010. Correlated changes in behavioral indicators of lameness in dairy cows following hoof trimming. Journal of Dairy Science 93, 5758-5763.

Chapinal, N., A.M. de Passillé, D. M. Weary, M. A. G. von Keyserlingk and J. Rushen. 2009. Using gait score, walking speed, and lying behavior to detect hoof lesions in dairy cows. Journal of Dairy Science 92, 4365-4374.

Chapinal, N., A. M. de Passillé, M. Pastell, L. Hänninen, L. Munksgaard, and J. Rushen. 2011. Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle. Journal of Dairy Science 94, 2895-2901.

CIGR, 1994. The design of dairy cow housing, Vol. section 2.

Clutton-Brock, J. 1981. Domesticated animals from early times. British Museum Press, London, UK.

Cook, N. B., R. L. Mentink, T. B. Bennett and K. Burgi. 2007. The effect of heat stress and lameness on time budgets of lactating dairy cows. Journal of Dairy Science 90, 1674-1682.

De Rosa, G., C. Tripaldi, F. Napolitano, F. Saltalamacchia, F. Grasso, V. Bisegna and A. Bordi. 2003. Repeatability of some animal-related variables in dairy cows and buffaloes. Animal Welfare 12, 625-629.

DeLong, E. R., D. M. DeLong and D. L. Clarke-Pearson. 1988. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 44 (3), 837-845.

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

Dippel, S., C. B. Tucker, C. Winckler and D. M. Weary. 2011. Effects of behaviour on the development of claw lesions in early lactation dairy cows. Applied Animal Behaviour Science 134, 16-22.

Dohoo, I. R. and K. E. Leslie. 1991. Evaluation of changes in somatic cell counts as indicators of new intramammary infections. Preventive Veterinary Medicine 10, 225–237.

Dohoo, I., W. Martin, and H. Stryhn. 2003. Veterinary Epidemiological Research. AVC Inc., Charlottetown, Canada

Dyer, R. M., N. K. Neerchal, U. Tasch, Y. Wu, P. Dyer and P. G. Rajkondawar. 2007. Objective Determination of claw pain and its relationship to limb locomotion score in dairy cattle. Journal of Dairy Science 90, 4592-4602.

EFSA. 2009. Effects of farming systems on dairy cow welfare and disease, pp. 1–284.

EFSA. 2012. Statement on the use of animal based measures to assess the welfare of animals. EFSA Journal 10 (6), 2767.

Espetvedt, M. N., C. Wolff, S. Rintakoski, A. Lind and O. Østerås. 2012. Completeness of metabolic disease recordings in Nordic national databases for dairy cows. Preventive Veterinary Medicine 105, 25-37.

Esslemont, R. J. and M. A. Kossaibati. 1997. Culling in 50 dairy herds in England. Veterinary Record 140, 36-39.

Ettema, J., S. Østergaard, A. R. Kristensen. 2010. Modelling the economic impact of three lameness causing diseases using herd and cow level incidences. Preventive Veterinary Medicine 95, 64–73.

EU. 1997. Treaty of Amsterdam amending the Treaty on European Union, the Treaties establishing the European Communities and certain related acts. Page 145 in Official Journal of the European Communities, Volume 40. Vol. 97/C 340/01. E. Union, ed., Official Journal of the European Communities.

European Food Safety Authority. 2009a. Scientific report of EFSA prepared by the Animal Health and Animal Welfare Unit on the effects of farming systems on dairy cow welfare and disease. Annex to the EFSA Journal (2009) 1143, 1-38. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/1143r.pdf

European Food Safety Authority. 2009b. Scientific Opinion of the Panel on Animal Health and Welfare on a request from European Commission on the overall effects of farming systems on dairy cow welfare and disease. The EFSA Journal (2009) 1143, 1-38. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/1143.pdf

European Food Safety Authority. 2009c. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on leg and locomotion problems in dairy cows. The EFSA Journal (2009) 1142, 1-57. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/1142.pdf

European Food Safety Authority. 2009d. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on udder problems in dairy cows. The EFSA Journal (2009) 1141, 1-60. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/1141.pdf

European Food Safety Authority. 2009e. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on metabolic and reproductive problems in dairy cows. The EFSA Journal (2009) 1140, 1-75. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/1140.pdf

European Food Safety Authority. 2009f. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on behaviour, fear and pain problems in dairy cows. The EFSA Journal (2009) 1139, 1-68. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/1139.pdf

European Food Safety Authority. 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010 8(6), 1637. 90 pp. doi:10.2903/j.efsa.2010.1637. Available online: www.efsa.europa.eu

European Food Safety Authority. 2012a. EFSA Panel on Animal Health and Welfare (AHAW); Scientific Opinion on the use of animal based measures to assess welfare of dairy cows. EFSA Journal EFSA Journal 2012 10(1), 2554. Available online: http://www.efsa.europa.eu/en/efsajournal/doc/2554.pdf

European Food Safety Authority. 2012b. Technical meeting on the use of animal based measures for the welfare of dairy cows, pigs and broilers. Supporting Publications 2012: EN-341. Available online: http://www.efsa.europa.eu/en/supporting/doc/341e.pdf

Farm Animal Welfare Advisory Committee, 1979. Press statement. Derived from http://www.fawc.org.uk/freedoms.htm)

FAWC, 1979. Press notice: Five Freedoms, http://www.fawc.org.uk/pdf/fivefreedoms1979.pdf. Released by FAWC, December 1979

Faye, B. and L. Perochon. 1995. La mortalité des vaches laitiéres dans l'enquête écopathologique Bretagne. Veterinary Research 26, 124-131.

Fleiss, J.L. 1981. Statistical methods for rates and proportions (2nd ed.). New York: John Wiley. ISBN 0-471-26370-2.

Flower, F.C. and D. M. Weary. 2006. Effect of hoof pathologies on subjective assessments of dairy cow gait. Journal of Dairy Science 89, 139-146.

Flower, F.C. and D. M. Weary. 2009. Gait assessment in dairy cattle. Animal 3, 87-95.

Frankena, K., J. G. C. J. Somers, W. G. P. Schouten, J. V. van Stek, J. H. M. Metz, E. N. Stassen and E. A, M. Graat. 2009. The effect of digital lesions and floor type on locomotion score in Dutch dairy cows. Preventive Veterinary Medicine 88, 150-157.

Fraser, D. and D. M. Broom. 1990a. Welfare measurement. Pages 266-279 in Farm animal behaviour and welfare. Third ed. Bailiére Tindall, London.

Fraser, D. and D. M. Broom. 1990b. Welfare Terminology and Concepts. Pages 256-265 in Farm Animal Behaviour and Welfare. Third Edition ed. D. Fraser and D. M. Broom, ed. Bailliére Tindall, London.

Fraser, D., D. M. Weary, E. A. Pajor, and B. N. Milligan. 1997. A Scientific Conception of Animal Welfare that Reflects Ethical Concerns. Animal Welfare 6(3), 187-205.

Gardner, I. A., D. W. Hird, W. W. Utterback, C. Danaye-Elmi, B. R. Heron, K. H. Christiansen and V. M. Sischo. 1990. Mortality, morbidity, case-fatality, and culling rates for California dairy cattle as evaluated by the national animal health monitoring system, 1986-87. Preventive Veterinary Medicine 8, 157-170.

Gibbons J, E. Vasseur, J. Rushen and A. M. de Passille. 2012. A training programme to ensure high repeatability of injury scoring of dairy cows. Animal Welfare 21, 379-388.

Gibbons, J. M., A. B. Lawrence and J. M. Haskell. 2009. Consistency of aggressive behaviour in dairy cows. Applied Animal Behaviour Science 121, 1-7.

Gomez, A. and N. B. Cook. 2010. Time budgets of lactating dairy cattle in commercial freestall herds. Journal of Dairy Science 93, 5772-5781.

Harris, B. L. 1989. New Zealand dairy cow removal reasons and survival rate. New Zealand Journal of Agricultural Research 32, 355-358.

Haskell, M. J., L. J. Rennie, V. A. Bowell, M. J. Bell and A. B. Lawrence. 2006. Housing System, Milk Production, and Zero-Grazing Effects on Lameness and Leg Injury in Dairy Cows. Journal of Dairy Science 89, 4259–4266.

Hillerton, J.E. 1999. Redefining mastitis based on somatic cell count. International Dairy Federation, Brussels, Belgium. Bulletin 345/1999, 4–6.

Hillerton, J.E. 2000. Detecting mastitis cow-side. National Mastitis Council. Annual Meeting 39, 48–53.

Houe, H. and T. B. Jensen. 2012. Danske databaser om husdyr. Rapport vedrørende VID-projekt 2010-01: "Databaser – Potentialer for dyrevelfærdsvurdering" (in Danish)

Houe, H., A. K. Ersbøll and N. Toft (editors). 2004. Introduction to Veterinary Epidemiology. Biofolia, Frederiksberg, Denmark. 398 pp.

Houe, H., I. A. Gardner and L. R. Nielsen. 2011. Use of information on disease diagnoses from databases for animal health economic, welfare and food safety purposes: strengths and limitations of recordings. Acta Veterinaria Scandinavica, 53 (suppl. 1), S7, 5 pp.

Houe, H., T. B. Jensen, A. C. Nielsen, H. Feldstedt, P. Sandøe, B. Forkman, N. Otten, S. N. Andreasen, T. D. Nielsen, N. Toft and J. T. Sørensen. 2012. Databases in Denmark: opportunities and constraints for welfare assessments. Book of Abstracts of the 13th International Symposium on Veterinary Epidemiology and Economics, 2012, Maastricht, 136.

Houe, H., P. Sandøe and P. T. Thomsen. 2011. Welfare assessments based on lifetime health and production data in Danish dairy cows. Journal of Applied Animal Welfare Science 14, 255-264.

Huxley, J. N., J. Burke, S. Roderick, D. C. J. Main and H. R. Whay. 2004. Animal welfare assessment benchmarking as a tool for health and welfare planning in organic dairy herds. The Veterinary Record 155, 3.

International Committee for Animal Recording (ICAR). 2012. International agreement of recording practices approved by the General Assembly held in Cork, Ireland, on June 2012.

Ito, K., M. A. G. von Keyserlingk, S. J. LeBlanc, D. M. Weary. 2010. Lying behavior as an indicator of lameness in dairy cows. Journal of Dairy Science 93, 3553-3560.

Jensen, K. K. and P. Sandøe. 2013. Hvor god er dyrevelfærden samlet set? - Om aggregering af velfærd. Pages 41-49 in Vurdering af dyrevelfærd i en husdyrbesætning. J. T. Sørensen, H. Houe, T. Rousing, and P. Sandøe, ed. Center for Bioetik og Risikovurdering, Frederiksberg C, Denmark (in Danish).

Leach, K. A., H. R. Whay, C. H. Maggs, Z. E. Barker, E. A. Paul, A. K. Bell and D. C. J. Main. 2010. Working towards a reduction in cattle lameness: 1. Understanding barriers to lameness control on dairy farms. Research in Veterinary Science 89, 311-317.

Lee, C. S., F. B. P. Wooding and P. Kemp. 1980. Identification, properties, and differential counts of Cell-populations using electron-microscopy of dry cow secretions, colostrum and milk from normal cows. Journal of Dairy Research 47, 39-50.

Ley, S. J., A. E. Waterman and A. Livingston. 1996. Measurement of mechanical thresholds, plasma cortisol and catecholamines in control and lame cattle: A preliminary study. Research in Veterinary Science 61, 172-173.

Lind, A., P. T. Thomsen, A. K. Ersbøll, M. N. Espetvedt, C. Wolff, S. Rintakoski and H. Houe. 2012. Validation of Nordic dairy cattle disease recording databases—Completeness for locomotor disorders. Preventive Veterinary Medicine 107, 204-213.

March. S., J. Brinkmann and C. Winkler. 2007. Effect of training on the inter-observer reliability of lameness scoring in dairy cattle. Animal Welfare 16, 131-133.

Menzies, F. D., D. G. Bryson, T. McCallion and D. I. Matthews. 1995. A study of mortality among suckler and dairy cows in Northern Ireland in 1992. Veterinary Record 137, 531-536.

Miguel-Pacheco, G. G., J. Kaler, J. Remnant, L. Cheyne, C. Abbott, A. P. French, T. P. Pridmore and J. N. Huxley. 2014. Behavioural changes in dairy cows with lameness in an automatic milking system. Applied Animal Behaviour Science 150, 1-8.

Milian-Suazo, F., H. N. Erb and R. D. Smith. 1988. Descriptive Epidemiology of Culling in Dairy Cows from 34 Herds in New York State. Preventive Veterinary Medicine 6, 243-251.

Molony, V. and J. E. Kent. 1997. Assessment of acute pain in farm animals using behavioral and physiological measurements. Journal of Animal Science 75, 266-272.

Mörk, M., A. Lindberg, S. Alenius, I. Vågsholm and A. Egenvall. 2009. Comparison between dairy cow disease incidence in data registered by farmers and in data from a disease-recording system based on veterinary reporting. Preventive Veterinary Medicine 88, 298-307.

Nielsen, S. S., N. Toft and I. A. Gardner. 2011. Structured approach to design of diagnostic test evaluation studies for chronic progressive infections in animals. Veterinary Microbiology 150, 115-125.

Olmos, G., L. Boyle, A. Hanlon, J. Patton, J. J. Murphy and J. F. Mee. 2009. Hoof disorders, locomotion ability and lying times of cubicle-housed compared to pasture-based dairy cows. Livestock Science 125, 199-207.

Otten, N. D. 2014. Identification of dairy herds with animal welfare problems. Page 163 in Ph.D.-thesis, Department of Large Animal Sciences, Faculty of Health and Medical Sciences. University of Copenhagen, Frederiksberg C, Denmark.

Pastell, M. E. and M. Kujala. 2007. A probabilistic neural network model for lameness detection. Journal of Dairy Science 90, 2283–2292.

Pavlenko, A., C. Bergsten, I. Ekesbo, T. Kaart, A. Aland and L. Lidfors. 2011. Influence of digital dermatitis and sole ulcer on dairy cow behaviour and milk production. Animal 5, 1259-1269.

Plesch, G., N. Broerkens, S. Laister, C. Winckler and U. Knierim. 2010. Reliability and feasibility of selected measures concerning resting behaviour for the on-farm welfare assessment in dairy cows. Applied Animal Behaviour Science 126, 19-26.

Presi, P. and M. Reist. 2011. Scientific Report prepared for EFSA: Review of methodologies applicable to the validation of animal based indicators of welfare. Available online: http://www.efsa.europa.eu/en/supporting/doc/171e.pdf

Prunier, A., L. Mounier, P. Le Neindre, C. Leterrier, P. Mormède, V. Paulmier, P. Prunet, C. Terlouw and R. Guatteo. 2013. Identifying and monitoring pain in farm animals: a review. Animal 7, 998-1010.

Radostits, O. M., C. C. Gay, K. W. Hinchcliff and P. D. Constable. 2007. Veterinary Medicine. A textbook of the diseases of cattle, horses, sheep, pigs and goats. 10th edition. Saunders Elsevier, 2156 pp.

Rintakoski, S., J. Taponen, O. A. T. Peltoniemi and A. M. K. Virtala. 2012. Validation of the Finnish national dairy disease register—Data transfer from cow health cards to the disease register. Journal of Dairy Science 95, 4309-4318.

RSPCA. 2011. RSPCA welfare standards for dairy cattle. Page 66. www.rspca.org.uk.

Rushen, J., E. Pombourcq and A. M. de Passillé. 2007. Validation of two measures of lameness in dairy cows. Applied Animal Behaviour Science 106, 173-177.

Rutherford, K. M. D., F. M. Langford, M. C. Jack, L. Sherwood, A. B. Lawrence and M. J. Haskell. 2008. Hock injury prevalence and associated risk factors on organic and nonorganic dairy farms in the United Kingdom. Journal of Dairy Science 91, 2265-2274.

Sandøe, P. and S. B. Christiansen. 2008. The changing face of animal ethics. Pages 1-14 in Ethics of animal use. Blackwell Publishing, Oxford.

Sargeant, J. M., K. E. Leslie, J. E. Shirley, B. J. Pulkrabek and G. H. Lim. 2001. Sensitivity and specificity of somatic cell count and California Mastitis Test for identifying intramammary infection in early lactation. Journal of Dairy Science 84, 2018–2024.

Saunders. 2007. Days in milk. Saunders comprehensive veterinary disctionary. Elsevier, Inc.

Schaeffer, L., C. Minder, I. McMillan and E. Burnside. 1977: Nonlinear techniques for predicting of 305-day lactation production of Holsteins and Jersey. Journal of Dairy Science 60, 1636–1644.

Schepers, A. J., T. J. G. M. Lam, Y. H. Schukken, J. B. M. Wilmink and W. J. A. Hanekamp. 1997. Estimation of variance components for somatic cell counts to determine thresholds for uninfected quarters. Journal of Dairy Science 80, 1833-1840.

Scott, E. M., A. M. Nolan, and J. L. Fitzpatrick. 2001. Conceptual and Methodological Issues Related to Welfare Assessment: A Framework for Measurement. Acta Agriculturae Scandinavica, Section A - Animal Science 51(1 supp 30), 5-10.

Sharma, N., N. K. Singh and M. S. Bhadwal. 2011. Relationship of somatic cell count and mastitis: An overview. Asian-Australian Journal of Animal Science 24, 429–438.

Sogstad, Å., T. Fjeldaas and O. Østerås. 2012. Locomotion score and claw disorders in Norwegian dairy cows, assessed by claw trimmers. Livestock Science 144, 157-162.

Song, X., T. Leroy, E. Vranken, W. Maertens, B. Sonck, and D. Berckmans. 2008. Automatic detection of lameness in dairy cattle - Vision-based trackway analysis in cow's locomotion. Computers and Electronics in Agriculture 64, 39–44.

Sørensen, J. T. and D. Fraser. 2010. On-farm welfare assessment for regulatory purposes: Issues and possible solutions. Livestock Science 131(1), 1-7.

Sørensen, J. T., P. Sandøe, and N. Halberg. 2001. Animal Welfare as One among Several Values to be Considered at Farm Level: The Idea of an Ethical Account for Livestock Farming. Acta Agriculturae Scandinavica, Section A – Animal Science 51(sup030), 11-16.

Spoolder, H., G. De Rosa, B. Hörning, S. Waiblinger, and F. Wemelsfelder. 2003. Integrating parameters to assess on-farm welfare. Animal Welfare 12 (4), 529-534.

Stevenson, M. A. and I. J. Lean. 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 483-488.

Tadich, N., E. Flor and L. Green. 2010. Associations between hoof lesions and locomotion score in 1098 unsound dairy cows. The Veterinary Journal 184, 60-65.

Tadich, N., C. Tejeda, S. Bastias, C. Rosenfeld and L. E. Green. 2013. Nociceptive threshold, blood constituents and physiological values in 213 cows with locomotion scores ranging from normal to severely lame. Veterinary Journal, article in press, http://dx.doi.org/10.1016/j.tvjl.2013.01.029.

Telezhenko, E., C. Bergsten. 2005. Influence of floor type on the locomotion of dairy cows. Applied Animal Behaviour Science 93, 183–197.

Thomsen, P. T. and N. P. Baadsgaard. 2006. Intra- and inter-observer agreement of a protocol for clinical examination of dairy cattle. Preventive Veterinary Medicine 75, 133-139.

Thomsen, P. T. and H. Houe. 2006. Dairy cow mortality. A review. Veterinary Quarterly 28, 122-129.

Thomsen, P. T., A. M. Kjeldsen, J. T. Sørensen and H. Houe. 2004. Mortality (including euthanasia) among Danish dairy cows (1990-2001). Preventive Veterinary Medicine 62, 19-33.

Thomsen, P. T., L. Munksgaard and J. T. Sørensen. 2012. Locomotion scores and lying behaviour are indicators of hoof lesions in dairy cows. The Veterinary Journal 193, 644-647.

Thomsen, P. T., L. Munksgaard and F. A. Tøgersen. 2008. Evaluation of a lameness scoring system for dairy cows. Journal of Dairy Science 91, 119-126.

Thomsen, P. T., S. Østergaard, J. T. Sørensen and H. Houe. 2007. Loser cows in Danish dairy herds: Definition, prevalence and consequences. Preventive Veterinary Medicine 79, 116-135.

Toft, N., J. F. Agger, H. Houe and J. Bruun. 2004. Measures of disease frequency. In: Houe H, Ersbøll AK and Toft N (editors). Introduction to Veterinary Epidemiology. Biofolia, Frederiksberg, Denmark, pp. 77-93.

Toma, B., B. Dufour, M. Sanaa, J. J. Bénet, A. Shaw, F. Moutou and A. Louza. 2001. Epidémiologie appliquée à la lutte collective contre les maladies animales transmissibles majeures. (Epidemiology applied to the collective control of major transmissible diseases of animals). Association pour l'étude de l'épidémiologie des maladies animales, Maisons-Alfort, France. 696 pp.

Tuyttens, F. A. M., M. Sprenger, A. Van Nuffel, W. Maertens and S. Van Dongen. 2009. Reliability of categorical versus continuous scoring of welfare indicators: lameness in cows as a case study. Animal Welfare 18, 399-405.

Van Hertem, T., S. Viazzi, M. Steensels, E. Maltz, A. Antler, V. Alchanatis, A. A. Schlageter-Tello, K. Lokhorst, E. C. B. Romanini, C. Bahr, D. Berckmans, I. Halachmi. 2014. Automatic lameness detection based on consecutive 3D-video recordings. Biosystems Engeneering 119, 108-116.

von Keyserlingk, M. A. G., A. Barrientos, K. Ito, E. Galo and D. M. Weary. 2012. Benchmarking cow comfort on North American freestall dairies: Lameness, leg injuries, lying time, facility design, and management for high-producing Holstein dairy cows. Journal of Dairy Science 95, 7399-7408.

Weary, D. M., J. M. Huzzey, and M. A. G. von Keyserlingk. 2009. Board-invited review: Using behavior to predict and identify ill health in animals. Journal of Animal Science 87, 770–777.

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

Whay, H. R., D. C. J. Main, A. J. F. Webster and L. E. Green. 2003. Assessment of the welfare of dairy cattle using animal based measurements: direct observations and investigation of farm records. The Veterinary Record 153, 197-202.

Whay, H. R., A. E. Waterman, A. J. F. Webster and J. K. O'Brien. 1998. The influence of lesion type on the duration of hyperalgesia associated with hind limb lameness in dairy cattle. The Veterinary Journal 156, 23-29.

Whay, H. R., D. C. J. Main, A. J. F. Webster, and L. E. Green. 2003c. Assessment of the welfare of dairy cattle using animal based measurements: direct observations and investigation of farm records. The Veterinary Record 153 (7), 197-202.

Whay, H. R., D. C. J. Main, L. E. Green, and A. J. F. Webster. 2003a. Animal based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. Animal Welfare 12, 205-217.

Whay, H. R., D. C. J. Main, L. E. Green, and A. J. F. Webster. 2003b. An animal based welfare assessment of group-housed calves on UK dairy farms. Animal Welfare 12 (4), 611-617.

Whay, H.R. 2002. Locomotion scoring and lameness detection in dairy cattle. In Practice 24, 444-449.

White, J. M. and J. R. Nichols. 1965. Reasons for disposal of Pennsylvania Holstein cattle. Journal of Dairy Science 48, 512-514.

Winckler, C. and S. Willen. 2001. The Reliability and Repeatability of a Lameness Scoring System for Use as an Indicator of Welfare in Dairy Cattle. Acta agriculturae Scandinavica, Section A, Animal science 51(sup030), 103-107.

Winckler, C., J. Brinkmann and J. Glatz. 2007. Long-term consistency of selected animal-related welfare parameters in dairy farms. Animal Welfare 16, 197-199.

