The utility of sensor technology to support reproductive management on dairy farms.