Wolff, C., M. Espetvedt, A. K. Lind, S. Rintakoski, A. Egenvall, A. Lindberg and U. Emanuelson, 2012. Completeness of the disease recording systems for dairy cows in Denmark, Finland, Norway and Sweden with special reference to clinical mastitis. BMC Veterinary Research 8.

Zurbrigg, K., D. Kelton, N. Anderson and S. Millman. 2005. Tie-stall design and its relationship to lameness, injury and cleanliness on 317 Ontario dairy farms. Journal of Dairy Science 88, 3201-3210.

Østerås, O. 2003. Naturlige variationer i celletal. Intern rapport nr. 177 – Maj 2003. Bilag til temamøde den 14. maj 2003 på Forskningscenter Foulum. Celletal og yversundhed (in Danish).

APPENDICES

Appendix A. The full list of adverse effects considered for pre-selection of the worst adverse effects

Adverse effects
1. Foot disorders & associated effects • lameness • locomotion problems • claw disorders • injuries • laminitis • infectious foot disorders • disease transmission (e.g. digital dermatitis)
2. Leg injuries & associated effects • leg injuries
3. Clinical Mastitis & associated conditions • trauma associated with mastitis • localised Mastitis • Systemic Mastitis • contagious mastitis • chronic mastitis • summer mastitis • udder trauma • teat disorders • discomfort
4. Metabolic disorders • ketosis • acute ruminal acidosis • sub-acute ruminal acidosis (SARA) • fatty liver • milk fever • prolonged metabolic disease/delayed treatment of disease • risk of improper treatment • metabolic stress/ metabolic stress due to disease
5. Reproductive disorders • less oestrus expression • reduced fertility • reproductive failure • dystocia • metritis • infections • induced abortion/abortion • perinatal mortality
6. Mortality (unassisted)

7. Mortality (euthanised)

Adverse effects extracted from EFSA reports (EFSA 2009a-f, 2012 a, b)
8. Mortality (slaughtered, including prior to end of expected productive life)
9. Exhaustion associated with prolonged high metabolic demand
10. Behavioural disruption-feeding • chronic hunger • reduced feed intake • excessive feed intake • suboptimal feed intake • reduced dry matter intake • social stress • pain • hunger • exhaustion • fear • frustration
11. Behavioural disruption-rest • too little rest • pain • fear
12. Behavioural disruption -flooring conditions/space (including fear, pain)
13. Thermal discomfort (including cold and heat)
14. Pain or fear (due to handling, milking, dehorning/other 'surgery', downer cows) or frustration (due to management factors)
15. Respiratory distress, pain, discomfort - air quality
16. Abomasal displacement
17. Other diseases • immunosuppression • risk of improper treatment • poor recovery (sick/injured animals)

18. Other Injuries

- frustration
- social stress
- inhibited social behaviour
- thirst/ suppressed drinking
- dehydration
- fear
- behaviour disorders
- altered behaviour

to goads

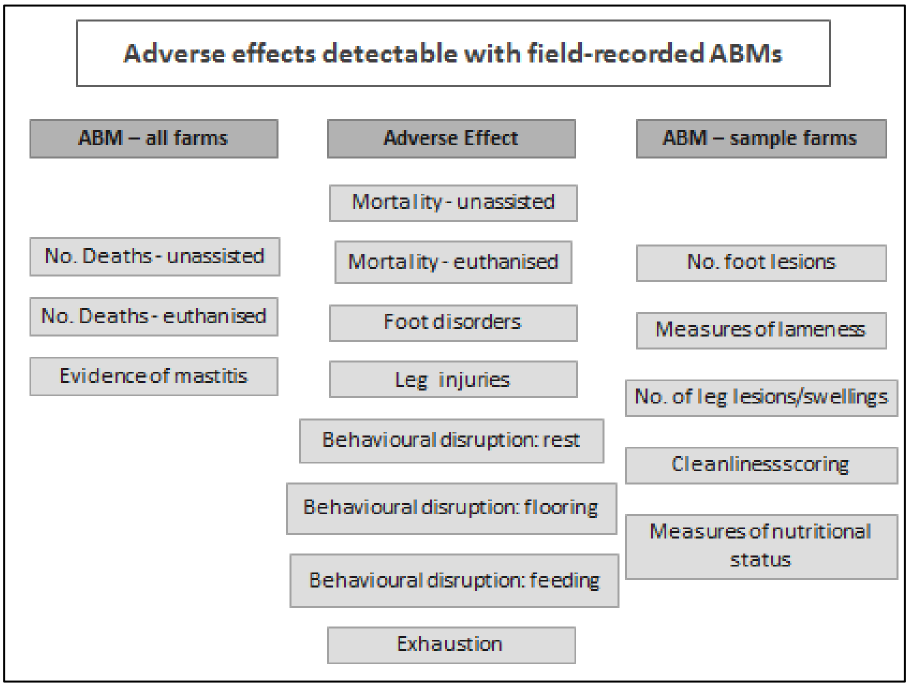
- reduced rumination
- inappetance
- inability to carry out maintenance behaviour
- increased constraint on time available for activities
- discomfort as a result of full udder
- inability to carry out normal behaviour
- behaviour disturbance

Appendix B. Preselected list of adverse effects

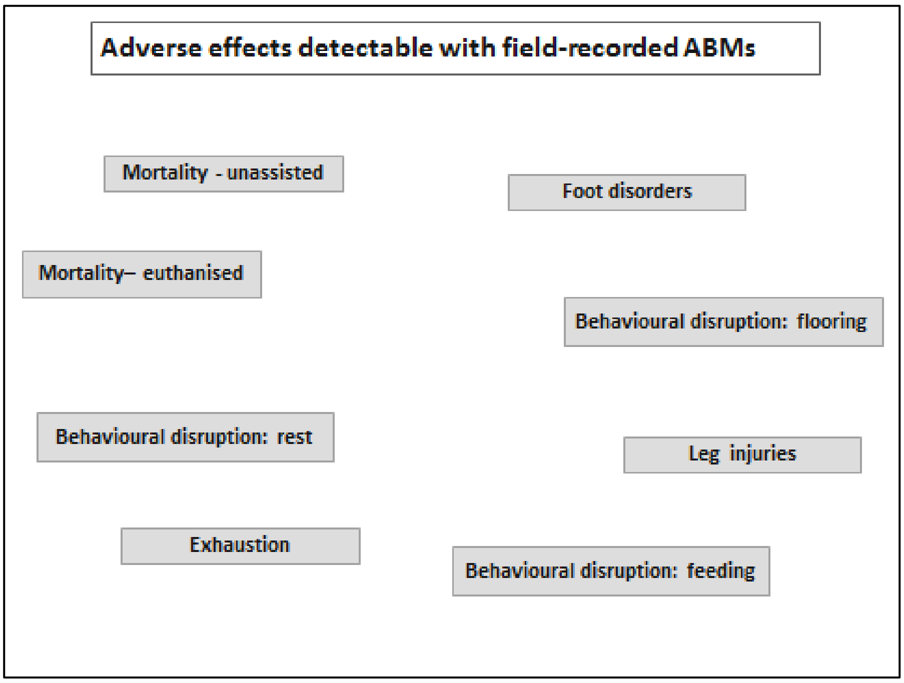
1	Foot disorders & associated effects e.g. lameness
2	Leg injuries & associated effects
3	Clinical Mastitis & associated conditions
4	Metabolic disorders e.g. SARA (sub-acute ruminal acidosis), lipomobilsation syndrome
5	Reproductive disorders e.g. reduced fertility, dystocia, infections
6	Mortality (unassisted)
7	Mortality (euthanised)
8	Mortality (slaughtered, including prior to end of expected productive life)
9	Exhaustion associated with prolonged high metabolic demand
10	Behavioural disruption-feeding (including social stress, pain, hunger, exhaustion, fear, frustration)
11	Behavioural disruption-rest (incl too little rest, pain, fear)
12	Behavioural disruption -flooring conditions/space (including fear, pain)
13	Thermal discomfort (including cold and heat)
14	Pain or fear (due to handling, milking, dehorning/other 'surgery', downer cows) or frustration (due to management factors)
15	Respiratory distress, pain, discomfort - air quality
16	Abomasal displacement

Appendix C. Schematic diagrams used to help elicit the relationships between ABMs and WAE (a), and between WAE (b).

a)



b)



Appendix D. The additional descriptions provided by the respondents for each ABM, and the frequency of recording each description, for each ABM collected routinely on most farms by qualified personnel.

ABM and additional descriptions provided	Frequency
Avoidance distance	**1**
% of animals that avoid human contact and escape	1
Avoidance test	**1**
Approach test	1
Behaviour	**1**
Withdrawal when observer approaches the manger, voluntary animal approach test, avoidance test	1
Calving ease	**1**
calving ease with a 5 points scoring system	1
Cleanliness score	**9**
% of dirty animals	1
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
Cleanliness score	1
Individual measure: 20 cows selected at random and visually assessed a) lower hind legs (above the coronary band), including the hock b) hind quarters – upper hind leg, flank and rear view, including tail (excluding udder) c) udder Scoring a, b, or c: 0 = Clean = No dirt or only minor splashing present (except teats which must have no minor splashing present); 1 = Dirty = An area of dirtiness (i.e. layer or plaques of fresh or dried dirt) amounting to at least palm size (10x15cm). Or, any dirt (including minor splashing) on or around the teats. Do not score stained hair; 2 = Very dirty = An area of dirtiness (i.e. layer or plaques of dirt) amounting to at least forearm length (40cm) in any dimension	1
Mud on body or on different parts of the body	1
Observation	1
Only at slaughter	1
WQ® protocol	1
Cows needing further care	**1**
Herd measure - Assess the whole herd and record number of any sick or injured cows that would benefit from further intervention. Further interventions could include further treatment, hospitalisation (i.e. removal from the main herd) or culling. Assess animals across the herd including the milking herd, dry cows, in-calf heifers, calves, hospital pens and animals that are due to leave the farm.	1
Evidence of discomfort when standing (e.g. time resting a foot)	**4**
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
time resting a foot	1
Evidence of mastitis (e.g. somatic cell count)	**12**
bulk tank somatic cell count as evidence of mastitis	1
Analysis quality control	1
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
Farm Records measure: Check farm records and record the number of individual clinical cases of mastitis that received treatment of any kind treated in the past 12 months.	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
individual milk control data on each individual cow every 5-6 weeks include data on individual cow somatic cell count	1
SCC	1
SCC - sometimes just at herd level	1

ABM and additional descriptions provided	Frequency
somatic cell count, Injuries to teats, udder and blind quarters	1
somatic cell counts (1.2 million)	1
SSC and health records	1
Faeces score	**1**
Shidmore et al., 1996	1
Fertility status index	**1**
Days open, calving interval, pregnancy rate, culling rate	1
Frequency of agonistic behaviour (e.g. numbers of chasing up from lying)	**3**
Assurance protocols/health records	1
Displacements from feeder	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Lameness Management	**1**
Herd measure - Assess and comment on the management of any score 3 cows seen during the visit, including any in a hospital pen. If no score 3 cows are identified ask about the management of the last score 3 cow on the farm.	1
Measure of hind legs in cubicle passage (e.g. number)	**2**
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Measure of nutritional status (e.g. body condition score)	**11**
content of urea of milk	1
% of animals with BCS < 2 and > 4,75	1
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
BCS	1
Edmonson et al., 1989	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Individual measure: 20 cows selected at random and visually assessed based on the Defra condition scoring method. i.e. Thin - score 1 or less than 2, Moderate - score 2,3 or less than 4, Fat - Score 4 to 5.	1
Observation, body condition score	1
Only at slaughter	1
through records in the farmer's management program	1
Measures of abnormal movement (e.g. number of slips)	**2**
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Measures of abnormal standing-up or sitting-down behaviour (e.g. rising with front legs first)	**4**
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Lying down and standing up movement	1
Measures of behaviour at feeding (e.g. number of displacements)	**3**
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
WQ® protocol	1

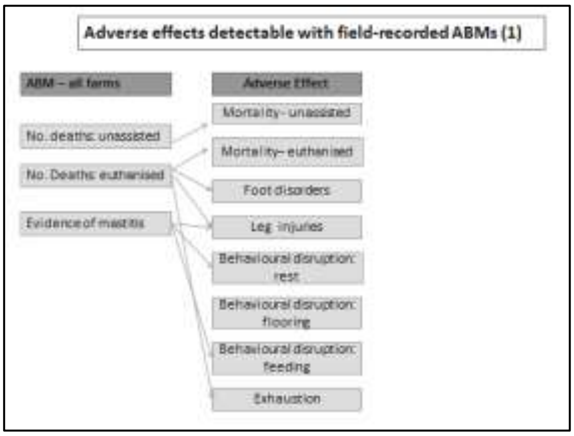
Measures of feed intake (e.g. feeding time, rumen fill)	**4**
Assurance protocols/health records	1
DMI	1
feeding time	1
information on concentrate intake through record in the farmer's management program	1
Measures of lameness (e.g. gait score)	**10**
% gait score 3-4 in the herd	1
Assurance protocols/health records	1
Foot score or locomotion score	1

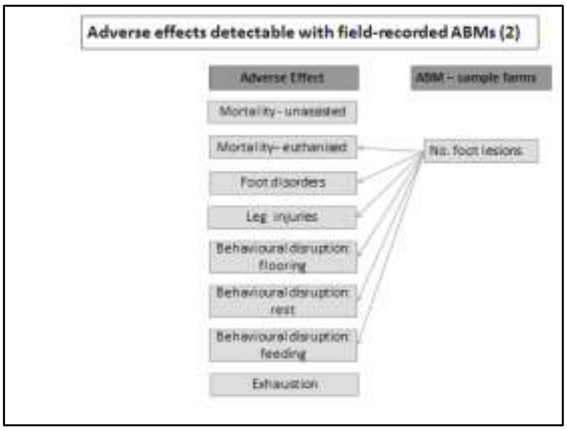
ABM and additional descriptions provided	Frequency
Gait score	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Individual measure: 20 cows selected at random and scores according to the DairyCo scoring method.	1
Locomotion score (scale 1-9). 180000 locomotion scores in the type classification system	1
Observation, locomotion score	1
Stepmetrix system	1
WQ® protocol	1
Measures of lying in passage (e.g. % animals)	**5**
% of animals lying	1
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Herd measure - Whilst assessing the herd, record the number of animals which are not lying correctly, i.e. lying partly (the edge of the cubicle is in contact with the hindquarters or udder) or completely outside the cubicle, or with any other lying difficulty such as dog- sitting or lying backwards. Do not include cows whose heads or bodies are across other cubicles.	1
Measures of overgrown/misshapen hooves	**5**
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Hoof trimmer records	1
presence of long claws	1
Trimming score	1
Measures of posture at rest (e.g. number of cows lying diagonally)	**4**
% of animals correctly resting in cubicles	1
Abnormal position of animals lying in cubicle	1
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Measures of standing in water/slurry (e.g. number of animals)	**3**
% of animals	1
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Number of broken tails	**1**
Herd measure: Whilst assessing the herd, record the number of animals that show evidence of a broken tail, including tails that are bent, short or injured. Investigate and record possible causes of any broken tails observed.	1
Number of collisions with equipment	**1**
Assurance protocols/health records	1
Number of dead animals (euthanised)	**5**
Assurance protocols/health records	1
Death	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
registration loading and unloading animals	1
through records in the farmer's management program	1
Number of dead animals (unassisted death)	**5**
Assurance protocols/health records	1
Death	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Registration loading and unloading animals	1
through records in the farmer's management program	1
Numbers of foot lesions or infectious foot conditions	**9**
% of foot lesions or infectious foot conditions	1
120, 000 animals are scored per year	1
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1

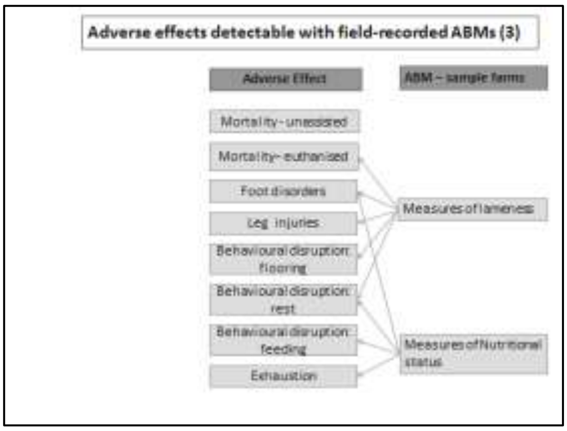
ABM and additional descriptions provided	Frequency
Foot lesions	1
Foot score	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Observation	1
records at the moment of claw care (once or twice yearly)	1
Numbers of hock, knee, skin lesions and swellings	**8**
% of animals with skin lesions	1
Assessment recording scheme of representative number of cows	1
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Individual measure: for hair loss, lesions and swellings 20 cows visually assessed from a distance not exceeding 2m. - a. Head, neck, shoulder, back;b. Flank, side, udderc. Hindquarter;d. Front leg (carpus)e. Hind leg and hock (include whole of tarsus to coronary band). Scoring:0 =No/slight skin damageNo lesions or hairless patches ≥2cm diameter1 = Hairless patchesOne or more hairless patches (may include scars) ≥2cm diameter2 = Lesions and/or swellingOne or more lesions (areas of skin damage i.e. wound or scab) or swellings ≥ 2cm diameter. (Score as a lesion/swelling even if accompanied by a hairless patch. Do not include scars)NB – swollen hocks = a thickening of the joint such that the usual joint anatomy becomes poorly defined or obscured	1
Injuries (neck, shoulders, spinal column, pelvis, ribs, knee, hock lesion and swollen)	1
Skin lesions and swellings	1
WQ® protocol	1
Response of cattle to stockperson	**1**
Herd measure - Check whether the person present for the assessment is the regular stockperson. Throughout the visit, observe the response of the cattle to the stockperson as they approach and interact with the cattle. As far as possible assess response to the stockperson alone, rather than the assessor. Score and comment. Scoring: 0 = Sociable (to the stockperson) 1 = Indifferent, 2 = Cautious	1
Teat score	**1**
Mein et al., 2001	1
Time spent resting	**3**
Assurance protocols/health records	1
Cow comfort index, Stall use index, Stall perching index (	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Time spent standing	**2**
Assurance protocols/health records	1
Health and welfare audit undertaken by independent vet from Royal Vet College	1
Verifying self-assessment	**1**
Records Measure - Check evidence of training in DairyCo mobility scoring - Verify and comment on the farm’s self-assessment of lameness by checking mobility scoring sheets, if they are being completed. Check frequency and scope (e.g. whole herd) of mobility scoring.	1

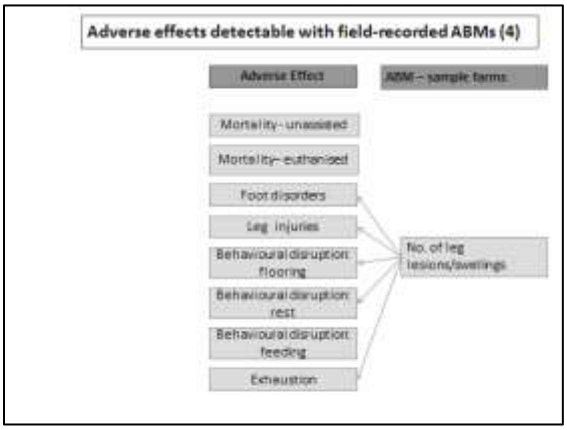
Appendix E. Schematic diagrams showing linkages between ABMs and WAEs (a), and between WAEs (b)

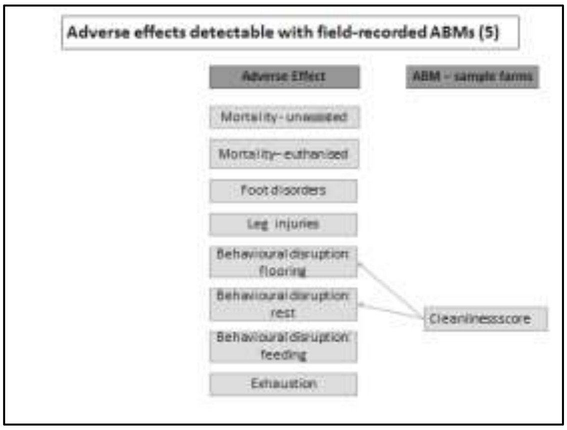
a)



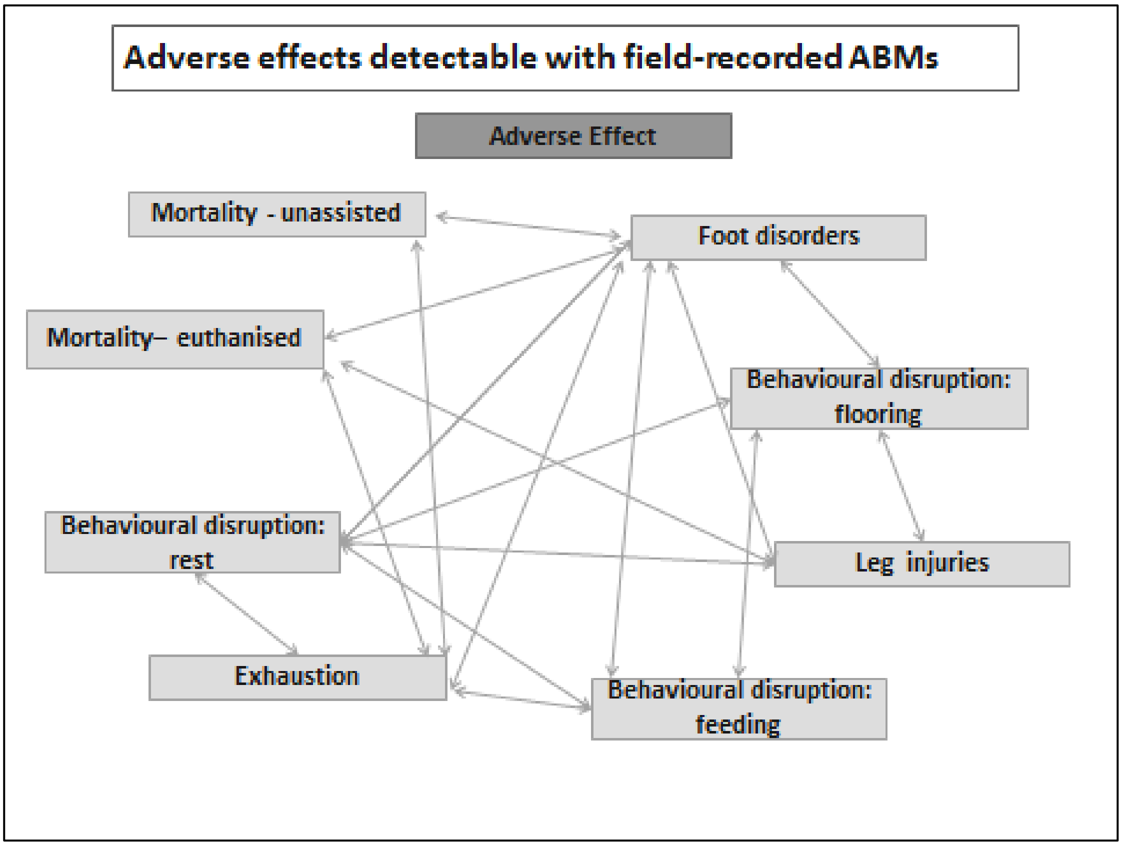








b)


Adverse effects detectable with field-recorded ABMs
Adverse Effect
Mortality - unassisted
Foot disorders
Mortality– euthanised
Behavioural disruption: flooring
Behavioural disruption: rest
Leg injuries
Exhaustion
Behavioural disruption: feeding

Appendix F. Sensitivity to different combinations of search words – mortality

Web of science, Topic, October 2013

Search String	Number hits	Comments/key component
'dairy cow*' or 'cow*' or 'cattle'	193,868	1: Relevant species
'sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement'	1,640,412	2: Validity and robustness
'Mortality' or 'dead' or 'death' or 'deaths' or 'survival' or euthan*'	1,510,290	3: ABM
('dairy cow*' or 'cow*' or 'cattle') and ('sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement')	9,830	1 and 2: Relevant species validity and robustness
('dairy cow*' or 'cow*' or 'cattle') and ('Mortality' or 'dead' or 'death' or 'deaths' or 'survival' or euthan*')	10,690	1 and 3: Relevant species and ABM
('sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement') and ('Mortality' or 'dead' or 'death' or 'deaths' or 'survival' or euthan*')	72,832	2 and 3; Validity/robustness and ABM
('dairy cow*' or 'cow*' or 'cattle') and ('Mortality' or 'dead' or 'death' or 'deaths' or 'survival' or euthan*') and ('sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement')	493	1 and 2 and 3: Relevant species and ABM and validity and robustness

Appendix G. Sensitivity to different combinations of search words – SCC

Web of science, Topic, October 2013

Search String	Number hits	Comments/key component
(‘somatic cell count or SCC’)	**18,007**	**3 ABM**
(‘dairy cow*’ or ‘cow*’ or ‘cattle’) and (‘somatic cell count or SCC’)	2,952	1 and 3: Relevant species and ABM
(‘sensitivity’ or ’specificity’ or ’reliability’ or ‘repeatability’ or ‘reproducibility’ or ‘agreement’) and (‘somatic cell count or SCC’)	1,656	2 and 3; Validity/robustness and ABM
(‘dairy cow*’ or ‘cow*’ or ‘cattle’) and (‘somatic cell count or SCC’) and (‘sensitivity’ or ’specificity’ or ’reliability’ or ‘repeatability’ or ‘reproducibility’ or ‘agreement’)	294	1 and 2 and 3: Relevant species and ABM and validity and robustness

Search from 1970-2013

Appendix H. Sensitivity to different combinations of search words – lameness

Web of Science, Topic, October 2013

Search String	Number hits	Comments/key component
'dairy cow*' or 'cow*' or 'cattle'	211,581	1: Relevant species
'lameness' or 'lame' or 'locomotion' or 'locomotory' or 'gait score'	45,735	2: ABM
'sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement'	1,838,356	3: Validity and robustness
('dairy cow*' or 'cow*' or 'cattle') and ('lameness' or 'lame' or 'locomotion' or 'locomotory' or 'gait score')	1,619	1 and 2: Relevant species and ABM
('dairy cow*' or 'cow*' or 'cattle') and ('sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement')	10,163	1 and 3: Relevant species validity and robustness
('dairy cow*' or 'cow*' or 'cattle') and ('lameness' or 'lame' or 'locomotion' or 'locomotory' or 'gait score') and ('sensitivity' or 'specificity' or 'reliability' or 'repeatability' or 'reproducibility' or 'agreement')	141	1 and 2 and 3: Relevant species and ABM and validity and robustness

Appendix J. Data from ILVO, Belgium provided for Objective 2

Purpose of the database?

Research on modern versus traditional and beginning versus end of stable period

Responsible person(s)/contact person?

Sophie De Graaf

Instituut voor Landbouw- en Visserijonderzoek, Eenheid - Dier
Scheldeweg 68
9090 Melle
+3292722609
Email: Sophie.degraaf@ilvo.vlaanderen.be
www.ilvo.vlaanderen.be

Frank Tyuttens

Email: frank.tuyttens@ilvo.vlaanderen.be

Who has the data ownership?

ILVO, Belgium

Data availability?

Available at the project intranet

Data are available at herd level only – some ABMs are collected on cow-level but no animalID was recorded.

Which variables are recorded according to legislation?

None

Who does the recording?

Researchers at ILVO

Notes

ABMs were recorded according to the Welfare Quality® protocol.

Variables – overview	Variable name in dataset	Variable type	Variable coding	Level
Overall				
Farm number	Frm	Discrete		
Date farm visit	Dat	Date		
Avoidance distance				
<200 cm	CntDst1	Continuous	Avoidance Distance at Feed rack test, distance where the cow pulls back when approached at the feed rack. No. Of animals per distance	Herd level
100 - 200 cm	CntDst2			
50 - 100 cm	CntDst3			
50 – 10	CntDst4			
10 – 0	CntDst5			
able to touch the animal	CntDst6			
Locomotion				
Not lame	CntLoc1	Continuous	Locomotory score. Score 1-5. No. Of animals per score	Herd level
Not lame, could be something wrong	CntLoc2			
Lame	CntLoc3			
severely lame	CntLoc4			
lame at more than 1 leg	CntLoc5			
locomotion not seen	CntLocNS			
Clinical scores				
Body condition score, ranging from 1 (very lean) to 5 (very fat)	BCS	Ordinal		Cow level
cow breed (dubbeldo or milk)	CowTyp	Discrete		Cow level
Hairless patches on the tarsus	TarHP_Oth	Continuous	No. Of lesions per animal	Cow level
Lesions on the tarsus	TarLes	Continuous	No. Of lesions per animal	Cow level
Swellings on the tarsus	TarSwe	Continuous	No. Of lesions per animal	Cow level
Hairless patches on the Hind Quarter	HQ_HP_Oth	Continuous	No. Of lesions per animal	Cow level
Lesions on the Hind Quarter	HQ_Les	Continuous	No. Of lesions per animal	Cow level

Variables – overview	Variable name in dataset	Variable type	Variable coding	Level
Swellings on the Hind Quarter	HQ_Swe	Continuous	No. Of lesions per animal	Cow level
Hairless patches on the Side	SidHP_Oth	Continuous	No. Of lesions per animal	Cow level
Lesions on the Side	SidLes	Continuous	No. Of lesions per animal	Cow level
Swellings on the Side	SidSwe	Continuous	No. Of lesions per animal	Cow level
Hairless patches on the back	BacHP_Oth	Continuous	No. Of lesions per animal	Cow level
Lesions on the back	BacLes	Continuous	No. Of lesions per animal	Cow level
Swellings on the back	BacSwe	Continuous	No. Of lesions per animal	Cow level
Hairless patches on the Carpus	CarHP_Oth	Continuous	No. Of lesions per animal	Cow level
Lesions on the Carpus	CarLes	Continuous	No. Of lesions per animal	Cow level
Swellings on the Carpus	CarSwe	Continuous	No. Of lesions per animal	Cow level
Hairless patches on the other parts of the body	OthHP_Oth	Continuous	No. Of lesions per animal	Cow level
Lesions on the other parts of the body	OthLes	Continuous	No. Of lesions per animal	Cow level
Swellings on the other parts of the body	OthSwe	Continuous	No. Of lesions per animal	Cow level
Overgrown claws	OveCla	Binary (Yes/No)	Overgrown claws on 1 or more legs	Cow level
Nasal discharge	NasDis	Binary (Yes/No)	Yes = Presence of parameter	Cow level
Ocular discharge	OcuDis	Binary (Yes/No)	Yes = Presence of parameter	Cow level
Increased Respiratory Rate	IncResRat	Binary (Yes/No)	Yes = Presence of parameter	Cow level
Diarrhea	Dia	Binary (Yes/No)	Yes = Presence of parameter	Cow level
Vulvar Discharge	VulDis	Binary (Yes/No)	Yes = Presence of parameter	Cow level
Broken Tail	BroTai	Binary (Yes/No)	Yes = Presence of parameter	Cow level
Feaces score, ranging from 1(very thin) to 5 (very thick)	Fae	Ordinal	1 = Very thin - 5 = Very thick	Cow level

Variables – overview	Variable name in dataset	Variable type	Variable coding	Level
Clinical scores (cont.)				
Dirty Abdomen	Abd			Cow level
Dirty leg	Leg			Cow level
Dirty Flank	Fla			Cow level
Dirty udder	Udd			Cow level
Dirty Teats	Tea			Cow level
Behavioural observations				
Scan time	ScaTim	Continuous	No. Of observations	
Number of animals standing at start	StaSta	Continuous	No. Of observations	
Number of animals lying at start	StaLyi	Continuous	No. Of observations	
Number of animals feeding at start	StaFee	Continuous	No. Of observations	
Number of animals lying out of the lying area at start	StaLyo	Continuous	No. Of observations	
Number of animals of which it is not seen whether they lie inside or outside the lying area at start	StaLyoNs	Continuous	No. Of observations	
Number of animals standing at ending observation	EndSta	Continuous	No. Of observations	
Number of animals lying at ending observation	EndLyi	Continuous	No. Of observations	
Number of animals feeding at ending observation	EndFee	Continuous	No. Of observations	
Number of animals lying out of the lying area at ending observation	EndLyo	Continuous	No. Of observations	
Number of animals of which it is not seen whether they lie inside or outside the lying area at ending observation	EndLyoNs	Continuous	No. Of observations	
Head butts	HB	Continuous	No. Of observations	
Chasing	CH	Continuous	No. Of observations	
Fighting	FI	Continuous	No. Of observations	
Sneezing	SN	Continuous	No. Of observations	
Tongue rolling	TR	Continuous	No. Of observations	
Displacement	DP	Continuous	No. Of observations	
Chasing upp	CHU	Continuous	No. Of observations	
Licking, social	LI	Continuous	No. Of observations	

Variables – overview	Variable name in dataset	Variable type	Variable coding	Level
Cough	CO	Continuous	No. Of observations	
Rubbing	BR	Continuous	No. Of observations	
Behavioural observations – lying				
Section of stable	IDSeg	Discrete		
Time it takes to lie down	TimLyi	Continuous	Seconds	
Collisions with housing equipment	Col	Binary	1 = Collision, 0 = No collision	
Behavioural observations – Herd scan				
Herd scan - number of animals standing in cubicles	StaCub	Continuous	No. Of observations	
Herd scan - number of animals standing	Sta	Continuous	No. Of observations	
Herd scan - number of animals lying	Lyi	Continuous	No. Of observations	
Herd scan - number of animals feeding	Fee	Continuous	No. Of observations	
Management				
Farm	Frm	Discrete	Farm identification	
Number of animal	averageanimalsyear	Continuous	Average number of cows (milking) at the farm	
Pasture access, days	pasturedays	Continuous	Number of days the cows are on pasture per year	
Pasture access, hours	pasturehours	Continuous	Number of hours the animals are on pasture	
Outdoor area access, days	outsideaccessday	Continuous	Number of days the cows have access to an outdoor area (not pasture)	
Outdoor area access, hours	outsideaccesshours	Continuous	Number of hours the animals have access to an outdoor area	
Dyctocia	dystocia	Continuous	Number of animals with a difficult delivery in the last year	
Downer cows	downer	Continuous	Number of downer cows in the last year	
Mortality	deaths	Continuous	Number of deaths in the last year	
Dehorned, visit	percdehorned	Continuous	Percentage of animals that are dehorned at the moment of visit	
Dehorned, total	dehornedatfarm	Continuous	Percentag of animals that are dehorned at the farm	
Dehorn, age	agedehorned	Continuous	Average age the animals are dehorned	

Variables – overview	Variable name in dataset	Variable type	Variable coding	Level
Dehorn, method	methoddehorned	Binary	Method of dehorning: Thermo or paste	
Dehorn, analgetic	dehornedanalgesics	Binary	Dehorning with analgesics	
Dehorn, anaestetic	dehornedanaestetics	Binary	Dehorning with anaestetics	
SCC	SCC400000	Continuous	Number of animals in the past three months that had a Somatic cell count of (more than) 400.000	
Housing	housing	Binary	Housing system: tie stall or louse housed	
Tiestall - all year	tiestallboundyearround	Binary	Only tie stall: whether the animals are tethered year round	
Tiestall – winter	tiestallboundwinter	Binary	Only tie stall: whether the animals are bound the whole winter season	
Tiestall – released	tiestallmovement	Binary	Only tie stall: whether the animals are released regularly	
Taildock, percent	taildock	Continuous	Percentage of animals that are tail docked	
Taildock, method	taildockmethod	Binary	Method of tail docking: rubber band or surgery	
Taildock, analgetic	taildockanalgesics	Binary	Tail docking with analgesics	
Taildock, anastetic	taildockanaestetics	Binary	Tail docking with anaestetics	
Drinking, number bowls	drinkwaterbowl	Continuous	Number of drinking water bowls	
Drinking, length	drinklengththr	Continuous	Length of throughs added up	
Drinking, number trough	drinknumberthro	Continuous	Number of throughs	
Drinking, flow	drindebiet	Binary	Whether the flow in the bowl and throughs is sufficient	
Drinking, clean	drinkclean	Binary	Whether the drinkers are clean	
Drinking, amount suff.	drinknumbersufficient	Binary	Whether the amountof drinkers is sufficient	

Appendix K. Data from IZSLER, Italy provided for Objective 2.

Purpose of the database?

This database comes from the IZSLER/CReNBA welfare assessment protocol for dairy cows kept in loose housing systems. The protocol was developed by Dr. Luigi Bertocchi and his co-workers. Data were collected to assess dairy cow welfare within some national research projects and also because some private productive chains asked to assess cow welfare in their own farms.
This database comprises the welfare results of 442 Northern Italy farms, assessed during the period 2011-2013.
CReNBA is the Italian acronym that stands for National Reference Centre for Animal Welfare, which is located in Istituto Zooprofilattico Sperimentale della Lombardia ed Emilia Romagna “Bruno Ubertini” (IZSLER, Brescia).

Responsible person(s)/contact person?

Dr. Luigi Bertocchi

Reparto Produzione Primaria
Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini" (IZSLER)
Centro di Referenza Nazionale per il Benessere Animale (CReNBA)
Via Bianchi, 9 25124 Brescia
Tel 030 2290250
Email:luigi.bertocchi@izsler.it

Dr. Francesca Fusi

Email: francesca.fusi@izsler.it

Who has the data ownership?

Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini" (IZSLER) - Centro di Referenza Nazionale per il Benessere Animale (CReNBA), Italy

Some further details can be found in Italian at this web site:

http://www.izsler.it/izs_home_page/archivio_news/00003111_Manuale_per_la_valutazione_del_benessere_e_della_biosicurezza_nell_allevamento_bovino_da_latte_a_stabulazione_libera.html

Data availability?

Available at the project intranet

Data on herd level only.

Which variables are recorded according to legislation?

- "Udder health - BTSCC (Bulk Tank Somatic Cell Count)" : geometric average BTSCC over a three-month period according to Regulation (EC) No 853/2004 (See Table below “Overview of variables – type and coding”).

• "Mutilations" according to the Italian national legislation,D. L.vo No 146/2001(Directive 98/58/EC) (See Table below “Overview of variables – type and coding”).

.

Who does the recording?

The ABMs data collection and the overall welfare assessment are carried out, during a farm visit, by a researcher or a veterinary practitioner after having attended and overcome a specific training course. The training course is carried out in IZSLER/CReNBA in Brescia (Italy) by Dr. Luigi Bertocchi and his co-workers.

Overview of variables – type and coding

Variables - overview	Variable type	Variable coding	Who does recordings	Level	Notes
Date	Date				
Farm number	Discrete				
Herd size	Continuous				
No. of lactating cows	Continuous				
Kg milk/cow/day	Continuous		Based on farmers' records	Farm level	Average daily milk production per cow on day of visit. Not energy corrected.
Cleanliness (lactating cows, dry cows, heifers)	Ordinal	1 = >20% of animals are dirty. 2 = ≥10% of animals are dirty. 3 = <10% of animals are dirty	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA	Farm level	Scoring performed on a binary scale (clean/not clean)
Avoidance test (lactating cows, dry cows, heifers)	Ordinal	1 = Difficulty of approach and the presence of stereotypies. 2 = Curious animals that come close and do not show stereotypies. 3 = Animals that come close and you can easily touch them	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA	Farm level	According to Welfare Quality® protocol (6.1.4.3.: Good human animal relationship). Stereotopies are such as tongue rolling/playing, bar biting, vacuum chewing, drinker playing etc. but primarily the outcome of the avoidance test is considered.
BCS (Body Condition Score) (lactating cows, dry cows, heifers)	Ordinal	1 = ≥10% of the animals outside BCS limits. 2= 5-10% of the animals outside BCS limits. 3= <5% of the animals outside BCS limits.	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA	Farm level	BCS limits: BCS values <2 and >4.25 are considered outside BCS limits.
Udder health	Ordinal	1= Geometric mean SCC >400.000 cells/ml. 2= 300.000 < Geometric mean SCC < 400.000 cells/ml. 3= Geometric mean SCC <300.000 cells/ml	Based on farmers' records	Farm level	SCC geometric mean over a three-month period.
No. of treatments for clinical mastitis	Ordinal	1 = >80% of lactating cows OR data not availiable. 2 = 40-80% of lactating cows. 3 = <40% of lactating cows	Based on farmers' records	Farm level	Farmers' records
Lameness	Ordinal	1 = >8% lame animals. 2 = 4-8% lame animals. 3 = <4% lame animals	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA	Farm level	Dichotomised score: 0, 1 = Not lame / 2, 3 = Lame. Original score from DairyCo Mobility Scoring System http://www.healthyhooves.eu/pdffiles/dairycomobilityscore.pdf

Variables - overview	Variable type	Variable coding	Who does recordings	Level	Notes
Skin lesions (lactating cows, dry cows, heifers)	Ordinal	1 = >30% of animals with skin lesionson hocks, bone tuberosity and soft tissue. 2 = 15-30% of animals with skin lesions on hocks, bone tuberosity and soft tissue. 3 = <15% of animals with skin lesions on hocks, bone tuberosity and soft tissue	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA	Farm level	Score: Welfare Quality® Protocol (6.1.3.1 Absence of injuries - Integument alterations)
Adult cows' mortality	Ordinal	1 = >5%. 2 = 2 - 5%. 3 = <2%	Based on farmers' records	Farm level	Farmers' records. Unassisted AND euthanised during last 12 months. Proportion of "adult cows found dead + adult cows euthanised"/ total adult cows (both lactating cows + dry cows)
Calves' mortality	Ordinal	1 = >10%. 2 = 4 - 10%. 3 = <4%	Based on farmers' records	Farm level	Farmers' records. Unassisted AND euthanised during last 12 months. Proportion of dead calves (unassisted and euthanised) / total number of calves during the latest 12 months.
Mutilations	Ordinal	1 = Mutilation not comply with law / cowswith and without horns mixed. 2 = Mutilations comply to current regulations. 3 = No mutilations		Farm level	Tail docking and disbudding (calves) / de-horning (cows) not in compliance with legislation (national legislative decree No 146/2001 (Directive 98/58/EC)
Housing system	Discrete (text)	Loose-housing / Tie-stall	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Exercise area / access to pasture	Binary (yes/no)		Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Calving pen	Ordinal	1 = No calving pen. 2 = Calving pen with deep litter and insufficient space (<6 m^2) OR with cubicles OR wrongly manage for times or litters hygiene. 3 = Calving pen with well managed deep litter	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Sickbay pen	Ordinal	1 = No sickbay pen. 2 = Sickbay pen not well managed OR with cubicles. 3 = Sickbay pen with well managed deep litter and easy access to milking	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		

Bedding material (lactating cows, dry cows, heifers)	Ordinal	1 = Absence of bedding material. 2 = Presence of mattress/floor mattress OR inappropriate bedding material. 3 = Presence of adequate bedding material.	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA

Variables - overview	**Variable type**	**Variable coding**	**Who does recordings**	**Level**	**Notes**
Floor type (lactating cows, dry cows, heifers)		1 = Slatted floor unsuitable or concrete floor smooth and slippery, 2 = Slatted floor suitable and wrinkled floor on at least 50% of total surface, 3 = Slatted floor suitable and wrinkled floor on total surface.	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Cleanliness of bedding (lactating cows, dry cows, heifers)	Ordinal	1 = Dirty bedding. 2 = Bedding not very clean and insufficiently managed. 3 = Clean bedding, top up or daily care and periodic replacement	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Adjustment of milking machines	Ordinal	1 = Absence of planned adjustment AND presence of deteriorated rubber parts or non-functioning devices. 2 = Random adjustment and rubber parts in good condition, but absence of documentation. 3 = Planned adjustment and presence of written records	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Hoof care	Ordinal	1 = No hoof care. 2 = Hoof care once/year. 3 = Hoof care twice/year			= Hoof trimming
Milking	Ordinal	1 = No hygiene, no correct oxytocine time. 2 = Good hygiene and oxytocine time respected. 3 = Good hygiene, pre-/post- teat dipping with clean cups and correct oxytocine time	Observer (researcher or veterinary practitioner) trained by IZSLER/CReNBA		
Cleanliness of bedding (lactating cows, dry cows, heifers)	Ordinal	1 = Dirty bedding. 2 = Bedding not very clean and insufficiently managed. 3 = Clean bedding, top up or daily care and periodic replacement	Observer (researcher or veterinarian) trained by IZLER		
Adjustment of milking machines	Ordinal	1 = Absence of planned adjustment AND presence of deteriorated rubber parts or non-functioning devices. 2 = Random adjustment and rubber parts in good condition, but absence of docomentation. 3 = Planned adjustment and presence of written records	Observer (researcher or veterinarian) trained by IZLER		

Hoof care	Ordinal	1 = No hoof care. 2 = Hoof care 1 time/year. 3 = Hoof care 2 times/year		= Hoof trimming
Milking	Ordinal	1 = No hygiene, no correct oxytocin time. 2 = Good hygiene and oxytocine time respected. 3 = Good hygiene, pre-/post- teat dipping with clean cups and correct oxytocine time	Observer (researcher or veterinarian) trained by IZLER	

Appendix L. Data from INRA, France.

Purpose of the database?

Research on modern versus traditional and beginning versus end of stable period

Responsible person(s)/contact person?

Luc Mounier

tel : +33 (0) 4 78 87 27 71 fax : +33 (0) 4 78 87 27 81

Email: luc.mounier@vetagro-sup.fr

Alice de Boyer des Roches

Tel : + 33 (0) 4 78 87 27 87 fax : + 33 (0) 4 78 87 27 81

Email: alice.deboyerdesroches@vetagro-sup.fr

VetAgro Sup Campus Vétérinaire de Lyon, Unité Gestion des élevages

1 avenue Bourgelat, F- 69280 Marcy l'étoile, France

http://vetagro-sup.fr/

INRA, UMR1213 Herbivores, Equipe Adaptation et Comportements sociaux

F-63122 Saint-Genès-Champanelle, France

http://www.inra.fr/

Who has the data ownership?

INRA, France

Data availability?

Available at the project intranet

Data are available at herd and/or individual levels.

Which variables are recorded according to legislation?

Mortality.

Who does the recording?

Researchers at INRA

Notes

ABMs were recorded according to the Welfare Quality® protocol.

Variables – overview	Variable name in dataset	Variable type	Variable coding	Level
Herd number	Unit_name OR Nom_ferme			
Date of WQ® assessment	DATE_OBS OR DATE_VISITE			
Date of SCC measure	DATE_CL_1	Date		
	DATE_CL_2			
SCC	NOMBR_CEL_CL_1	Continuous	SCC *1000	Cow level
	NOMBR_CEL_CL_2	Continuous		
	NOMBR_CEL_CL_3			
Proportion of non-lame cows	Perc_not_lame_cows	Continuous	Locomotion scoring performed using the WQ® scoring system: Non-lame = 0, Moderately lame = 2, Severely lame = 3	Herd level
Proportion of moderately lame cows (or lame cows in tie stalls)	Perc_moderately_lame_cows_(if_loose_housed)_or_%_lame_cows_(if_tied)	Continuous		Herd level
Proportion of severely lame cows	Perc_severely_lame_cows	Continuous		Herd level
Proportion of dead cows during last 12 months	Perc_mortality_during_the_last_12_months	Continuous		Herd level

Appendix M. Overview of variables in datasets analysed in Objective 5

Overview of variables – IZSLER/CRenBA

Variables - overview	Variable type	Variable name in dataset	Variable coding in the analyses	Who does recordings	Level	Notes
Herd number	Discrete	Herd				
Herd size	Continuous	HerdSize				Number of animals
Milk production (Kg milk/cow/day)	Continuous	MilkYield		Based on farmers' records	Farm level	Average daily milk production per cow on day of visit; not energy corrected.
Housing system	Categorical	House	Tethering , loose house, loose house with access to outdoor exercise area	Observer (researcher or veterinarian practitioner) trained by IZSLER/CReNBA	Farm level	The observations should be carried out on all animals. Tethering for even a single group of animals. Loose housing system for all groups of animals. Loose housing system and availability of an exercise area or pasture for adult cattle. The exercise area must have a total surface of the resting area which could be both with cubicles and straw litter, provided that it is covered.
Udder health	Categorical	SCC	Geometric mean BTSCC <300.000 cells/ml., Geometric mean BTSCC 300.000-400.000 cells/ml., Geometric mean BT SCC >400.000cells/ml	Based on farmers' records	Farm level	BTSCC geometric mean. Based on the available data on the date of the visit.
Floor type	Categorical	Floor	Unsuitable, >50% good surface, only good floor.	Observer (researcher or veterinarian practitioner) trained by IZSLER/CReNBA	Farm level	Criteria: The animals must be able to move, even quickly, without slipping. Presence of unsuitable slatted floor or concrete floor smooth and slippery. Presence of suitable slatted floor and good roughness on at least 50 % of the surface. Presence of suitable slatted floor and good roughness on all floors/surface
Adult cows' mortality	Categorical	Mort	5%, 2 - ≤5%, <2%	Based on farmers' records	Farm level	Annual mortality rate for adult cows, including unassisted, euthanized and emergency slaughtered
Lameness	Categorical	Lame	>8%, 4-8%, <4	Observer (researcher or veterinarian practitioner) trained by IZSLER/CReNBA	Farm level	Percentages of lame animals: Number of cows (both lactating and dry cows - including cows in sick pen)/ number of total adult cows (both lactating and dry cows)

Overview of variables – Welfare Quality®

Variables overview	Variable type in dataset	Variable name in dataset	Variable coding in the dataset	Variable coding in the analyses	Data source	Level	Notes
Herd number	Discrete	Herd					Id number of the herd
Herd size	Continuous	HerdSize	Number of cows				Number of cows
Udder health	Continuous	SCC	Percentage of cows with a Somatic cell count of (more than) 400.000	French dataset: <10 % of cows with SCC > 400.000, >=10 % of cows with SCC > 400.000 Belgian dataset: <20 % of cows with SCC > 400.000, >=20 % of cows with SCC > 400.000		Farm level	Percentage of animals that had a Somatic cell count of (more than) 400.000 in the past three months out of the total herd size
Adult cows' mortality	Continuous	Mort	Percentage of dead animals	French dataset: <5%, >=5% Belgian dataset mortality <3%, >=3%	Farmers' records	Farm level	Percentage of animals dead animals in the last year out of the total herd size
Lameness	Continuous	Lame	Percentage of lame and severely lame cows	French dataset: <8%, >=8% Belgian dataset: <22%, >=22%	Observer	Farm level	Percentages of lame cows was calculated as the sum of the scores: " lame", "severely lame" and "lame on more the one leg" out of the total numbers of cows observed
Access to pasture	Binary/continuons	Pasture	French dataset: Grazing/no grazing Belgian dataset: Number of days on pasture yearly	Yes,no		Farm level	Access to pasture. In the Belgian dataset < 10 days = no grazing
Breed	Binary	Breed	Breed type	French dataset: Holstein (milking), Montbéliarde (Double purpose) Belgian dataset: Milkin, Double purpose		Farm level	
House	Binary	House	House type	French dataset: Cubicles, Straw Belgian dataset: Tie stall, Loose housed		Farm level	

Overview of variables – Otten and Burow

Variables - overview	Variable type in dataset	Variable name in dataset	Variable coding	Variable coding in the analysis	Data source	Level	Notes
Herd number	Discrete	Herd					Id number of the herd
Herd size	Continuous	HerdSize			Danish Cattle Database		Number of cows
Udder health	Continuous	SCC	Mean of bulk tank somatic cell count over a year	<300.000, >=300.000	Danish Cattle Database	Farm level	Mean of bulk tank somatic cell count over a year
Adult cows' mortality	Continuous	Mort	Percentage of dead animals	<5% , >=5%	Danish Cattle Database	Farm level	Percentage of dead animals; euthanized or dying unassisted during the last 12 months
Lameness	Continuous	Lame	Lame_mod: Percentage of moderately lame cows Lame_sev : Percentage of severely lame cows	<28% , >=28%	Observer	Farm level	Moderately lame cows are defined by impaired stride and/or rhythm with reduced weight bearing on one limb. Severely lame cows are defined by no weight bearing or more than one limb affected by lameness
Production type	binary	Organic	Conventional, organic		Danish Cattle Database	Farm level	
Breed	Categorical	Breed	Danish Holstein (DH) / Red Danish Dairy (RDD) / Jersey / Crossbreed (CB)		Danish Cattle Database	Farm level	

Appendix N. Table sent to the members of the consortium for the final selection of the factors of variation.

The goal of this exercise is to reduce the number of parameters from this exhaustive list and keep fewer parameters (factors defining a population and/or causing the highest variability/impact on the ABMs and WAEs and/or easy to collect)			
(1) Please, on the column 'a', put the degree of relevance of the parameter with the three ABMs together (the strength of the association between this parameter and the three WAEs). Score from 1 (low relevance) to 5 (very high relevance)			
(2) Please, on the column 'b', put a cross if for you this parameter is needed to characterise a population (do you think that this parameter has to be kept?)			
(3) Please, on the column 'c', put a cross if you think that this parameter is easy to record on farm (by a "non specialist") AND easy to collect routinely and to keep updated in a database (not only parameters already collected but also those that could be collected in the near future)			
A risk factor is here considered as any factor having an association with the outcome of the ABM / apparition of a WAE.	a	b	c
Entity: Animal class			
Parameter type: Genetic			
* Breed			
* Milk production / yield (individual, herd)			
* Udder conformation (level of the udder from the hock…)			
* Weight			
Parameter type: Physiology			
* Age			
* Body condition score (skinny/normal/fat…)			
* Days in milk / after calving			
* Lactation and reproductive status (open/dry/pregnant)			
* Parity (primiparous/multiparous, number of parities)			
Parameter type: Health			
* Reproductive problems (high levels of abortion, retained placentas, dystocia, metritis...)			
* Injuries, accidents, pain (wounds, swellings…)			
* Udder status (teat end quality, cleanliness, hyperkeratosis, contamination by pathogens)			
* Hoof status (overgrown, trimming)			
Entity: Housing class			
Parameter type: Housing system			
* Housing system (tie-stalls, free-stalls with deep litter, free-stalls with cubicle)			
* Milking system (Automatic MS, milking parlour, traditional milking)			
Parameter type: Space organisation			

* Density (overstocking: cubicles per animal; feeding space per animal, drinkers per animal…)				
* Outside / exercise area (availability, size…)				
* Access to pasture (duration, period, distance, presence of shed…)				
* Calving pen (multiple animals / single animal…)				
* Beef cow unit				
* Contact with other species (poultry, cat)				
Parameter type: Material				
* Ergonomics (cubicle: brisket board, neck rail; tie*stall: electric cow trainers; steps in front of the manger…)				
* Bedding material (cubicle, calving pen…)				
Parameter type: Floor				
* Floor type (slatted, concrete, mattress)				
* Floor quality (broken ground, slippery ground, stairs, slope…)				
Parameter type: Ambience				
* Indoor environment (ventilation, Temperature Humidity Index)				
Entity: Management class				
Parameter type: Stockmanship				
* Behaviour towards the animals (at milking, herding…)				
* Empathy, threshold to recognise disease and pain, call veterinarian				
* Ability to detect/report diseases and welfare issues (detection of risk injury, of lameness…)				
* Organic dairy production				
* Available workforce at milking				
* Relation with the outside (member of milk control program, contract with foot trimmer, classes attendance)				
* Rules for mixing animals (introduction of transition cows in the herd, heifers…)				
* Calving interval				
Parameter type: Hygiene/Maintenance				
* Cleanliness of alleys and lying area				
* Cleanliness of water trough and manger				
* Manure used for bedding				
Parameter type: Herd health				
* Animal care/hygiene (cleanliness, hoof care, footbath, hair clipping)				
* Biosecurity (rules for outside persons, quarantine duration…)				
* Veterinary treatments (vaccines, antibiotics, foot care…)				
Parameter type: Milking				
* Order of milking (high producing cows, high SCC cows)				

* Procedure (frequency, waiting time, individual care)			
* Milking machine cleanliness (controls, dipping or backflushing units…)			
* Hygiene procedures during milking (teat dipping, gloves, teat preparation…)			
Parameter type: Dry-cows			
* Management of dry-cows (dry-cow therapy, specific surveillance, pre-calving hoof care)			
* Housing of dry-cows (comfort, cleanliness…)			
Parameter type: Feeding			
* Water quality and availability			
* Concentrate consumption (proportion, amount)			
* Forage or fibre availability / % in the ration			
* Supplementation with minerals and vitamins (selenium, vitamin E/A…)			
* Cows blocked at feeding (duration, prolonged standing…)			
* Feeding frequency			
* TMR feeding (Total Mixed Ration)			
* Negative Energy Balance (confirmed by tests e.g. MUN Milk Urea Nitrogen test)			
* Nutrition of calves (quality of milk: mastitic milk, antibiotics…)			
* Specific nutrition according to lactation or physiological status (feeding transitions)			
Parameter type: Culling			
* Culling rules (lesions, udder conformation, DIM…)			
* Replacement rate and origin (heifers bought outside…)			
* Economics (price of milk, replacement heifers, culled cows…)			
Entity: General factors class			
Parameter type: General factors			
* Herd size			
* Outdoor conditions (THI…)			
* Geographical region			
* Milk quota			

Appendix O. Table used to evaluate the availability of the final parameters in each country.

Parameter	1. In the database (y/n)	2. On herd level (y/n)	3. On animal level (y/n)	4. Number of farms a year	5. Number of animals a year	6. Number of sampling moments a year	7. Scale used in database	8. Method used to collect data
Parity								
Housing system								
Floor type								
Days in milk								
Access to pasture								
Milk production								
Herd size								
Breed								
Geographical region								
Organic dairy production								
Lameness								
Somatic cell count								
Mortality								

Appendix P. Table listing all parameters found in the Objective 3 literature search for risk factors/factors of variation associated with the 3 ABMs selected in objective 2 and associated WAEs (mortality, mastitis/SCC and lameness) with supporting references.

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
Animal	Genetic	1	*breed*	Alvasen et al., 2012; Dechow et al., 2011; Hare et al., 2006; Miller et al., 2008; Raboisson et al., 2011; Thomsen et al., 2006	Elbers et al., 1998; Ivemeyer et al., 2009; Ivemeyer et al., 2011; Myllys & Rautala, 1995; (Nickerson et al., 1995); Prendiville et al., 2010; Sharma et al., 2011; Walsh et al., 2007	Alban, 1994; Alban et al., 1996
		2	*personality (fearfulness/adaptation) / workability*			
		3	*cow conformation (BCS)*		Breen et al., 2009; Berry et al., 2007; Compton et al., 2007	Wells et al., 1993
		4	*high milk yield (milk production)*	Alvasen et al., 2012; Batra et al., 1971;Burow et al., 2011; Dechow and Goodling, 2008; Dematawewa & Berger, 1998; Harris, 1989; McConnel et al., 2008; Miller et al., 2008; Pinedo et al., 2010; Smith et al., 2000; Thomsen et al., 2006	Agabriel et al., 1997; DeGraves and Fetrow, 1993; DeVries et al., 2012; Eberhart et al., 1982; Emanuelson & Funke, 1991; Fenlon et al., 1995; Goodhope & Meek, 1980; Green et al., 2006; Igono et al., 1988; Khaitsa et al., 2000; Koldeweij et al., 1999; Moxley et al., 1978; Myllys & Rautala, 1995; Raubertas & Shook, 1982; Rogers et al., 1998; Shook, 1993; Tyler et al., 1989; Valde et al., 1997; Waage et al., 1998;Wenz et al., 2007; Wilson et al., 1997	Alban et al., 1996; Archer et al. 2010; Green et al., 2002; Warnick et al., 2001;
		5	*fat-protein inversion*	Dechow & Goodling, 2008; Dechow et al., 2011		
		6	*udder volume*			
		7	*udder conformation*		Barkema et al., 1998b; Bareille et	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
					al., 1998; Breen et al., 2009; Dohmen et al., 2010	
		8	*udder hygiene score*		Reneau et al., 2005;	
		9	*size*			
		10	*weight*	Erb et al., 1985		Wells et al., 1993
		11	*ease of calving*	Thomsen et al., 2004		
		12	*Estimated Transmitting Ability for milk calculated from pedigree*	Erb et al., 1985	De Haas et al., 2002	
	physiology	13	*age*	Dematewewa et al., 1998; Erb et al., 1985; Faye & Perochon, 1995; Hadley et al., 2006; Harris, 1989; Miller et al., 2008; McConnel et al., 2009; Thomsen et al., 2004	Beckley & Johnson, 1966; Blackburn, 1966; Detilleux et al., 2013; Eberhart et al., 1979; Erb et al., 1985; Reichmuth, 1975; Sheldrake et al., 1983	
		14	*lactation stage /days after calving / days in milk*	Dechow & Goodling, 2008; Faye and Perochon, 1995; Hadley et al., 2006; Hertl et al., 2011; Menzies et al., 1995; Milian-Suazo et al., 1988; Miller et al., 2008; Pinedo et al., 2010; Raboisson et al., 2011; Stevenson & Lean, 1998; Thomsen et al., 2004	DeVliegher et al., 2004; DeVries et al., 2012; Dohoo & Meek, 1982; Laevens et al., 1997; McDonald & Anderson, 1981; Myllis & Rautala, 1995; O'Driscoll et al., 2008; Olde Riekerink et al., 2007; Osteras & Lund, 1988 a; Sargeant et al., 1998; Smith et al., 1985; Steeneveld et al., 2008;	Boettcher et al., 1998; Webster 2002
		15	*pregnancy state*	Pinedo et al., 2010		

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		16	*parity*	Dematewewa et al., 1998; McConnel et al., 2009; Miller et al., 2008; Pinedo et al., 2010; Raboisson et al., 2011; Raboisson et al., 2011; Thomsen et al., 2004;	Barkema et al., 1998 a; Blackburn, 1966; Compton et al., 2007; Devries et al., 2012; Fadlelmoula et al., 2007; Green et al., 2007; Lindstrom et al., 1981; Ivemeyer et al., 2011; McCarthy et al., 2007; Mitchell et al., 1976; Morse et al., 1988;Olde Riekerink et al., 2007; Pearson and Mackie, 1979; Sargeant et al., 1998; Steeneveld et al., 2008; Smith et al., 1985;Schutz et al., 1990; Sheldrake et al., 1983; Valde et al., 2004; van den Borne et al., 2010; Walsh et al., 2007; Watters et al., 2013; Wilesmith et al., 1986; Whist et al., 2006;	Alban et al., 1996; Boettcher et al., 1998; Vokey et al., 2001; Webster, 2001;
		17	*dry period duration*		Peeler et al., 2000; Pytlewski et al., 2012; Pinedo et al., 2011; Skrzypek et al., 2004	
		18	*endocrinal changes*			
		19	*Diurnal variation*		Sharma et al., 2011; White & Rattray, 1965	
	animal health	20	*stress load*			
		21	*reproductive problems (high levels of abortion, retained placentas, dystocia, metritis)*	McConnel et al., 2008; Dematawewa & Berger, 1997		
		22	*immune state*			

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		23	*production diseases, physical injuries…*	Agger, 1983; Bar et al., 2008; Burow et al., 2011;Enemark, 2008; Dechow & Goodling, 2008; Dechow et al., 2011; Esslemont & Kossaibati, 1997; Faye & Perochon, 1995; Fulwider et al., 2007; Grohn et al., 1998 (culling); Hertl et al., 2011; McConnel et al., 2008; McConnel et al., 2009, 2010; Menzies et al., 1995; Nocek, 1997; Seegers et al., 2003 (REV); Thomsen et al., 2006; Watson et al., 2008; Thomsen et al., 2004;	Detilleux et al., 2013; Elbers et al., 1998; Fulwider et al., 2007; Hultgren et al., 2004; Sogstad et al., 2006; Svensson et al., 2006; Waage et al., 1998	Fulwider et al., 2007
		24	*Mamary gland infection level (mastitis)*		Craven & Williams, 1985; Dohoo & Meek, 1982; Meek et al., 1980; Miller et al., 1990; Wilson et al., 1997	
		25	*Baseline SCC*		Watters et al., 2013	
		26	*pain*	Andersen & Muir, 2005		
		27	*twinning*			
		28	*Teat End Callosity (score) (TEC)*		Breen et al., 2009	
		29	*teat-end hyperkeratosis (HK)*		Breen et al., 2009; Lewis et al., 2000; Sieber & Farnsworth, 1981	
		30	*cleanliness*		Agabriel et al., 1997; Barkema et al., 1999- 1998b; Compton et al., 2007; DeVries et al., 2012; Dohmen et al., 2010; Reneau et al., 2005; Schreiner and Ruegg, 2003; Valde et al., 1997; Ward et al., 2002;	
		31	*claw and leg health*			Laursen et. Al 2009
Housing	**housing system**	32	*tie-stalls*	Alvasen et al., 2012 ; DeChow et al., 2011	Dechow et al., 2011; Khaitsa et al., 2000; Osteras & Lund, 1988 a; Valde et al., 1997; Wronski et	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
					al., 2007	
		33	*loose-housing*	Dechow et al., 2011	Barnouin et al., 2004; Bartlett et al., 1992 b	Faye & Lescourret, 1989
		34	*no outside area*	McConnel et al., 2008		Rouha-Mülleder et.al., 2009
		35	*freestalls as primary housing facility*	Alvasen et al., 2012; McConnel et al., 2008	(Dufour et al., 2011); Bartlett et al., 1992; Khaitsa et al., 2000; Smith et Ely, 1997; Wenz et al., 2007;	
		36	*freestalls with deep litter*	Thomsen et al., 2006		Livesey et al., 1998; Webster, 2002
		37	*freestalls with cubicle*	Thomsen et al., 2006	Detilleux et al., 2013	Colam-Ainsworth et al., 1989; Livesey et al., 1998; Webster, 2001
		38	*manure packed system (vs other types of housing)*		Wenz et al., 2007	
		39	*compost bedded pack barns*		Barberg et al., 2007	Barberg et al., 2007
	housing design: space organisation	40	*space allowance or density*	Dechow et al., 2011	O'Driscoll et al., 2008	Rouha-Mülleder et.al., 2009
		41	*presence of cubicles - overstocking*		Bareille et al., 1998; Fregonesi & Leaver, 2001	Cook et. Al., 2004; Fregonesi & Leaver, 2001; Leonard et al., 1996; Rouha-Mülleder et.al., 2009
		42	*feeding area*			
		43	*exercice area*			Wells et al., 1995
		44	*access to pasture*	Thomsen et al., 2006; Burow et al., 2011	Ivemeyer et al., 2009; O'Driscoll et al., 2008	Somers et al., 2003
		45	*free access to an enclosure from the cow shed*		Barnouin et al., 2004	
		46	*free access to pasture*	Burow et al., 2011		
		47	*Lying space accessibility*		Bareille et al., 1998	
		48	*facilities for diseased animals*			
		49	*multiple-animal area for maternity cows*	McConnel et al., 2008		

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		50	*individual-animal area for maternity cows*	McConnel et al., 2008		
	housing design: material	51	*brisket board*			
		52	*neck rail*		Fulwider et al., 2007	Rouha-Mülleder et.al., 2009
		53	*cow trainers*			Alban et al., 1996; Bergsten & Pettersson, 1992
		54	*size of cubicles*		Fulwider et al., 2007	
		55	*feeding area design (size, number, steps, …)*			
	housing design: floor	56	*floor type (slatted floor…)*		Osteras & Lund, 1988 a; Ruud et al., 2010; Valde et al., 1997	
		57	*bedding material*	Dechow et al., 2011; Weigel et al., 2003	Bewley et al., 2001; (Dufour et al., 2011); Elbers et al., 1998; Fulwider et al., 2007; Jayarao et al., 2004; Ivemeyer et al., 2009; Wenz et al., 2007; Wilson et al., 1997	Barrientos et al., 2013; Cook, 2003 a; Cook et al., 2004; Flower et al., 2007; Fulwider et al., 2007; Hultgren & Bergsten, 2001; Nordlund et. Al., 2004; Norring et. Al., 2008; Rouha-Mülleder et.al., 2009; Vokey et al., 2001; Weary & Taszkun, 2000
		58	*concrete*		Valde et al., 1997	Cook et. Al., 2004; Faye & Lescourret, 1989; Somers et al., 2003; Wells et al., 1995
		59	*Conventional straw yards*		Fregonesi & Leaver, 2001; Peeler et al., 2000; Ward et al., 2002	Somers et al., 2003
		60	*Shavings or sawdust in the calving area*		Nyman et al., 2009	
	handling facilities and walking tracks quality	61	*functionality and use of handling facilities*			
		62	*broken ground*			
		63	*slatted*			
		64	*slippery ground*			
		65	*stairs*			
	housing design: ambience	66	*ventilation system*		Barkema et al., 1998a; Pytlewski et al., 2012	
		67	*lightning*			

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
Management		68	*isolation*			
		69	*indoor humidity*			
		70	*indoor temperature*			
	milking system	71	*type of milking parlour*			
		72	*automatic milking system*	Burow et al., 2011	Berglund et al., 2002; Dohmen et al., 2010; Kelton et al., 2001; Klungel et al., 2000; Kruip et al., 2002; Hovinen & Pyorala, 2011; Hovinen et al., 2009; Rasmussen et al., 2001-2002; Svennersten-Sjaunja and Pettersson, 2008; van der Vorst et al., 2002; Zecconi et al., 2003;	
		73	*Technical milking machine and management variables*			
		74	*traditional milking system*	Burow et al., 2011		
	stockmanship	75	*stockpeople behaviours towards cows: herding practices (hurry…), handling and milking practices*			Chesterton et al, 1989; Clarkson & Ward, 1991
		76	*empathy*			
		77	*human-animal relationship*		Ivemeyer et al., 2011	
		78	*lower threshold for euthanasia*	Thomsen & Sorensen, 2008		
		79	*attitude towards work*		Hutton et al., 1990 ; Wronski et al, 2007	
		80	*action procedure implementation*			
		81	*proper drug usage*			
		82	*member of a milk control program*	Raboisson et al., 2011	Moxley et al., 1978	
		83	*efficient estrus detection*		Khaitsa et al., 2000	
	cleaning	84	*cleanliness of alley*		DeVries et al., 2012; Köster et al., 2006	Chesterton et al., 1989; Wells et al., 1995

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		85	*cleanliness of stalls*		Bareille et al., 1998; Bartlett et al., 1992; DeVries et al., 2012; Köster et al., 2006; Mtaallah et al., 2002; Pytlewski et al., 2012; Schreiner & Ruegg, 2003;	
		86	*clean water troughs*		Goodger et al., 1988	
		87	*disinfectant used in backflushed solution*		Hutton et al., 1991	
		88	*clean exercise area (e.g. scraped < 1time/day)*		Barnouin et al., 2004; Mtaallah et al., 2002	
		89	*unhygienic conditions*		Elbers et al., 1998; Hutton et al., 1990; Sharma et al., 2011; Wilson et al., 1997	
		90	*cleanliness of milking machine*		Agabriel et al., 1997	
		91	*cleanliness of bedding and facilities*			Barrientos et al., 2013
	herd-health	92	*detection of risk of injuries*	Sandgren et al., 2009	Hutton et al., 1990; Koster et al., 2006	
		93	*training for detecting/reporting diseases and welfare issues*			
		94	*docked tail*		Wenz et al., 2007	
		95	*frequency of veterinary examinations*		Wilson et al., 1997	
		96	*nutritional supplementation*	McConnel et al., 2008		
		97	*hoof care*	Dechow et al., 2011	Osteras & Lund, 1988 a; Sato et al., 2008	Faye & Lescourret, 1989; Fjeldaas et al., 2006
		98	*regular adjustment of milking machine*			
		99	*separate management of ruminant species*			
		100	*hair clipping (tail, udder...)*		Barkema et al., 1998b; (Dufour et al., 2011); Rodrigues et al., 2005	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		101	*health control during dry period*			
		102	*vaccination (biosecurity)*	McConnel et al., 2008	Giraudo et al., 1997; Green et al., 2007; Nickerson et al., 1999; Noguera et al., 2011; Tenhagen et al., 2001; Wenz et al., 2007	
		103	*treatment with antibiotics (at least once a year)*	McConnel et al., 2008	Bastan et al., 2010; Duffield et al., 2008; McDougall et al., 2004; Sharma et al., 2007; Sharma, 2008; Wronski et al., 2007	
		104	*cow brought onto the operation (biosecurity)*		Wenz et al., 2007	
		105	*administering bST (Bovine Somatrophin/bovine growth hormone)*	McConnel et al., 2008		
		106	*declared mastitis*		Agabriel et al., 1997	
		107	*quarantine policy*		Ivemeyer et al., 2011	
		108	*calving pen used for sick cows*		De Vliegher et al., 2004	
		109	*mastitis detection / checking*		Barnouin et al., 2004 ; Lievaart et al., 2007 Barnouin et al., 2004; Wilson et al., 1997	
		110	*(clinical mastitis) management*		Barkema et al., 1998a; Barnouin et al., 2004; Detilleux et al., 2013; De Vliegher et al., 2004; Rodrigues et al., 2005	Leach et al., 2012
		111	*footbath at the farm*			Amory et al., 2006
		112	*fly control*			
	milking procedures	113	*wearing gloves during milking*		Bach et al., 2008; (Dufour et al., 2011); Hutton et al., 1990; Rodrigues et al., 2005;	
		114	*milking frequency*		Kruip et al., 2002	
		115	*no written milking procedures*		Rodrigues et al., 2005	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		116	*milking in different places in summer and winter*		Agabriel et al., 1997	
		117	*high SCC cows and clinical mastitis cases milked with a specific unit*		Barnouin et al., 2004; Mtaallah et al., 2002	
		118	*use hired milkers*		Bartlett et al., 1992	
		119	*no stripping squirts of milk before milking (machine stripping)*		Detilleux et al., 2013; Elbers et al., 1998; Mtaallah et al., 2002	
		120	*premilking teat preparation*		Barkema et al., 1998b; Barnouin et al., 2004; Bartlett et al., 1992; Detilleux et al., 2013; Erskine et al., 1987a; Koster et al., 2006; Nyman et al., 2009; Wilson et al., 1997	
		121	*post-milking teat dipping/disinfection (PMTD)*		Bareille et al., 2000; Barkema et al., 1998a; Barnouin et al., 2004; Bodoh et al., 1976; Detilleux et al., 2013; Dohoo & Meek, 1982; (Dufour et al., 2011); Elbers et al., 1998; Erskine & Eberhard, 1991; Erskine et al., 1987a; Fadlelmoula et al., 2007; Fenlon et al., 1995; Goodhope et al., 1980; Hutton et al., 1991; Khaitsa et al., 2000; Koster et al., 2006; Moxley et al., 1978; Mtaalllah et al., 2002; Olde Riekerink et al., 2012; Oliver et al., 2001; Peeler et al., 2000; Singh & Singh, 2002; Whist et al., 2006; Wronski et al., 2007	
		122	*Vacuum not turned off before unit removal*		Hutton et al., 1991	
		123	*High SCC and/or clinical mastitis cows milked last*		(Dufour et al., 2011) Barnouin et al., 2004; Hutton et al., 1990; Wilson et al., 1995	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		124	*using automatic take-offs (automatic cluster removal)*		Barkema et al., 1998b; (Dufour et al., 2011); Hutton et al., 1990; Jayarao et al., 2004; Rodrigues et al., 2005; Smith et Ely, 1997; Wenz et al., 2007	
		125	*Disinfection between milkings*		Fadlelmoula et al., 2007	
		126	*yearly inspection of the milking system*		Agabriel et al., 1997; (Dufour et al., 2011); Erskine et al., 1987a; Fenlon et al., 1995; Hutton et al., 1990; Peeler et al., 2000; Pytlewski et al., 2012; Rodrigues et al., 2005	
		127	*waiting time before/after milking*		Barkema et al., 1998a; Barnouin et al., 2004; DeVries et al., 2010; DeVries et al., 2011; DeVries et al., 2012; (Dufour et al., 2011); Peeler et al., 2000; Watters et al., 2013	Nordlund et. Al., 2004
		128	*surveillance of dry-cow udders for mastitis*		Barkema et al., 1998a; Bareille et al., 2000; (Dufour et al., 2011)	
		129	*dry-cow therapy ("prophylactic dry cow treatment")*		Barkema et al., 1998a; Bodoh et al., 1976; (Dufour et al., 2011); Dingwell et al., 2003; Dohoo & Meek, 1982; Erskine et al., 1987a; Erskine & Eberhart, 1991; Goodhope & Meek, 1980; Green et al., 2007; Hutton et al., 1991; Laven & Lawrence, 2008; Nickerson, 2009; Rodrigues et al., 2005; Runciman et al., 2010; Smith et al., 1985; Wenz et al., 2007; Whist et al., 2006; Wronski et al., 2007	
		130	*high-producing cows milked first*		Hutton et al., 1990	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		131	*number of milkers per shift/month/units per person*		Goodhope & Meek, 1980; Mtaallah et al., 2002; Rodrigues et al., 2005	
		132	*Cow-by-cow premilking preparation and attachment sequence*		Barkema et al., 1998a	
		133	*use paper towels (vs. cloth towels)*		Bach et al., 2008	
		134	*Restraint measures at milking*		Svensson et al., 2006	
		135	*residual calf suckling in dual-purpose breed*		Gonzales-Sedano et al., 2010	
		136	*vibration and noise*		Gygax & Nosal, 2006	
	Dynamics	137	*herd size*	Alvasen et al., 2012; Batra et al., 1971; Dechow & Goodling, 2008; Dechow et al., 2011; Hadley et al., 2006; McConnel et al., 2008; Miller et al., 2008; Pinedo et al., 2010;Smith et al., 2000; Thomsen et al., 2006	Agabriel et al., 1997; Bodoh et al., 1976; Fadlelmoula et al., 2007; Osteras & Lund, 1988 a; Valde et al., 1997; Wenz et al., 2007; Wilesmith et al., 1986; Wronski et al., 2007	Alban, 1994; Cramer et. Al., 2009
		138	*replacement rate*	Raboisson et al., 2011; Thomsen et al., 2006	Fenlon et al., 1995; Goodhope & Meek, 1980; Peeler et al., 2000	
		139	*low percentage of cows culled less than 50 DIM (Days In Milk)*	McConnel et al., 2008		
		140	*calving interval*	Alvasen et al., 2012; McConnel et al., 2008; Raboisson et al., 2011		
		141	*autumn calving peak*			
		142	*superovulatory drugs*			
		143	*increased age at first calving*		Waage et al., 1998	
	around calving	144	*isolation of heifers before or during calving*		Nyman et al., 2009; Svensson et al., 2006	
		145	*suckling calf on early lactation*		Wenz et al., 2007	
		146	*hoof condition at calving*			Cramer et al., 2009

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		147	*treatment of heifers and calves*		Bareille et al. , 2000; De Vliegher et al., 2004; Nyman et al., 2009; Svensson et al., 2006	
		148	*season of calving*		Agabriel et al., 1997	
		149	*management changes after calving*			
		150	*transition cow management (three weeks before calving to three weeks after calving)*		Khaitsa et al., 2000	
		151	*routine drenching (injection of liquid in the rumen) of fresh cows (recently calved)*	McConnel et al., 2008		
		152	*time remaining open (cow not in calf)*			
		153	*dry cows housed in a different location then milking cows (vs another area of the same shed)*		Barnouin et al., 2004	
		154	*cleaning the calving pen after each calving*		Bareille et al, 2000; Barkema et al., 1998a; De Vliegher et al, 2004; (Dufour et al., 2011); Elbers et al., 1998; Green et al., 2007; Peeler et al., 2000	
		155	*separation from calf*			Rouha-Mülleder et.al., 2009
	group	156	*grouping primiparous*			
		157	*rules for regrouping animals (age, health…)*			
		158	*having a beef cow unit (housing system?)*	Raboisson et al., 2011		
		159	*mixing groups*		Dohoo & Meek, 1982	
	movement	160	*access to pasture*		Wronski et al., 2007	Olmos et. Al., 2009
		161	*Time spent on pasture / grazing time*	Alvasen et al., 2012; Burow et al., 2011; Dechow et al., 2011; Thomsen et al., 2006	Green et al., 2007	Alban et al., 1996

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		162	*restricted grazing*			
		163	*summer pasture (pasture-grazing during summer)*	Thomsen et al., 2006	Waage et al., 1998	
		164	*frequency of exercise*			
		165	*organic dairy*	Alvasen et al., 2012; Thomsen et al., 2006	Haskell et al., 2009	Sanders et al., 2009
		166			Muller & Sauerwein., 2010	
	feeding	167	*water quality*			
		168	*concentrate consumption*	Norgaard et al., 1999		Livesey et al., 1998; Nocek, 1997
		169	*occasional diet supplementation / Providing supplemental minerals and vitamins for lactating cows*		Barkema et al., 1998a; Barnouin et al., 1998b; Chew et al., 1982; (Dufour et al., 2011); Erskine et al., 1987b; Hogan et al., 1993; LeBlanc et al., 2002; Mutoni et al., 2012; Pytlewski et al., 2012; Sharma & Maiti, 2005; Smith et al., 1984; Weiss et al, 1990-1997; Wenz et al., 2005; Wenz et al., 2007; Wronski et al., 2007;	Amory et al., 2006; Faye & Lescourret, 1989; Hedges et al., 2001;
		170	*forage or fibre availability*		Nyman et al., 2009; Waage et al., 1998; Wronski et al., 2007	Faye & Lescourret, 1989; Offer et al, 2003; Wells et al., 1995
		171	*feeding frequency*		Agabriel et al., 1997	
		172	*TMR feeding (Total Mixed Ration)*	Dechow et al., 2011; McConnel et al., 2008		
		173	*using forage test results to balance rations*	McConnel et al., 2008		
		174	*using MUN to determining ration composition (Milk Urea Nitrogen Test)*	McConnel et al., 2008		
		175	*negative energy balance*	Dechow & Goodling, 2008; McConnel et al., 2010		Collard et al., 2000
		176	*hay in contact with cats or poultry (Botulism risk)*	Galey et al., 2000; Livesey et al., 2004; Otter et al., 2006; Payne et al., 2009; Wobeser et al., 1997		

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
		177	*nutrition of calves and heifers*		Bareille et al., 2000; Barkema et al., 1998a; Barnouin et al., 2004; De Vliegher et al., 2004; Ivemeyer et al., 2009; (Kesler, 1981); Lievaart et al., 2007; Nyman et al., 2009; Roberson et al., 1994a; Svensson et al., 2006; Svensson et al., 2006	
		178	*nutrition of transition cows*		Barkema et al., 1998a; Barnouin et al., 2004	
		179	*nutrition of dry cows*		Barkema et al., 1998a; Erskine et al., 1987b	
		180	*nutrition of milking cows*		Bareille et al., 1998; Smith & Ely, 1997	
	culling policy	181	*rules for culling (rigorous culling policy)*	Dechow & Goodling, 2008; McConnel et al., 2008 Pinedo et al., 2010	Bareille et al., 1998; Barkema et al., 1998b; Barnouin et al., 2004;; Rodrigues et al., 2005; Whist et al., 2006	
		182	*milk quota*			
		183	*disease control programmes*			
		184	*economics (milk price, replacement heifers price, cull cow price…)*	Bar et al., 2008; Hadley et al., 2006		
		185	*economic incentives to lower SCC*			
Sampling		186	*Sampling methods*		Barkema et al., 1997; Brooks et al., 1983; Buelow et al., 1996; Djabri et al., 2002; Greer & Pearson, 1976; Holdaway et al., 1996; Jaartsveld et al., 1983; Madsen, 1979; Miller et al., 1986; Natzke et al., 1972; Ostensson et al., 1988; Paape & Tucker, 1966; Timms et al., 1986; Woolford et al., 1998;	

Entity classes	Type of parameters		Parameter	Mortality	Mastitis/high SCC	Lameness
Environment		187	*Number of samples*		Buelow et al., 1996; Mattila et al., 1986; Poutrel & Rainard, 1982; Sears et al., 1990	
		188	*outdoor temperature*			
		189	*cold flooring*		Ewbank, 1968	
		190	*season of the year*	Alvasen et al., 2012; Faye & Pérochon, 1995; Hertl et al., 2011; McConnel et al., 2009; Miller et al., 2008; Pinedo et al., 2010; Vitali et al., 2009	Agabriel et al., 1997; Barnouin et al., 2004; Bodoh et al., 1976; Dohoo & Meek, 1982; Igono et al., 1988; Khate & Yadav, 2010; Olde Riekerink et al., 2007; Skrzypek et al., 2004; Smith et al., 1985 (IMI); Waage et al., 1998	Alban et al., 1996; Cook, 2003a
		191	*rain (wet climate)*			
		192	*geographical region*	McConnel et al., 2008; Raboisson et al., 2011	Sato et al., 2008; Waage et al., 1998; Wenz et al., 2007	
		193	*warm humid conditions / heat related stress / Temperature Humidity Index*	Crescio et al., 2010; Smith et al., 2000; Vitali et al., 2009	Igono et al., 1988; Morse et al. 1988; Osteras & Lund, 1988 a	
		194	*low reporting*			
		195	*inhabitant density*	Raboisson et al., 2011		

Reference list for Appendix P

Agabriel C, Coulon JB, Sibra C, Journal C and Hauwuy A, 1997. Facteurs de variation de la numération cellulaire du lait en exploitation. Annales de Zootechnie 46, 13–19.

Agger JF, 1983. Production disease and mortality in dairy cows. Analysis of records from disposal plants from 1960-1982, in: 5. International Conference on Production Disease in Farm Animals, Uppsala (Sweden), 10-12 Aug 1983.

Alban L, 1994. Lameness in Danish dairy cows: frequency and possible risk factors. Preventive veterinary medicine 22, 213–225.

Alban L, 1995. Lameness in Danish dairy cows: frequency and possible risk factors. Preventive Veterinary Medicine 22, 213–225. doi:10.1016/0167-5877(94)00411-B

Alban L, Agger JF and Lawson LG, 1996. Lameness in tied Danish dairy cattle: the possible influence of housing systems, management, milk yield, and prior incidents of lameness. Preventive Veterinary Medicine 29, 135–149.

Alvåsen K, Jansson Mörk M, Hallén Sandgren C, Thomsen PT and Emanuelson U, 2012. Herd-level risk factors associated with cow mortality in Swedish dairy herds. Journal of dairy science 95, 4352–62.

Amory JR, Kloosterman P, Barker ZE, Wright JL, Blowey RW and Green LE, 2006. Risk factors for reduced locomotion in dairy cattle on nineteen farms in The Netherlands. Journal of dairy science 89, 1509–15.

Anderson DE and Muir WW, 2005. Pain management in ruminants. The Veterinary clinics of North America. Food animal practice 21, 19–31.

Archer SC, Green MJ and Huxley JN, 2010. Association between milk yield and serial locomotion score assessments in UK dairy cows. Journal of dairy science 93, 4045–53.

Bach A, Valls N, Solans A and Torrent T, 2008. Associations between nondietary factors and dairy herd performance. Journal of dairy science 91, 3259–67.

Bar D, Gröhn YT, Bennett G, González RN, Hertl JA, Schulte HF, Tauer LW, Welcome FL and Schukken YH, 2008. Effects of repeated episodes of generic clinical mastitis on mortality and culling in dairy cows. Journal of dairy science 91, 2196–204.

Barberg AE, Endres MI, Salfer JA and Reneau JK, 2007. Performance and welfare of dairy cows in an alternative housing system in Minnesota. Journal of dairy science 90, 1575–83.

Bareille N, Kiebre-Toe MB, Beaudeau F and Seegers H, 2000. Risk factors for elevated milk somatic cell count during early lactation in dairy heifers, in: 7èmes Rencontres Autour Des Recherches Sur Les Ruminants. Paris, France, 6-7 Décembre, 2000. pp. 99–102.

Bareille N, Seegers H, Fourichon C, Beaudeau F and Malher X, 1998. Survenue et expression des mammites cliniques et subcliniques en troupeaux bovins laitiers□: facteurs de risque lies. Rencontres Recherches Ruminants 5, 297–300.

Barkema HW, Schukken YH, Lam TJ, Beiboer ML, Benedictus G and Brand A, 1999. Management practices associated with the incidence rate of clinical mastitis. Journal of dairy science 82, 1643–54.

Barkema HW, Schukken YH, Lam TJ, Beiboer ML, Benedictus G and Brand A, 1998a. Management practices associated with low, medium, and high somatic cell counts in bulk milk. Journal of dairy science 81, 1917–27.

Barkema HW, Schukken YH, Lam TJ, Beiboer ML, Wilmink H, Benedictus G and Brand A, 1998b. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. Journal of dairy science 81, 411–9.

Barkema HW, Schukken YH, Lam TJ, Galligan DT, Beiboer ML and Brand A, 1997. Estimation of interdependence among quarters of the bovine udder with subclinical mastitis and implications for analysis. Journal of dairy science 80, 1592–9.

Barnouin J, Chassagne M, Bazin S and Boichard D, 2004. Management practices from questionnaire surveys in herds with very low somatic cell score through a national mastitis program in France. Journal of dairy science 87, 3989–99.

Barrientos AK, Chapinal N, Weary DM, Galo E and von Keyserlingk MAG, 2013. Herd-level risk factors for hock injuries in freestall-housed dairy cows in the northeastern United States and California. Journal of dairy science 96, 3758–65.

Bartlett P, Miller G, Lance S and Heider L, 1992. Managerial determinants of intramammary coliform and environmental streptococci infections in Ohio dairy herds. Journal of Dairy Science 75, 1241–1252.

Bastan A, Cengiz M, Cengiz S, Polat B, Colak A, Akan M, Isfendiyar D and Acar DB, 2010. Effects of precalving antibiotic treatment on mastitis and individual somatic cell count in heifers. Journal of Animal and Veterinary Advances 9, 1245–1249.

Batra TR, Burnside EB and Freeman MG, 1971. Canadian dairy cow disposals II. Effect of herd size and production level on dairy cow disposal patterns. Canadian Journal of Animal Science 51, 85–87.

Beckley MS and Johnson T, 1966. Five year study of a California mastitis test on a commercial dairy herd. Journal of dairy science 49, 746.

Berglund I, Pettersson G and Svennersten-Sjaunja K, 2002. Automatic milking: effects on somatic cell count and teat end-quality. Livestock Production Science 78, 115–124.

Bergsten C and Petterson B, 1992. The cleanliness of cows tied in stalls and the health of their hooves as influenced by the use of electric trainers. Preventive veterinary medicine 13, 229–238.

Berry DP, Lee JM, Macdonald K a, Stafford K, Matthews L and Roche JR, 2007. Associations among body condition score, body weight, somatic cell count, and clinical mastitis in seasonally calving dairy cattle. Journal of dairy science 90, 637–48.

Bewley J, Palmer RW and Jackson-Smith DB, 2001. A comparison of free-stall barns used by modernized Wisconsin dairies. Journal of dairy science 84, 528–41.

Blackburn PS, 1966. The variation in cell count of cow's milk throughout lactation and from one lactation to the next. Journal of dairy research 33, 193–198.

Bodoh GW, Battista WJ, Schultz LH and Johnston RP, 1976. Variation in Somatic Cell Counts in Dairy Herd Improvement Milk Samples. Journal of Dairy Science 59, 1119–1123.

Boettcher PJ, Dekkers JC, Warnick LD and Wells SJ, 1998. Genetic analysis of clinical lameness in dairy cattle. Journal of dairy science 81, 1148–56.

Breen JE, Bradley A J and Green MJ, 2009. Quarter and cow risk factors associated with a somatic cell count greater than 199,000 cells per milliliter in United Kingdom dairy cows. Journal of dairy science 92, 3106–15.

Brooks BW, Barnum D A and Meek A H, 1983. An observational study of Corynebacterium bovis in selected Ontario dairy herds. Canadian journal of comparative medicine. Revue canadienne de médecine comparée 47, 73–8.

Buelow KL, Goodger WJ, Collinsb MT, Clayton MK, Nordlund K V and Thomas CB, 1996. A model to determine sampling strategies and milk inoculum volume for detection of intramammary Staphylococcus aureus infections in dairy cattle by bacteriological culture. Preventive Veterinary Medicine mix, 343–355.

Burow E, Thomsen PT, Sørensen JT and Rousing T, 2011. The effect of grazing on cow mortality in Danish dairy herds. Preventive veterinary medicine 100, 237–41.

Chesterton RN, Pfeiffer DU, Morris RS and Tanner CM, 1989. Environmental and behavioural factors affecting the prevalence of foot lameness in New Zealand dairy herds - a case-control study. New Zealand veterinary journal 37, 135–42.

Chew BP, Hollen LL, Hillers JK and Herlugson ML, 1982. Relationship between vitamin A and beta-carotene in blood plasma and milk and mastitis in Holsteins. Journal of dairy science 65, 2111–2118.

Clarkson DA and Ward WR, 1991. Farm tracks, stockman's herding and lameness in dairy cattle. Veterinary Record 129, 511–512.

Colam-Ainsworth P, Lunn GA, Thomas RC and Eddy RG, 1989. Behaviour of cows in cubicles and its possible relationship with laminitis in replacement dairy heifers. The Veterinary record 125, 573–5.

Collard BL, Boettcher PJ, Dekkers JC, Petitclerc D and Schaeffer LR, 2000. Relationships between energy balance and health traits of dairy cattle in early lactation. Journal of dairy science 83, 2683–2690.

Compton CWR, Heuer C, Parker K and McDougall S, 2007. Risk factors for peripartum mastitis in pasture-grazed dairy heifers. Journal of dairy science 90, 4171–80.

Cook NB, 2003. Prevalence of lameness among dairy cattle in Wisconsin as a function of housing type and stall surface. Journal of the American Veterinary Medical Association 223, 1324–8.

Cook NB, Bennett TB and Nordlund K V, 2004. Effect of free stall surface on daily activity patterns in dairy cows with relevance to lameness prevalence. Journal of dairy science 87, 2912–22.

Cramer G, Lissemore KD, Guard CL, Leslie KE and Kelton DF, 2009. Herd-level risk factors for seven different foot lesions in Ontario Holstein cattle housed in tie stalls or free stalls. Journal of dairy science 92, 1404–11.

Craven N and Williams MR, 1985. Defences of the bovine mammary gland against infection and prospects for their enhancement. Veterinary immunology and immunopathology 10, 71–127.

Crescio MI, Forastiere F, Maurella C, Ingravalle F and Ru G, 2010. Heat-related mortality in dairy cattle: A case crossover study. Preventive veterinary medicine 97, 191–7.

De Haas Y, Barkema HW and Veerkamp RF, 2002. Genetic parameters of pathogen-specific incidence of clinical mastitis. Animal Science 74, 233–242.

De Vliegher S, Laevens H, Barkema HW, Dohoo IR, Stryhn H, Opsomer G and de Kruif A, 2004. Management practices and heifer characteristics associated with early lactation somatic cell count of Belgian dairy heifers. Journal of dairy science 87, 937–47. doi:10.3168/jds.S0022-0302(04)73238-5

Dechow CD and Goodling RC, 2008. Mortality, culling by sixty days in milk, and production profiles in high- and low-survival Pennsylvania herds. Journal of dairy science 91, 4630–9.

Dechow CD, Smith EA and Goodling RC, 2011. The effect of management system on mortality and other welfare indicators. Animal Welfare 20, 145–158.

DeGraves FJ and Fetrow J, 1993. Economics of mastitis and mastitis control. The Veterinary Clinics of North America. Food Animal Practice 9, 421–424.

Dematawena CMB and Berger PJ, 1997. Effect of Dystocia on Yield, Fertility, and Cow Losses and an Economic Evaluation of Dystocia Scores for Holsteins. Journal of Dairy Science 80, 754–761.

Dematawewa CM and Berger PJ, 1998. Genetic and phenotypic parameters for 305-day yield, fertility, and survival in Holsteins. Journal of dairy science 81, 2700–9.

Detilleux J, Theron L, Duprez J-N, Reding E, Humblet M-F, Planchon V, Delfosse C, Bertozzi C, Mainil J and Hanzen C, 2013. Structural equation models to estimate risk of infection and tolerance to bovine mastitis. Genetics, selection, evolution 45, 6.

Devries TJ, Aarnoudse MG, Barkema HW, Leslie KE and von Keyserlingk MAG, 2012. Associations of dairy cow behavior, barn hygiene, cow hygiene, and risk of elevated somatic cell count. Journal of dairy science 95, 5730–9.

Devries TJ, Deming JA, Rodenburg J, Seguin G, Leslie KE and Barkema HW, 2011. Association of standing and lying behavior patterns and incidence of intramammary infection in dairy cows milked with an automatic milking system. Journal of dairy science 94, 3845–55.

DeVries TJ, Dufour S and Scholl DT, 2010. Relationship between feeding strategy, lying behavior patterns, and incidence of intramammary infection in dairy cows. Journal of dairy science 93, 1987–97.

Dingwell RT, Kelton DF and Leslie KE, 2003. Management of the dry cow in control of peripartum disease and mastitis. The Veterinary clinics of North America. Food animal practice 19, 235–65.

Djabri B, Bareille N and Beaudeau F, 2002. Quarter milk somatic cell count in infected dairy cows□: a meta-analysis. Veterinary Research 33, 335–357.

Dohmen W, Neijenhuis F and Hogeveen H, 2010. Relationship between udder health and hygiene on farms with an automatic milking system. Journal of dairy science 93, 4019–33.

Dohoo IR and Meek AH, 1982. Somatic cell counts in bovine milk. The Canadian veterinary journal. La revue vétérinaire canadienne 23, 119–25.

Duffield TF, Rabiee AR and Lean IJ, 2008. A meta-analysis of the impact of monensin in lactating dairy cattle. Part 3. Health and reproduction. Journal of dairy science 91, 2328–41.

Dufour S, Fréchette A, Barkema HW, Mussell A and Scholl DT, 2011. Invited review: effect of udder health management practices on herd somatic cell count. Journal of dairy science 94, 563–79.

Eberhart RJ, Hutchinson LJ and Spencer SB, 1982. Relationships of bulk tank somatic cell counts to prevalence of intra-mammary infection and to indexes of herd production. Journal of food protection 45, 1125–1128.

Eberhart RJ, Natzke RP, Newbould FHS, Nonnecke BJ and Thompson P, 1979. Coliform mastitis - A review. Journal of dairy science 62, 1–22.

Elbers ARW, Miltenburg JD, De Lange D, Crauwels AP, Barkema HW and Schukken YH, 1998. Risk factors for clinical mastitis in a random sample of dairy herds from the southern part of The Netherlands. Journal of dairy science 81, 420–6.

Emanuelson U and Funke H, 1991. Effect of milk yield on relationship between bulk milk somatic cell count and prevalence of mastitis. Journal of dairy science 74, 2479–83.

Enemark JMD, 2008. The monitoring, prevention and treatment of sub-acute ruminal acidosis (SARA): a review. Veterinary journal (London, England□: 1997) 176, 32–43. doi:10.1016/j.tvjl.2007.12.021

Erb HN, Smith RD, Oltenacu PA, Guard CL, Hillman RB, Powers PA, Smith MC and White ME, 1985. Path model of reproductive disorders and performance, milk fever, mastitis, milk yield, and culling in Holstein cows. Journal of dairy science 68, 3337–49.

Erskine RJ and Eberhart RJ, 1991. Post-milking teat dip use in dairy herds with high or low somatic cell counts. Journal of the American Veterinary Medical Association 199, 1734–6.

Erskine RJ, Eberhart RJ, Hutchinson LJ and Scholz RW, 1987a. Blood selenium concentrations and glutathione peroxidase activities in dairy herds with high and low somatic cell counts. Journal of the American Veterinary Medical Association 190, 1417–21.

Erskine RJ, Eberhart RJ, Hutchinson LJ and Spencer SB, 1987b. Herd management and prevalence of mastitis in dairy herds with high and low somatic cell counts. Journal of the American Veterinary Medical Association 190, 1411–6.

Esslemont RJ and Kossaibati MA, 1997. Culling in 50 dairy herds in England. Veterinary Record 140, 36–39.

Ewbank R, 1968. An experimental demonstration of the effect of surface cooling upon the health of the bovine mammary gland. Veterinary Record 84, 685–686.

Fadlelmoula AA, Fahr AA, Anacker G and Swalve HH, 2007. The Effect of Management Factors on Somatic-cell Counts and Specific Mastitis Causing Pathogens in Large Scale Dairy Units. Research Journal of Animal and Veterinary Sciences 2, 24–27.

Faye B and Lescourret F, 1989. Environmental factors associated with lameness in dairy cattle. Preventive veterinary medicine 7, 267–287.

Faye B and Pérochon L, 1995. La mortalité des vaches laitières dans l'enquête écopathologique Bretagne. Veterinary Research 26, 124–131.

Fenlon DR, Logue DN, Gunn J and Wilson J, 1995. A study of mastitis bacteria and herd management practices to identify their relationship to high somatic cell counts in bulk tank milk. The British veterinary journal 151, 17–25.

Fjeldaas T, Sogstad AM and Østerås O, 2006. Claw trimming routines in relation to claw lesions, claw shape and lameness in Norwegian dairy herds housed in tie stalls and free stalls. Preventive veterinary medicine 73, 255–71.

Flower FC, de Passillé AM, Weary DM, Sanderson DJ and Rushen J, 2007. Softer, higher-friction flooring improves gait of cows with and without sole ulcers. Journal of dairy science 90, 1235–42.

Fregonesi JA and Leaver JD, 2001. Behaviour, performance and health indicators of welfare for dairy cows housed in strawyard or cubicle systems. Livestock Production Science 68, 205–216.

Fulwider WK, Grandin T, Garrick DJ, Engle TE, Lamm WD, Dalsted NL and Rollin BE, 2007. Influence of free-stall base on tarsal joint lesions and hygiene in dairy cows. Journal of dairy science 90, 3559–66.

Galey FD, Terra R, Walker R, Adaska J, Etchebarne MA, Puschner B, Fisher E, Whitlock RH, Rocke T, Willoughby D and Tor E, 2000. Type C Botulism in Dairy Cattle from Feed Contaminated with a Dead Cat. Journal of Veterinary Diagnostic Investigation 12, 204–209.

Giraudo JA, Calzolari A, Rampone H, Rampone A, Giraudo AT, Bogni C, Larriestra A and Nagel R, 1997. Field trials of a vaccine against bovine mastitis. 1. Evaluation in heifers. Journal of dairy science 80, 845–53.

González-Sedano M, Marın-Mejia B, Maranto MI, Leme de Magalhães-Labarthe AC and Alonso-Diaz MA, 2010. Effect of residual calf suckling on clinical and sub-clinical infections of mastitis in dual-purpose cows: epidemiological measurements. Research in veterinary science 89, 362–6.

Goodger WJ, Galland JC and Christiansen VE, 1988. Survey of milking management practices on large dairies and their relationship to udder health and production variables. Journal of dairy science 71, 2535–42.

Goodhope RG and Meek AH, 1980. Factors associated with mastitis in Ontario dairy herds: a case control study. Canadian journal of comparative medicine. Revue canadienne de médecine comparée 44, 351–7.

Green LE, Hedges VJ, Schukken YH, Blowey RW and Packington AJ, 2002. The impact of clinical lameness on the milk yield of dairy cows. Journal of dairy science 85, 2250–6.

Green LE, Schukken YH and Green MJ, 2006. On distinguishing cause and consequence: do high somatic cell counts lead to lower milk yield or does high milk yield lead to lower somatic cell count? Preventive veterinary medicine 76, 74–89.

Green MJ, Bradley AJ, Medley GF and Browne WJ, 2007. Cow, farm, and management factors during the dry period that determine the rate of clinical mastitis after calving. Journal of dairy science 90, 3764–76.

Greer DO and Pearson JK, 1976. Reproducibility of electronic cell counts in milk; a study of 5 further factors. The Journal of dairy research 43, 371–9.

Gröhn YT, Eicker SW, Ducrocq V and Hertl JA, 1998. Effect of diseases on the culling of Holstein dairy cows in New York State. Journal of dairy science 81, 966–78.

Gygax L and Nosal D, 2006. Short communication: contribution of vibration and noise during milking to the somatic cell count of milk. Journal of dairy science 89, 2499–502.

Hadley GL, Wolf CA and Harsh SB, 2006. Dairy cattle culling patterns, explanations, and implications. Journal of dairy science 89, 2286–96.

Hare E, Norman HD and Wright JR, 2006. Survival rates and productive herd life of dairy cattle in the United States. Journal of dairy science 89, 3713–20.

Harris BL, 1989. New Zealand dairy cow removal reasons and survival rate. New Zealand Journal of Agricultural Research 32, 355–358.

Haskell MJ, Langford FM, Jack MC, Sherwood L, Lawrence AB and Rutherford KMD, 2009. The effect of organic status and management practices on somatic cell counts on UK dairy farms. Journal of dairy science 92, 3775–80.

Hedges J, Blowey RW, Packington AJ, O'Callaghan CJ and Green LE, 2001. A longitudinal field trial of the effect of biotin on lameness in dairy cows. Journal of dairy science 84, 1969–75.

Hertl JA, Schukken YH, Bar D, Bennett GJ, González RN, Rauch BJ, Welcome FL, Tauer LW and Gröhn YT, 2011. The effect of recurrent episodes of clinical mastitis caused by gram-positive and gram-negative bacteria and other organisms on mortality and culling in Holstein dairy cows. Journal of dairy science 94, 4863–77.

Hogan JS, Weiss WP and Smith KL, 1993. Role of vitamin E and selenium in host defense against mastitis. Journal of dairy science 76, 2795–803.

Holdaway RJ, Holmes CW and Steffert IJ, 1996. A comparison of indirect methods for diagnosis of subclinical intramammary infection in lactating dairy cows. Australian Journal of Dairy Technology 51, 64–71.

Hovinen M and Pyörälä S, 2011. Invited review: udder health of dairy cows in automatic milking. Journal of dairy science 94, 547–62. doi:10.3168/jds.2010-3556

Hovinen M, Rasmussen MD and Pyörälä S, 2009. Udder health of cows changing from tie stalls or free stalls with conventional milking to free stalls with either conventional or automatic milking. Journal of dairy science 92, 3696–703. doi:10.3168/jds.2008-1962

Hultgren J and Bergsten C, 2001. Effects of a rubber-slatted flooring system on cleanliness and foot health in tied dairy cows. Preventive veterinary medicine 52, 75–89.

Hultgren J, Manske T and Bergsten C, 2004. Associations of sole ulcer at claw trimming with reproductive performance, udder health, milk yield, and culling in Swedish dairy cattle. Preventive veterinary medicine 62, 233–51.

Hutton CT, Fox LK and Hancock DD, 1990. Mastitis control practices: differences between herds with high and low milk somatic cell counts. Journal of dairy science 73, 1135–43.

Hutton CT, Fox LK and Hancock DD, 1991. Risk factors associated with herd-group milk somatic cell count and prevalence of coagulase-positive staphylococcal intramammary infections. Preventive Veterinary Medicine 11, 25–35.

Igono MO, Johnson HD, Steevens BJ, Hainen WA and Shanklin MD, 1988. Effect of season on milk temperature, milk growth hormone, prolactin, and somatic cell counts of lactating cattle. International Journal of Biometeorology 32, 194–200.

Ivemeyer S, Knierim U and Waiblinger S, 2011. Effect of human-animal relationship and management on udder health in Swiss dairy herds. Journal of dairy science 94, 5890–902.

Ivemeyer S, Walkenhorst M, Heil F, Notz C, Maeschli A, Butler G and Klocke P, 2009. Management factors affecting udder health and effects of a one year extension program in organic dairy herds. Animal 3, 1596–604.

Jaartsveld FHJ, Van Puffelen E, Oskam J, Tielen MJM, Verstegen MWA and Albers GAA, 1983. Somatic cell counts in milk of dairy cows in relation to stage of lactation, age, production level and presence of pathogens. Netherlands Milk and Dairy Journal 37, 79–90.

Jayarao BM, Pillai SR, Sawant AA, Wolfgang DR and Hegde NV, 2004. Guidelines for monitoring bulk tank milk somatic cell and bacterial counts. Journal of dairy science 87, 3561–73.

Kelton DF, Rodenburg J and Hand K, 2001. Udder health and milk quality on Ontario dairy farms utilizing voluntary milking systems, in: 2nd International Symposium on Mastitis and Milk Quality. pp. 410–414.

Kesler EM, 1981. Feeding mastitic milk to calves: review. Journal of dairy science 64, 719–23.

Khaitsa ML, Wittum TE, Smith KL, Henderson JL and Hoblet KH, 2000. Herd characteristics and management practices associated with bulk-tank somatic cell counts in herds in official Dairy Herd Improvement Association programs in Ohio. American journal of veterinary research 61, 1092–8.

Khate K and Yadav BR, 2010. Incidence of mastitis in Sahiwal cattle and Murrah buffaloes of a closed organized herd. The Indian Journal of Animal Sciences 80.

Klungel GH, Slaghuis BA and Hogeveen H, 2000. The effect of the introduction of automatic milking systems on milk quality. Journal of dairy science 83, 1998–2003.

Koldeweij E, Emanuelson U and Janson L, 1999. Relation of milk production loss to milk somatic cell count. Acta veterinaria Scandinavica 40, 47–56.

Köster G, Tenhagen B-A, Scheibe N and Heuwieser W, 2006. Factors Associated with High Milk Test Day Somatic Cell Counts in Large Dairy Herds in Brandeburg. II. Milking Practices. Journal of veterinary medicine. A, Physiology, pathology, clinical medicine 53, 209–214.

Kruip TAM, Morice H, Robert M and Ouweltjes W, 2002. Robotic milking and its effect on fertility and cell counts. Journal of dairy science 85, 2576–81.

Laevens H, Deluyker H, Schukken YH, De Meulemeester L, Vandermeersch R, De Muêlenaere E and De Kruif A, 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. Journal of dairy science 80, 3219–26.

Laursen M V, Boelling D and Mark T, 2009. Genetic parameters for claw and leg health, foot and leg conformation, and locomotion in Danish Holsteins. Journal of dairy science 92, 1770–7.

Laven RA and Lawrence KE, 2008. Efficacy of blanket treatment of cows and heifers with an internal teat sealant in reducing the risk of mastitis in dairy cattle calving on pasture. New Zealand veterinary journal 56, 171–5.

Leach KA, Tisdall DA, Bell NJ, Main DCJ and Green LE, 2012. The effects of early treatment for hindlimb lameness in dairy cows on four commercial UK farms. Veterinary journal (London, England□: 1997) 193, 626–32.

LeBlanc SJ, Duffield TF, Leslie KE, Bateman KG, TenHag J, Walton JS and Johnson WH, 2002. The effect of prepartum injection of vitamin E on health in transition dairy cows. Journal of dairy science 85, 1416–26.

Leonard FC, O'Connel JM and Farrell O, 1996. Effect of overcrowding on claw health in first-calved Friesan heifers. British Veterinary Journal 152, 459–472.

Lewis S, Cockcroft PD, Bramley RA and Jackson PGG, 2000. The likelihood of subclinical mastitis in quarters with different types of teat lesions in the dairy cow. Cattle Practice 8, 293–299.

Lievaart JJ, Barkema HW, Kremer WDJ, van den Broek J, Verheijden JHM and Heesterbeek JAP, 2007. Effect of herd characteristics, management practices, and season on different categories of the herd somatic cell count. Journal of dairy science 90, 4137–44.

Lindström UB, Kenttämies H, Arstila J and Tuovila R, 1981. Usefulness of cell counts in predicting bovine mastitis. Acta Agriculturae Scandinavica 31, 199–203.

Livesey CT, Harrington T, Johnston AM, May SA and Metcalf JA, 1998. The effect of diet and housing on the development of sole haemorrhages, white line haemorrhages and heel erosions in Holstein heifers. Animal Science 67, 9–16.

Madsen PS, 1979. Influence of storage and preservation of milk samples on microscopic and Fossomatic somatic cell counts. Nordisk veterinaermedicin 31, 449–54.

Mattila T, Pyörälä S and Sandholm M, 1986. Veterinary Research Comparison of milk antitrypsin, albumin, N-Acetyl-B-D-Glucosaminidase, somatic cells and bacteriological analysis as indicators of bovine subclinical mastitis. Veterinary Research Communications 10, 113–124.

McCarthy S, Berry DP, Dillon P, Rath M and Horan B, 2007. Effect of strain of Holstein–Friesian and feed system on udder health and milking characteristics. Livestock Science 107, 19–28.

McConnel CS, Garry FB, Hill AE, Lombard JE and Gould DH, 2010. Conceptual modeling of postmortem evaluation findings to describe dairy cow deaths. Journal of dairy science 93, 373–86.

McConnel CS, Garry FB, Lombard JE, Kidd JA, Hill AE and Gould DH, 2009. A necropsy-based descriptive study of dairy cow deaths on a Colorado dairy. Journal of dairy science 92, 1954–62.

McConnel CS, Lombard JE, Wagner BA and Garry FB, 2008. Evaluation of factors associated with increased dairy cow mortality on United States dairy operations. Journal of dairy science 91, 1423–32.

McDonald JS and Anderson AJ, 1981. Total and differential somatic cell counts in secretions from noninfected bovine mammary glands: the early nonlactating period. American Journal of Veterinary Research 42, 1360–1365.

McDougall S, Young L and Anniss FM, 2004. Production and health of pasture-fed dairy cattle following oral treatment with the ionophore lasalocid. Journal of dairy science 87, 2967–76.

Menzies FD, Bryson DG, McCallion T and Matthews DI, 1995. A study of mortality among suckler and dairy cows in Northern Ireland in 1992. Veterinary Records 137, 531–536.

Milian-Suazo F, Erb HN and Smith RD, 1988. Descriptive epidemiology of culling dairy cows from 34 herds in New York State. Preventive Veterinary Medicine 6, 243–251.

Miller RH, Kuhn MT, Norman HD and Wright JR, 2008. Death losses for lactating cows in herds enrolled in dairy herd improvement test plans. Journal of dairy science 91, 3710–5.

Miller RH, Paape MJ and Acton JC, 1986. Comparison of milk somatic cell counts by Coulter and Fossomatic Counters. Journal of dairy science 69, 1942–6.

Miller RH, Paape MJ, Peters RR and Young MD, 1990. Total and differential somatic cell counts and N-acetyl-beta-D-glucosaminidase activity in mammary secretions during dry period. Journal of dairy science 73, 1751–5.

Mitchell RG, Mather RE, Swallow WH and Randy HA, 1976. Effects of a corticosteroid and diuretic agent on udder edema and milk yield in dairy cows. Journal of dairy science 59, 109–112.

Morse D, DeLorenzo MA, Wilcox CJ, Collier RJ, Natzke RP and Bray DR, 1988. Climatic effects on occurrence of clinical mastitis. Journal of dairy science 71, 848–53.

Moxley JE, Kennedy BW, Downey BR and Bowman JST, 1978. Survey of Milking Hygiene Practices and Their Relationships to Somatic Cell Counts and Milk Production. Journal of Dairy Science 61, 1637–1644.

Mtaallah B, Oubey Z and Hammami H, 2002. Estimation des pertes de production en lait et des facteurs de risque des mammites subcliniques à partir des numérations cellulaires de lait de tank en élevage bovin laitier. Revue de Médecine Vétérinaire 153, 251–260.

Müller U and Sauerwein H, 2010. A comparison of somatic cell count between organic and conventional dairy cow herds in West Germany stressing dry period related changes. Livestock Science 127, 30–37.

Mutoni G, Prasad S, De K, Pal S, Mukherjee J, Kapila S, Kapila R, Kaur H, Mohanty AK and Dan AK, 2014. Effect of supplementation of vitamin E , copper and zinc around peripartum on udder health , milk yield and composition of Sahiwal cows. Livestock Research for Rural Development 24, 1–6.

Myllys V and Rautala H, 1995. Characterization of clinical mastitis in primiparous heifers. Journal of dairy science 78, 538–45.

Natzke RP, Everett RW, Guthrie RS, Keown JF, Meek AM, Meril WG, Roberts SJ and Schmidt GH, 1972. Mastitis control program: effect on milk production. Journal of dairy science 42, 774–783.

Nickerson SC, 2009. Control of heifer mastitis: antimicrobial treatment - an overview. Veterinary microbiology 134, 128–35.

Nickerson SC, Owens WE and Boddie RL, 1995. Mastitis in dairy heifers: initial studies on prevalence and control. Journal of dairy science 78, 1607–18.

Nickerson SC, Owens WE, Tomita GM and Widel PW, 1999. Vaccinating dairy heifers with a Staphylococcus aureus bacterin reduces mastitis at calving. Large animal practice 20, 16–28.

Nocek JE, 1997. Bovine acidosis: implications on laminitis. Journal of dairy science 80, 1005–28.

Noguera M, March R, Guix R, Prenafeta A and Foix A, 2011. Evaluation of the efficacy of a new vaccine against bovine mastitis caused by CNS field trial results, in: Proc. NMC Annual Meeting, Arlington, VA. Natl. Mastitis Counc., Madison, WI. pp. 187–188.

Nordlund K V, Cook NB and Oetzel GR, 2004. Investigation Strategies for Laminitis Problem Herds. Journal of dairy science 87, 27–35.

Nørgaard NH, Lind KM and Agger JF, 1999. Cointegration analysis used in a study of dairy-cow mortality. Preventive veterinary medicine 42, 99–119.

Norring M, Manninen E, de Passillé AM, Rushen J, Munksgaard L and Saloniemi H, 2008. Effects of sand and straw bedding on the lying behavior, cleanliness, and hoof and hock injuries of dairy cows. Journal of dairy science 91, 570–6.

Nyman A-K, Emanuelson U, Gustafsson AH and Persson Waller K, 2009. Management practices associated with udder health of first-parity dairy cows in early lactation. Preventive veterinary medicine 88, 138–49.

O'Driscoll K, Boyle L, French P, Meaney B and Hanlon A, 2008. The effect of out-wintering pad design on dirtiness score, somatic cell score and mastitis incidence in dairy cows. Animal□: an international journal of animal bioscience 2, 912–20.

Offer JE, Leach KA, Brocklehurst S and Logue DN, 2003. Effect of Forage Type on Claw Horn Lesion Development in Dairy Heifers. The Veterinary Journal 165, 221–227.

Olde Riekerink RGM, Barkema HW, Stryhn H, 2007. The effect of season on somatic cell count and the incidence of clinical mastitis. Journal of dairy science 90, 1704–15. doi:10.3168/jds.2006-567

Olde Riekerink RGM, Ohnstad I, van Santen B and Barkem AHW, 2012. Effect of an automated dipping and backflushing system on somatic cell counts. Journal of dairy science 95, 4931–8.

Oliver SP, Gillespie BE, Lewis MJ, Ivey SJ, Almeida RA, Luther DA, Johnson DL, Lamar KC, Moorehead HD and Dowlen HH, 2001. Efficacy of a new premilking teat disinfectant containing a phenolic combination for the prevention of mastitis. Journal of dairy science 84, 1545–9.

Olmos G, Boyle L, Hanlon A, Patton J, Murphy JJ and Mee JF, 2009. Hoof disorders, locomotion ability and lying times of cubicle-housed compared to pasture-based dairy cows. Livestock Science 125, 199–207.

Ostensson K, Hageltorn M and Aström G, 1988. Differential cell counting in fraction-collected milk from dairy cows. Acta veterinaria Scandinavica 29, 493–500.

Osterås O and Lund A, 1988. Epidemiological analyses of the associations between bovine udder health and housing. Preventive veterinary medicine 6, 79–90.

Otter A, Livesey C, Hogg R, Sharpe R and Gray D, 2006. Risk of botulism in cattle and sheep arising from contact with broiler litter. Veterinary Record 159, 186–187.

Paape MJ, Tucker HA and 1966. Somatic cell content variation in fraction-collected milk. Journal of dairy science 49, 265–7.

Payne JH, Hogg RA, Otter A, Roest HIJ and Livesey CT, 2011. Emergence of suspected type D botulism in ruminants in England and Wales (2001 to 2009), associated with exposure to broiler litter. The Veterinary record 168, 640.

Pearson JK and Mackie DP, 1979. Factors associated with the occurrence, cause and outcome of clinical mastitis in dairy cattle. Veterinary Record 105, 456–463.

Peeler EJ, Green MJ, Fitzpatrick JL, Morgan KL and Green LE, 2000. Risk factors associated with clinical mastitis in low somatic cell count British dairy herds. Journal of dairy science 83, 2464–72.

Pinedo P, Risco C and Melendez P, 2011. A retrospective study on the association between different lengths of the dry period and subclinical mastitis, milk yield, reproductive performance, and culling in Chilean dairy cows. Journal of dairy science 94, 106–15.

Pinedo PJ, DeVries A and Webb DW, 2010. Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. Journal of dairy science 93, 2250–61.

Poutrel B and Rainard P, 1982. Predicting the probability of quarter infection (by major pathogens) from somatic cell concentration. American journal of veterinary research 43, 1296–9.

Prendiville R, Pierce KM and Buckley F, 2010. A comparison between Holstein-Friesian and Jersey dairy cows and their F1 cross with regard to milk yield, somatic cell score, mastitis, and milking characteristics under grazing conditions. Journal of dairy science 93, 2741–50.

Pytlewski J, Antkowiak I, Adamski M, Kučera J and Skrzypek R, 2012. Factors Associated with Hygienic Quality of Bulk Tank Milk Produced in Central Poland. Annals of Animal Science 12, 227–235.

Raboisson D, Cahuzac E, Sans P and Allaire G, 2011. Herd-level and contextual factors influencing dairy cow mortality in France in 2005 and 2006. Journal of dairy science 94, 1790–803.

Rasmussen MD, Bjerring M, Justesen P and Jepsen L, 2002. Milk quality on Danish farms with automatic milking systems. Journal of dairy science 85, 2869–78.

Rasmussen MD, Blom JY, Nielsen LAH and Justesen P, 2001. Udder health of cows milked automatically. Livestock Production Science 72, 147–156.

Raubertas RF and Shook GE, 1982. Relationship Between Lactation Measures of Somatic Cell Concentration and Milk Yield. Journal of Dairy Science 65, 419–425.

Reichmuth J, 1975. Somatic cell counting - Interpretation of results, in: Seminar on Mastitis Control, 7–11 April, Reading. International Dairy Federation, Brussels. pp. 93–109.

Reneau JK, Seykora AJ, Heins BJ, Endres MI, Farnsworth RJ and Bey RF, 2005. Association between hygiene scores and somatic cell scores in dairy cattle. Journal of the American Veterinary Medical Association 227, 1297–301.

Roberson JR, Fox LK, Hancock DD, Gay CC and Besser TE, 1994. Coagulase-positive Staphylococcus intramammary infections in primiparous dairy cows. Journal of dairy science 77, 958–69.

Rodrigues ACO, Caraviello DZ and Ruegg PL, 2005. Management of Wisconsin dairy herds enrolled in milk quality teams. Journal of dairy science 88, 2660–71.

Rogers GW, Banos G, Sander Nielsen U and Philipsson J, 1998. Genetic Correlations Among Somatic Cell Scores, Productive Life, and Type Traits from the United States and Udder Health Measures from Denmark and Sweden. Journal of Dairy Science 81, 1445–1453.

Rouha-Mülleder C, Iben C, Wagner E, Laaha G, Troxler J and Waiblinger S, 2009. Relative importance of factors influencing the prevalence of lameness in Austrian cubicle loose-housed dairy cows. Preventive veterinary medicine 92, 123–33.

Runciman DJ, Malmo J and Deighton M, 2010. The use of an internal teat sealant in combination with cloxacillin dry cow therapy for the prevention of clinical and subclinical mastitis in seasonal calving dairy cows. Journal of dairy science 93, 4582–91.

Ruud LE, Bøe KE and Osterås O, 2010. Associations of soft flooring materials in free stalls with milk yield, clinical mastitis, teat lesions, and removal of dairy cows. Journal of dairy science 93, 1578–86.

Sanders AH, Shearer JK and De Vries A, 2009. Seasonal incidence of lameness and risk factors associated with thin soles, white line disease, ulcers, and sole punctures in dairy cattle. Journal of dairy science 92, 3165–74.

Sandgren CH, Lindberg A and Keeling LJ, 2009. Using a national dairy database to identify herds with poor welfare. Animal Welfare 18, 523–532.

Sargeant JM, Scott HM, Leslie KE, Ireland MJ and Bashiri A, 1998. Clinical mastitis in dairy cattle in Ontario: frequency of occurrence and bacteriological isolates. The Canadian veterinary journal. La revue vétérinaire canadienne 39, 33–8.

Sato K, Bartlett PC, Alban L, Agger JF and Houe H, 2008. Managerial and environmental determinants of clinical mastitis in Danish dairy herds. Acta veterinaria Scandinavica 50, 4.

Schreiner DA and Ruegg PL, 2003. Relationship between udder and leg hygiene scores and subclinical mastitis. Journal of dairy science 86, 3460–5.

Schutz MM, Hansen LB, Steuernagel GR and Kuck A L, 1990. Variation of Milk, Fat, Protein, and Somatic Cells for Dairy Cattle. Journal of Dairy Science 73, 484–493.

Sears PM, Smith BS, English PB, Herer PS and Gonzalez RN, 1990. Shedding pattern of Staphylococcus aureus from bovine intramammary infections. Journal of dairy science 73, 2785–9.

Seegers H, Fourichon C and Beaudeau F, 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Veterinary Research 34, 475–491.

Sharma N, 2008. Foot and mouth disease-mastitis cascade in dairy cattle: a field study. International Journal of Zoological Research 4, 64–67.

Sharma N, Gupta SK, Sharma U and Hussain K, 2007. Treatment of clinical mastitis in a buffaloes: a case report. Buffalo Bulletin 26, 56–58.

Sharma N and Maiti SK, 2005. Effect of dietary supplementation of vitamin E and selenium in sub clinical mastitis in dairy cows. Indian Journal of Veterinary Medicine 25, 76.

Sharma N, Singh NK and Bhadwal MS, 2011. Relationship of somatic cell count and mastitis: An overview. Journal of animal science 24, 429–438.

Sheldrake RF, Hoare RJ and McGregor GD, 1983. Lactation stage, parity, and infection affecting somatic cells, electrical conductivity, and serum albumin in milk. Journal of dairy science 66, 542–7.

Shook GE, 1993. Genetic improvement of mastitis through selection on somatic cell count. The Veterinary clinics of North America. Food animal practice 9, 563–81.

Sieber RL and Farnsworth RJ, 1981. Prevalence of chronic teat-end lesions and their relationship to intramammary infection in 22 herds of dairy cattle. Journal of the American Veterinary Medical Association 178, 1263–7.

Singh S and Sing M, 2002. Post Milking Teat Dip Effect on Somatic Cell Count , Milk Production and Composition in Cows and Buffaloes. Asian Australasian Journal of Animal Sciences 15, 1517–1522.

Skrzypek R, Wójtowski J and Fahr R-D, 2004. Factors affecting somatic cell count in cow bulk tank milk--a case study from Poland. Journal of veterinary medicine. A, Physiology, pathology, clinical medicine 51, 127–31.

Smith JW and Ely LO, 1997. The Influence of Feeding and Housing Systems on Production, Reproduction, and Somatic Cell Count Scores of Southern Holstein Herds. The Professional Animal Scientist 13, 155–161.

Smith JW, Ely LO and Chapa AM, 2000. Effect of region, herd size, and milk production on reasons cows leave the herd. Journal of dairy science 83, 2980–7.

Smith KL, Harrison JH, Hancock DD, Todhunter DA and Conrad HR, 1984. Effect of vitamin E and selenium supplementation on incidence of clinical mastitis and duration of clinical symptoms. Journal of dairy science 67, 1293–300.

Smith KL, Todhunter DA and Schoenberger PS, 1985. Environmental Pathogens and Intramammary Infection During the Dry Period. Journal of Dairy Science 68, 402–417.

Sogstad ÅM, Osterås O and Fjeldaas T, 2006. Bovine Claw and Limb Disorders Related to Reproductive Performance and Production Diseases. Journal of Dairy Science 89, 2519–2528.

Somers JGCJ, Frankena K, Noordhuizen-Stassen EN and Metz JHM, 2003. Prevalence of claw disorders in Dutch dairy cows exposed to several floor systems. Journal of dairy science 86, 2082–93.

Steeneveld W, Hogeveen H, Barkema HW, van den Broek J and Huirne RBM, 2008. The influence of cow factors on the incidence of clinical mastitis in dairy cows. Journal of dairy science 91, 1391–402.

Stevenson MA and Lean IJ, 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 482–488.

Svennersten-Sjaunja KM and Pettersson G, 2008. Pros and cons of automatic milking in Europe. Journal of animal science 86, 37–46.

Svensson C, Linder A and Olsson S-O, 2006. Mortality in Swedish dairy calves and replacement heifers. Journal of dairy science 89, 4769–77.

Tenhagen BA, Edinger D, Baumgärtner B, Kalbe P, Klünder G and Heuwieser W, 2001. Efficacy of a herd-specific vaccine against Staphylococcus aureus to prevent post-partum mastitis in dairy heifers. Journal of veterinary medicine. A, Physiology, pathology, clinical medicine 48, 601–7.

Thomsen PT and Houe H, 2006. Dairy cow mortality. A review. The Veterinary quarterly 28, 122–9.

Thomsen PT, Kjeldsen AM, Sørensen JT and Houe H, 2004. Mortality (including euthanasia) among Danish dairy cows (1990-2001). Preventive veterinary medicine 62, 19–33.

Timms LL, Sommer DA and Schultz LH, 1986. Role of minor pathogens in mastitis. Journal of dairy science 68, 202.

Tyler JW and Thurmond MC, Lasslo L, 1989. Relationship between test-day measures of somatic cell count and milk production in California dairy cows. Canadian journal of veterinary research 53, 182–7.

Valde JP, Hird DW, Thurmond MC and Osterås O, 1997. Comparison of ketosis, clinical mastitis, somatic cell count, and reproductive performance between free stall and tie stall barns in Norwegian dairy herds with automatic feeding. Acta veterinaria Scandinavica 38, 181–92.

Valde JP, Lawson LG, Lindberg A, Agger JF, Saloniemi H and Osterås O, 2004. Cumulative risk of bovine mastitis treatments in Denmark, Finland, Norway and Sweden. Acta veterinaria Scandinavica 45, 201–10.

Van den Borne BHP, van Schaik G, Lam TJGM and Nielen M, 2010. Variation in herd level mastitis indicators between primi- and multiparae in Dutch dairy herds. Preventive veterinary medicine 96, 49–55.

Van der Vorst Y and de Koning K, 2002. Automatic milking systems and milk quality in three European countries, in: First North American Conference on Robotic Milking, March 20-22, 2012. pp. 1–13.

Vitali A, Segnalini M, Bertocchi L, Bernabucci U, Nardone A and Lacetera N, 2009. Seasonal pattern of mortality and relationships between mortality and temperature-humidity index in dairy cows. Journal of dairy science 92, 3781–90.

Vokey FJ, Guard CL, Erb HN and Galton DM, 2001. Effects of alley and stall surfaces on indices of claw and leg health in dairy cattle housed in a free-stall barn. Journal of dairy science 84, 2686–99.

Waage S, Sviland S and Odegaard S A, 1998. Identification of risk factors for clinical mastitis in dairy heifers. Journal of dairy science 81, 1275–84.

Walsh S, Buckley F, Berry DP, Rath M, Pierce K, Byrne N and Dillon P, 2007. Effects of breed, feeding system, and parity on udder health and milking characteristics. Journal of dairy science 90, 5767–79.

Ward WR, Hughes JW, Faull WB, Cripps PJ, Sutherland JP and Sutherst JE, 2002. Observational study of temperature, moisture, pH and bacteria in straw bedding, and faecal consistency, cleanliness and mastitis in cows in four dairy herds. Veterinary Record 151, 199–206.

Warnick LD, Janssen D, Guard CL and Gröhn YT, 2001. The effect of lameness on milk production in dairy cows. Journal of dairy science 84, 1988–97.

Watson EN, David GP and Cook AJC, 2008. Review of diagnostic laboratory submissions of adult cattle “found dead” in England and Wales in 2004. Veterinary Record 163, 531–535.

Watters MEA, Meijer KMA, Barkema HW, Leslie KE, von Keyserlingk MAG and Devries TJ, 2013. Associations of herd- and cow-level factors, cow lying behavior, and risk of elevated somatic cell count in free-stall housed lactating dairy cows. Preventive veterinary medicine 111, 245–55.

Weary DM and Taszkun I, 2000. Hock lesions and free-stall design. Journal of dairy science 83, 697–702.

Webster AJ, 2001. Effects of housing and two forage diets on the development of claw horn lesions in dairy cows at first calving and in first lactation. Veterinary journal (London, England : 1997) 162, 56–65.

Webster AJF, 2002. Effects of housing practices on the development of foot lesions in dairy heifers in early lactation. Veterinary Record 151, 9–12.

Weigel KA, Palmer RW and Caraviello DZ, 2003. Investigation of factors affecting voluntary and involuntary culling in expanding dairy herds in Wisconsin using survival analysis. Journal of dairy science 86, 1482–6.

Weiss WP, Hogan JS, Smith KL and Hoblet KH, 1990. Relationships among selenium, vitamin E, and mammary gland health in commercial dairy herds. Journal of dairy science 73, 381–90.

Weiss WP, Hogan JS, Todhunter DA and Smith KL, 1997. Effect of vitamin E supplementation in diets with a low concentration of selenium on mammary gland health of dairy cows. Journal of dairy science 80, 1728–37. 8

Wells SJ, Trent AM, Marsh WE, McGovern PG and Robinson RA, 1993. Individual cow risk factors for clinical lameness in lactating dairy cows. Preventive veterinary medicine 17, 95–109.

Wells SJ, Trent AM, Marsh WE, Williamson NB and Robinson RA, 1995. Some risk factors associated with clinical lameness in dairy herds in Minnesota and Wisconsin. The Veterinary record 136, 537–40.

Wenz JR, Garry FB, Lombard JE, Elia R, Prentice D and Dinsmore RP, 2005. Short communication: Efficacy of parenteral ceftiofur for treatment of systemically mild clinical mastitis in dairy cattle. Journal of dairy science 88, 3496–9.

Wenz JR, Jensen SM, Lombard JE, Wagner BA and Dinsmore RP, 2007. Herd management practices and their association with bulk tank somatic cell count on United States dairy operations. Journal of dairy science 90, 3652–9.

Whist a C, Østerås O and Sølverød L, 2006. Clinical mastitis in Norwegian herds after a combined selective dry-cow therapy and teat-dipping trial. Journal of dairy science 89, 4649–59.

White F and Rattray EAS, 1965. Diurnal variation in the cell content of cows' milk. Journal of comparative pathology 75, 253–61.

Wilesmith JW, Francis PG and Wilson CD, 1986. Incidence of clinical mastitis in a cohort of British dairy herds. Veterinary Record 118, 49–55.

Wilson DJ, Das HH, Gonzalez RN and Sears PM, 1997. Association between management practices, dairy herd characteristics, and somatic cell count of bulk tank milk. Journal of the American Veterinary Medical Association 210, 1499–502.

Wobeser G, Baptiste K, Clark EG and Deyo AW, 1997. Type C botulism in cattle in association with a botulism die-off in waterfowl in Saskatchewan. The Canadian veterinary journal 38, 782.

Woolford MW, Williamson JH and Henderson HV, 1998. Changes in electrical conductivity and somatic cell count between milk fractions from quarters subclinically infected with particular mastitis pathogens. The Journal of dairy research 65, 187–98.

Wroński M, Jarmuż W and Skrzypek R, 2007. Factors associated with the levels of somatic cells and microorganisms in bulk tank milk. Polish Journal of Food and Nutrition Sciences 57, 481–485.

Zecconi A, Piccinini R, Casirani G, Binda E and Migliorati L, 2003. Effects of automatic milking system on teat tissues, intramammary infections and somatic cell counts. Italian Journal of Animal Science 2, 275–282.

Appendix Q. Relevance scores attributed by the experts to each parameter in the shortlist of factors of variation identified in Objective 3.

Consortium member:	1	2	3	4	5	6	7	8	9	10	11	Mean	SD	Median
Breed	4	3	3	3	2	4	4	3	1	3	3	3.0	0.9	3
Milk production/yield	3	3	4	4	3	4	4	4	2	3	4	3.5	0.7	4
Udder conformation	4	1	2	2	3	2	2	1	2	2	2	2.1	0.8	2
Weight	3	1	2	2	2	2	3	5	1	2	2	2.3	1.1	2
Age	4	4	4	3	4	3	5	4	3	4	4	3.8	0.6	4
Body condition score	4	3	5	5	3	5	4	5	2	2	2	3.6	1.3	4
Days in milk/after calving	2	5	5	4	3	4	3	3	3	4	4	3.6	0.9	4
Lactation and reproductive status	2	2	5	4	2	3	4	2	3	2	3	2.9	1.0	3
Parity	4	4	4	5	3	5	4	4	4	5	4	4.2	0.6	4
Reproductive problems	5	2	4	4	5	3	3	5	3	2	4	3.6	1.1	4
Injuries, accidents, pain	5	2	4	5	5	4		4	3	2	4	3.8	1.1	4
Udder status	3	2	4	5	5	4	2	2	4	3	4	3.5	1.1	4
Hoof status	5	3	5	4	4	4	3	3	4	5	4	4.0	0.8	4
Housing system	4	4	2	5	3	5	4	4	5	5	5	4.2	1.0	4
Milking system	2	2	2	3	3	5	3	2	4	3	3	2.9	0.9	3
Density	5	4	4	4	3	5	4	5	5	5	4	4.4	0.7	4
Outside/exercise area	5	4	3	2	2	3	4	4	1	3	3	3.1	1.1	3
Access to pasture	5	4	3	4	1	5	4	4	2	3	4	3.5	1.2	4
Calving pen	5	2	2	3	3	4	3	3	3	4	3	3.2	0.9	3
Beef cow unit	2	2	2	1	1	1	2	1	1	1	1	1.4	0.5	1
Contact with other species	1	1	2	1	1	1	2	2	1	1	1	1.3	0.5	1
Ergonomics	5	3	4	4	2	3	4	3	2	4	2	3.3	1.0	3
Bedding material	4	3	3	4	3	4	3	3	3	4	4	3.5	0.5	3
Floor type	4	4	5	4	3	4	3	3	3	4	4	3.7	0.6	4
Floor quality	4	3	5	4	3	4	3	3	3	4	4	3.6	0.7	4
Indoor environment	5	3	1	4	2	3	3	3	3	4	3	3.1	1.0	3
Behaviour towards the animals	5	3	3	4	1	1	3	5	1	3	4	3.0	1.5	3
Empathy, threshold to recognise disease and pain, call veterinarian	5	3	5	4	3	4	5	4	4	5	4	4.2	0.8	4
Ability to detect/report diseases and welfare issues	4	3	5	4	3	4	5	4	5	5	4	4.2	0.8	4
Organic dairy production	4	3	2	3	1	4	3	3	2	1	2	2.5	1.0	3
Available workforce at milking	3	2	4	2	2	3	2	3	1	2	2	2.4	0.8	2
Relation with the outside	4	2	5	3	3	3	4	3	4	3	2	3.3	0.9	3
Rules for mixing animals	3	2	3	2	2	2	2	4	2	3	3	2.5	0.7	2
Calving interval	3	2	4	4	2	1	3	3	1	3	2	2.5	1.0	3
Cleanliness of alleys and lying area	4	3	3		2	4	3	4	4	3	3	3.3	0.7	3
Cleanliness of water trough and manger	4	2	5		1	3	2	4	1	2	3	2.7	1.3	2.5
Manure used for bedding	4	2	4		3	3	3	3	4	4	3	3.3	0.7	3
Animal care/hygiene	5	3	5	4	3	4	3	3	4	4	4	3.8	0.8	4
Biosecurity	3	3	3	2	1	2	4	3	4	3	3	2.8	0.9	3
Veterinary treatments	5	3	4	5	3	3	5	3	4	4	4	3.9	0.8	4
Order of milking	3	2	5	3	2	4	2	2	2	3	3	2.8	1.0	3
Milking procedure	4	2	4	3	2	4	2	2	4	3	3	3.0	0.9	3
Milking machine cleanliness	3	2	5	3	4	4	2	2	3	3	3	3.1	0.9	3
Hygiene procedures during milking	4	2	5	3	4	4	2	3	4	3		3.4	1.0	3.5
Management of dry cows	4	3	4	3	3	4	4	3	5	4	3	3.6	0.7	4
Housing of dry cows	4	2	4	4	1	3	4	2	4	4	3	3.2	1.1	4
Water quality and availability	5	3	5	4	2	4	3	5	2	3	3	3.5	1.1	3
Concentrate consumption	3	2	5	4	2	3	4	3	3	4	3	3.3	0.9	3

Consortium member:	1	2	3	4	5	6	7	8	9	10	11	Mean	SD	Median
Forage or fibre availability / % in the ration	3	2	5	2	2	3	4	4	3	4	4	3.3	1.0	3
Supplementation with minerals and vitamins	4	2	4	2	1	3	4	3	1	3	3	2.7	1.1	3
Cows blocked at feeding	4	2	4	3	1	3	3	3	1	1	3	2.5	1.1	3
Feeding frequency	4	1	3	3	1	3	3	3	2	2	2	2.5	0.9	3
TMR feeding	3	2	3	2	1	3	3	3	2	2	2	2.4	0.7	2
Negative energy balance	5	2	4	4	2	3	4	4	3	4	3	3.5	0.9	4
Nutrition of calves	4	1	4	1			3	4	1	3	2	2.6	1.3	3
Specific nutrition according to lactation or physiological status	3	2	4	4	2	3		4	3	4	2	3.1	0.9	3
Culling rules	4	2	4	5	3	4	3	3	4	4	4	3.6	0.8	4
Replacement rate and origin	3	3	3	3	2	3	3	2	3	2	4	2.8	0.6	3
Economics	3	2	4	2	3	4	3		4	4	3	3.2	0.8	3
Herd size	4	4	3	5	1	4	4	3	1	2	4	3.2	1.3	4
Outdoor conditions	4	3	1	5	3	3	4		5	4	4	3.6	1.2	4
Geographical region	4	2	2	4	1	4	4	3	1	3	2	2.7	1.2	3
Milk quota	4	2	3	3	1	3	2	4	1	1	2	2.4	1.1	2

Appendix R. Opinion of the experts on the feasibility to collect and to keep the parameters updated (in bold: parameters selected by at least 8 members)

Consortium member:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Sum**
Breed	x	x	x	x	x	x	x	x	x	x	x	**11**
Milk production/yield	x	x	x	x	x	x	x	x	x	x	x	**11**
Udder conformation												0
Weight			x									1
Age	x			x	x	x	x	x	x	x	x	**9**
Body condition score			x				x					2
Days in milk/after calving	x	x	x	x	x		x	x	x	x	x	**10**
Lactation and reproductive status	x		x	x			x			x	x	6
Parity	x	x	x	x	x	x	x	x	x	x	x	**11**
Reproductive problems	x		x					x				3
Injuries, accidents, pain	x		x									2
Udder status												0
Hoof status	x			x								2
Housing system	x	x	x	x	x	x	x		x	x	x	**10**
Milking system	x	x	x	x	x	x	x		x	x	x	**10**
Density	x		x	x			x		x	x		6
Outside/exercise area	x		x	x			x	x	x	x	x	**8**
Access to pasture	x	x	x	x		x	x	x	x	x	x	**10**
Calving pen	x		x	x			x		x	x		6
Beef cow unit	x			x	x		x		x	x	x	7
Contact with other species							x					1
Ergonomics	x		x	x			x	x		x		6
Bedding material	x		x	x		x	x	x			x	7
Floor type	x	x	x	x		x	x	x	x	x	x	**10**
Floor quality	x		x					x		x		4
Indoor environment	x									x		2
Behaviour towards the animals												0
Empathy, threshold to recognise disease and pain, call veterinarian												0
Ability to detect/report diseases and welfare issues												0
Organic dairy production	x	x		x	x		x	x	x	x	x	**9**
Available workforce at milking	x		x				x		x		x	5
Relation with the outside			x	x			x			x	x	5
Rules for mixing animals							x		x			2
Calving interval	x	x		x	x			x			x	6
Cleanliness of alleys and lying area												0
Cleanliness of water trough and manger			x									1
Manure used for bedding												0
Animal care/hygiene			x									1
Biosecurity							x					1
Veterinary treatments				x	x		x	x	x	x		6
Order of milking			x									1
Milking procedure	x											1
Milking machine cleanliness	x		x	x					x			4
Hygiene procedures during milking	x		x	x								3
Management of dry cows	x		x				x					3
Housing of dry cows	x		x				x		x			4
Water quality and availability	x											1

Consortium member:	1	2	3	4	5	6	7	8	9	10	11	Sum
Concentrate consumption			x									1
Forage or fibre availability / % in the ration			x									1
Supplementation with minerals and vitamins	x		x									2
Cows blocked at feeding							x				x	2
Feeding frequency	x		x	x			x				x	5
TMR feeding	x		x				x		x	x		5
Negative energy balance												0
Nutrition of calves												0
Specific nutrition according to lactation or physiological status	x								x			2
Culling rules			x									1
Replacement rate and origin		x	x	x	x					x	x	6
Economics			x	x								2
Herd size	x	x	x	x	x	x	x	x	x	x	x	**11**
Outdoor conditions	x			x					x	x		4
Geographical region	x	x	x	x	x	x	x	x	x	x	x	**11**
Milk quota	x	x			x		x	x	x	x	x	**8**

Appendix S. Opinion of the experts on the ability of the parameters to characterise a population (in bold: parameters selected by at least 8 members)

Consortium member:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Sum**
Breed	x	x	x	x	x	x	x	x		x	x	**10**
Milk production /yield			x	x	x	x	x	x		x	x	**8**
Udder conformation												0
Weight												0
Age					x						x	2
Body condition score			x					x				2
Days in milk/after calving			x									1
Lactation and reproductive status			x			x					x	3
Parity			x			x		x			x	4
Reproductive problems			x					x				2
Injuries, accidents, pain			x									1
Udder status												0
Hoof status			x									1
Housing system	x	x	x	x		x	x	x	x	x	x	**10**
Milking system	x		x	x		x	x		x	x		7
Density			x									1
Outside exercise area			x					x				2
Access to pasture	x	x	x	x			x	x	x	x	x	**9**
Calving pen			x									1
Beef cow unit	x											1
Contact with other species												0
Ergonomics			x									1
Bedding material			x	x								2
Floor type			x				x					2
Floor quality			x									1
Indoor environment												0
Behaviour towards the animals												0
Empathy, threshold to recognise disease and pain, call veterinarian												0
Ability to detect/report diseases and welfare issues												0
Organic dairy production	x			x		x	x	x	x	x	x	**8**
Available workforce at milking			x									1
Relation with the outside			x									1
Rules for mixing animals												0
Calving interval								x				1
Cleanliness of alleys and lying area												0
Cleanliness of water trough and manger			x									1
Manure used for bedding												0
Animal care/hygiene			x									1
Biosecurity												0
Veterinary treatments												0
Order of milking			x									1
Milking procedure												0
Milking machine cleanliness			x									1
Hygiene procedures during milking			x									1
Management of dry cows			x									1
Housing of dry cows			x									1
Water quality and availability												0
Concentrate consumption			x									1

Consortium member:	1	2	3	4	5	6	7	8	9	10	11	Sum
Forage or fibre availability / % in the ration			x					x				2
Supplementation with minerals and vitamins												0
Cows blocked at feeding												0
Feeding frequency			x									1
TMR feeding			x									1
Negative energy balance												0
Nutrition of calves												0
Specific nutrition according to lactation or physiological status												0
Culling rules			x									1
Replacement rate and origin			x									1
Economics			x									1
Herd size	x	x	x	x	x	x	x	x	x	x		**10**
Outdoor conditions	x			x					x	x		4
Geographical region	x	x	x	x	x	x	x	x	x	x	x	**11**
Milk quota	x							x				2

Appendix T. Results from the data providers on the availability of the factors of variation in the national databases

Parameters	**1. In the database (y/n)**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	Y	Y	Y		Y
Housing system	N	N	Y/N		N
Floor type	N	N	N		N
Days in milk	Y	Y	Y		Y
Access to pasture	N	N	N		N
Milk production	Y	N	Y		Y
Herd size	Y	Y	Y		Y
Breed	Y	Y	Y		Y
Geographical region	Y	Y	Y		Y
Organic dairy production	N	N	Y		N
Lameness	N	N	N		N
Somatic cell count	Y	Y	Y		Y
Mortality	Y	Y	Y		Y
Parameters	**2. On herd level (y/n)**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	N	Y	Y		Y
Housing system	N	Y	Y/N		N
Floor type	N	N	N		N
Days in milk	N	Y	Y		Y
Access to pasture	N	Y	N		N
Milk production	Y	N			Y
Herd size	Y	Y	Y		Y
Breed	N	Y	Y		Y
Geographical region	Y	Y	Y		Y
Organic dairy production	N	N	Y		N
Lameness	N	N	N		N
Somatic cell count	Y	Y	Y		Y
Mortality	Y	Y	Y		Y
Parameters	**3. On animal level (y/n)**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	Y	Y	Y		Y
Housing					

system					
Floor type					
Days in milk	Y	Y	Y		Y
Access to pasture					
Milk production	Y	N			Y
Herd size					
Breed	Y	Y	Y		Y
Geographical region					
Organic dairy production					
Lameness	N	N	N		N
Somatic cell count	Y	Y	Y		Y
Mortality	Y	Y	Y		Y
Parameters	**4. Number of farms a year***				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	~3.600	all	~4000		46855
Housing system	N	Very low (15%)			N
Floor type	N	N	N		N
Days in milk	~3.600	Y	~4000		46855
Access to pasture	N	Very low (15%)	N		N
Milk production	~3.100	N	~4000		46855
Herd size	~3.600	all	~4000		46855
Breed	~3.600	all	~4000		46855
Geographical region	~3.600	most	~4000		46855
Organic dairy production	N	N	N		
Lameness	N	N	N		
Somatic cell count	~3.100	5.500 (for two regions: Lombardia and Emilia Romagna)	~4000		46855
Mortality	~3.600	Y	~4000		12000
Parameters	**5. Number of animals a year**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	~575.000	all	~330000		7 M calvings
Housing system					
Floor type					
Days in milk	~575.000	Y (see below)	~330000		2.5 M lactations
Access to					

pasture					
Milk production	~510.000	N	~330000		22.3 M daily milk productions
Herd size					
Breed	~575.000	all	~330000		6.9 M identities
Geographical region					
Organic dairy production					
Lameness		N			
Somatic cell count	~510.000	N	~330000		22.3 M
Mortality	~575.000	Y	~330000		290000

Parameters	**6. Number of sampling moments a year**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	Continuously	Continuously	Continuously		every calving
Housing system					
Floor type					
Days in milk	Continuously	-	Continuously		every lactation
Access to pasture		-			
Milk production	Depends on farm – 6 or 11 per year	N	Depends on farm - between 6-12 per year		according to ICAR dairy performances protocols
Herd size	Continuously	Continuously	Continuously		Continuously
Breed	Continuously	Continuously	Continuously		
Geographical region		-	Continuously		
Organic dairy production		N	Continuously		
Lameness		N			
Somatic cell count	Depends on farm – but 6 or 11 per year	24 (twice a month)	Continuously		every daily milk production
Mortality	Continuously	Immediate communication of new deaths	Continuously		

Parameters	**7. Scale used in database**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	Numerical	Numerical	Numerical		Numerical
Housing system	N	N	N		
Floor type	N	N	N		
Days in milk	Numerical	Date of start/date of end of lactation	Numerical		Numerical

Access to pasture	N	N	N		
Milk production	Numerical - Different possibilities exist (kg milk produced, kg milk delivered, kg energy corrected milk – per year cow, per farm, etc.)	N	Numerical		Numerical
Herd size	Numerical - Different possibilities exist (CowYears, feeding days, milking days, counting animal numbers…)	Numerical	Numerical		Numerical
Breed	Categorical	Name of the breed	Numerical		National code table
Geographical region	Numerical (Postal code)	GPS location	Numerical		District
Organic dairy production		N	Categorical		
Lameness		N			
Somatic cell count	Numerical (SCC * 1.000)	Numerical	Numerical		Numerical
Mortality	Numerical - Different possibilities exist (number of deaths (from different causes), different rates and risks can be calculated)	Date of birth/ date of death	Numerical		Numerical

Parameters	**8. Method used to collect data**				
Country	**Denmark**	**Italy**	**Sweden**	**Belgium**	**France**
Parity	Calving routinely reported by farmer (mandatory)	Communication by the farmer			
Housing system		Communication by the farmer			

Floor type		n	
Days in milk	Calving routinely reported by farmer (mandatory)	Communication by the farmer/ automatic attribution	
Access to pasture		Communication by the farmer	
Milk production	Electronic milking control performed by farmer 6-11 times/year	N	
Herd size	Calvings/transfer reported by farmer (mandatory)	Communication by the farmer/OVs	
Breed	Calving routinely reported by farmer (mandatory) including information about breed	Communication by the farmer	
Geographical region		Communication by the farmer	
Organic dairy production		n	
Lameness		n	
Somatic cell count	Electronic milking control performed by farmer 6-11 times/year	Opto-fluorometric instrument	Fossomatic
Mortality	Deaths/transfer reported by farmer. Transfers/deaths are reported as dead at farm, euthanized or slaughtered (mandatory).	Communication by the farmer/ OVs	CDB reports

Appendix U. Datamodel

Element Code	Element Name	Type[9]	M	Controlled terminology	Description
A.01	localOrgId	xs:string (100)			Unique identification of the local or regional or national organisation that provided the information
A.02	localOrgCountry	xs:string (2)		COUNTRY	Country where the local organisation is placed. (ISO 3166-1-alpha-2).
B.01	progId	xs:string (100)		"CFT_EFSA_AHAW_2012_01"	Unique identification code of the programme or project for which the sampling unit was taken.
B.04	progType	xs:string (5)		K005A Official (National) programme K009A Official (EU) programme K012A Industry/ private programme K013A Survey	Type of programme recording the indicators
B.07	sampPoint	xs:string (5)		E100A Primary production	Point, in the food chain, where the indicator was recorded
B.08	progInfo	*CompoundType*[21]			Additional specific information and comments on the sampling programme depending on specific requirements of the different data collection domains such as if the programme is used for the verification of the *Salmonella* reduction target, number of animal under the control program, total number of samples tested, etc.
C.02	sampUnitType	xs:string (5)		G198A holding G199A animal G202A herd	Define the level at which the reported indicator is reported
	animalId	xs:string (250)			Report animal ID where indicator is reported at animal level to allow all indicators for an animal to be linked
	herdId	xs:string (250)			Report herd ID for allow indicators for a herd to be linked

	sampHoldingId	*xs:string (250)*			Report holding ID for allow indicators for a holding to be linked
D.01	sampId	xs:string (100)	M		Identification code of the sample taken.
D.03	sampCountry	xs:string (2)	M	COUNTRY	Country where the holding is located
D.04	sampArea	xs:string (5)		NUTS	Area where the holding is located (Nomenclature of territorial units for statistics - NUTS - coding system valid only for EEA and Switzerland).
D.06	sampY	xs:integer (4)	M		Year of sampling. In case the sampling has been performed over a period of time the start date (as year) of sampling should be reported.
D.07	sampM	xs:integer (2)			Month of sampling. In case the sampling has been performed over a period of time the start date (as month) of sampling should be reported.
D.08	sampD	xs:integer (2)			Day of sampling. In case the sampling has been performed over a period of time the start date (as day) of sampling should be reported.
E.01	sampMatType	xs:string (5)	M	S000A animal sample	Type of sample taken (e.g. food, food stimulants, animal, feed, environment; food contact material), identifying the sub-domain of the matrix catalogue to be used.
E.02	sampMatCode	*CompoundType*[21]	M	A0C9L = Dairy cows	Description of the sample taken characteristics using the FoodEx2 catalogue.
	breed	xs:string (250)		Holstein black Holstein red Jersey Fleckvieh Brown Swiss Montbéliarde Scandinavian red	Breed of dairy cows in herd
	prod	xs:string (5)		A0C6Q= Intensive production A0C6Y = Conventional non-intensive production Z0216 = Other production	

	housing	xs:string (250)		Tie-stall system Cubicle system Deep litter system Free range system	Housing system used on the holding
	flooring	xs:string (250)		slatted/concrete/ mattress	Flooring used in housing
	bedding	xs:string (250)		sand / sawdust / straw / compost / deep litter / chalk / none	Bedding used in housing
	pasture	xs:string (1)		YES/NO	Dairy cows have access to pasture
E.03	sampMatText	xs:string (250)			Description of the sample taken characteristics using free text.
F.03	analysisY	xs:integer (4)	M		Year when the analysis was completed.
F.04	analysisM	xs: integer (2)			Month when the analysis was completed.
F.05	analysisD	xs: integer (2)			Day when the analysis was completed.
G.01	anMatCode	*CompoundType*[21]		A02LT = Milk	Encoding of the matrix only required in case of somatic cell count
K.02	paramCode	*CompoundType*[21]	M	RF-XXXX-XXX-XXX = values specified in paramText	Indicate type of numerical value reported
K.03	paramText	xs:string (250)		**Animal level measurements:** **SSC** = Somatic cell count (cells/mL) **DIM** = Days in milk (days) **Yield** = Average milk yield per cow (kg/d) **Parity** = number of offspring **Herd level measurements:** **HerdSize** = number of dairy cows in the herd **DaysPasture** = Number of days on pasture per year (days) **Yield** = Average milk yield per cow (kg/d) **Parity** = number of offspring **Density** = cubicles per animal	Additional information on indicator or herd measurement

				AMR = On farm mortality (Percentage of animals (euthanized and emergency slaughtered) on the farm during the year) **HSCC** = High Somatic Cell Count (Percentage of cows with Somatic cell count > 400000 for three months / No cows tested during three months) **BMSCC** =Bulk Milk Somatic Cell Count (Percentage of bulk tank measures with Somatic cell count > 400000 for three months / No bulk tank measures tested during three months) **HLAME** = Lameness at herd level (Percentage of lame animals in the herd per year)	
L.05	anMethText	xs:string (250)		"Welfare Quality®"	Method of measuring indicators for example type of lameness scoring
L.06	anMethInfo	*CompoundType*[21]			Additional specific information and comments on the analytical method depending on specific requirements of the different data collection domains such as disk concentration and diameter for antimicrobial resistance diffusion method, method sensitivity and method specificity, migration time, migration temperature, etc....
M.01	resId	xs:string (100)	M		Identification code result a row of the data table in the transmitted file. The result identification code must be maintained at organisation level and it will be used in further updated/deletion operation from the senders.

M.03	resUnit	xs:string (5)	UNIT	Unit of measurement for the values reported in Result value
M.10	resVal	xs:double		Numerical value for specific measurement as categorised in paramCode and expressed in the unit specified by the element Result unit.
M.15	resQualValue	xs:string (3)	POSNEG	This field should be completed only if the result value is qualitative e.g. positive/ present or negative/ absent. In this case the element Result value should be left blank.

Original Column names and definitions

Country& database	Name in database	Definition
Italy SCC	Date of record	Date when sample was taken
	Laboratory	Name of the laboratory where the sample was sent to (by default IZSLER Centro di Referenza Nazionale per la Qualità del Latte Bovino (Italian National Reference Centre for Bovine Milk Quality))
	Farm code	Code number for the identification of farms
	Bulk Tank SCC	The number of somatic cells
	BTSCC Geometric mean	The rolling geometric mean calculated for each sampling with reference to the samples of the same farm during the previous 91 days
	Region	The region of the farm
	Province	The province of the farm
	Country	The country of the farm
Italy AMR	Country	Italy
	Region	Italian Region
	AnimalID	Not specified since data are at herd level
	HerdID	Every farm having at least one cow (female bovine having calved at least once)
	HoldingID	ID of the holding
	Year	Year of sample
	Month	Month of sample
	Day	Day of sample
	Breed	'Most prevalent breed' considering all cows present in 2012 on the farm and the days they spent on the farm in 2012
	Production type	Not compulsory; not available in some cases. If available three options: 'transhumance', 'intensive', 'extensive'.
	Housing	not available
	Flooring	not available
	Bedding	not available
	Herd size	Average number of cows (female bovine having calved at least once) in the farm in 2012; number of days spent on the farm in 2012 for each cow is considered
	Access to pasture	Not a compulsory data; not available in some cases. If available three options: 'transhumance', 'intensive', 'extensive'
	Milk Yield	not available (it is in Luigi and Fracnesca DB, for some regions)
	Parity	Sum of overall calvings (until 31/12/2012) in the life of each cow present in the farm in 2012
	DIM	not available
	SSC	not available (it is in Luigi and Fracnesca DB, for some regions)
	Density	not available
	AMR	Ratio between number of cows dead on farm in 2012 and 'Herd_size'
	LAME	not available

Denmark SCC	ID	Line ID
	Cow_ID	Cow ID
	Herd_ID	Herd ID
	Date_milkcontrol	Day month year
	Milkcontrol_number	Number of milk control per farm
	Kg_milk	Kg of milk during the sample day
	SCC_1000	Somatic cell count / 1000
Denmark AMR	Herd_ID	Herd ID
	Country	Country abbreviation
	Postal_code	Postal code in four numbers
	Breed_code	Code for the breed
	Breed_name	Name of the breed
	Number_cows	Average number of cows in 2012
	Number_euthanasie	Number of cows euthanized in 2012
	Number_death	Number of cows that died in 2012
	Sum_dead	Sum of number_euthanasie and number_dead
	Mortality	Ratio between Number_cows and Sum_dead
Belgium	UBN	Unique farm number (Uniek bedrijfsnummer)
	datum MPR	Date the sample was taken
	aantal dieren in controle voor SCC	Number of animals that were tested on SCC
	% hoog celgetal	Percentage of animals with a cell count higher than 400000
	% nieuwe infecties gem celgetal	Percentage of new infections average SCC

GLOSSARY AND ABBREVIATIONS

ABM: Animal based measure.

Adverse effect: A negative response of an animal towards factors/hazards acting upon the animal.

AMR: Annual mortality rate.

Animal based measure (ABM): A response of an animal or an effect on an animal used to assess its welfare. It can be taken directly on the animal or indirectly and includes the use of animal records. It can result from a specific event, e.g. an injury, or be the cumulative outcome of many days, weeks or months, e.g. body condition.

AUC: Area under the curve.

Cut-off: A cut-off value when an animal based measure or WAE is considered to be 'positive' i.e. indicative of a defined welfare outcome.

Data element: All variables present in a database. This can range from welfare indicators to year of sampling.

DCF: Data Collection Framework. This is a web interface developed by EFSA accessible by most common web browsers through which data providers can submit their files. The system provides automatic feedback on errors in structure and content, and confirmation of successful submissions.

Entity class: An overall category covering a number of distinct factors. For instance, the entity class "management" covers "feeding routine", "biosecurity standards" etc.

Intra-observer reliability (IOR): Level of agreement between repeated 'measurements' of the animal based measure on the same 'sample' by the same assessor, on different occasions. Also termed intra-observer agreement.

Inter-observer agreement (IOA): Degree of agreement between measurements or observations conducted in replicates by different people.

Measure: A form of evaluation rather than an intervention intended to deal with a problem.

Measurement: The result of the above evaluation, e.g. size and depth of wound, percentage of lame animals.

MS: EU member states.

Population: The totality of individuals of the same kind that share or have in common certain attributes.

RCD: Routinely Collected Data. These are data that are not collected in light of a certain project but routinely, for instance by authorities checking milk quality.

Reliability: A general term referring to the ability of the animal based measures to be applied under various conditions, by different personnel while still providing similar (correct) results.

Repeatability: See Intra-observer reliability (IOR)

Reproducibility: See Inter-observer agreement (IOA)

Risk factors: ‘Any factor associated with the increase of appearance or development of a phenomenon’ (Toma et al., 1996).

Robustness: How the measure is affected by changes in the environment, who is taking the measure and when it is taken. It encompasses concepts such as intra-observer reliability (repeatability) and inter-observer agreement (reproducibility), which are the agreement between repeated measurements of the welfare consequence on the same sample by the same assessor (intra-observer) or a different assessor (inter-observer) respectively.

ROC: Receiver operating characteristic.

SCC: Somatic cell count.

SD: Standard deviation.

SSD: Standard Sample description. This is an EFSA-document that provides specifications (standardised data elements) aimed at harmonising the collection of analytical data.

Sensitivity (Se): The minimum level of welfare outcome (changes) that will be detected by the animal based measure. The proportion of the animals/herds with a given welfare problem that will be correctly identified by the ABM as having welfare problems.

Specificity (Sp): The extent to which an animal based measure is specific for one welfare outcome, or relates (respond to) several outcomes. The proportion of the animals/herds not having a given welfare problem that will be correctly identified by the ABM as not having welfare problems.

Target condition: The disease/condition/behaviour that a test (e.g. ABM) originally was designed to measure.

Threshold: See ‘Cut-off’

TMR: Total mixed ration

Validity: The fitness of an animal based measure that has been properly developed, optimised and standardised for an intended purpose. Validation includes estimates of the analytical and diagnostic performance characteristics of the measure/indicator (i.e. sensitivity and specificity).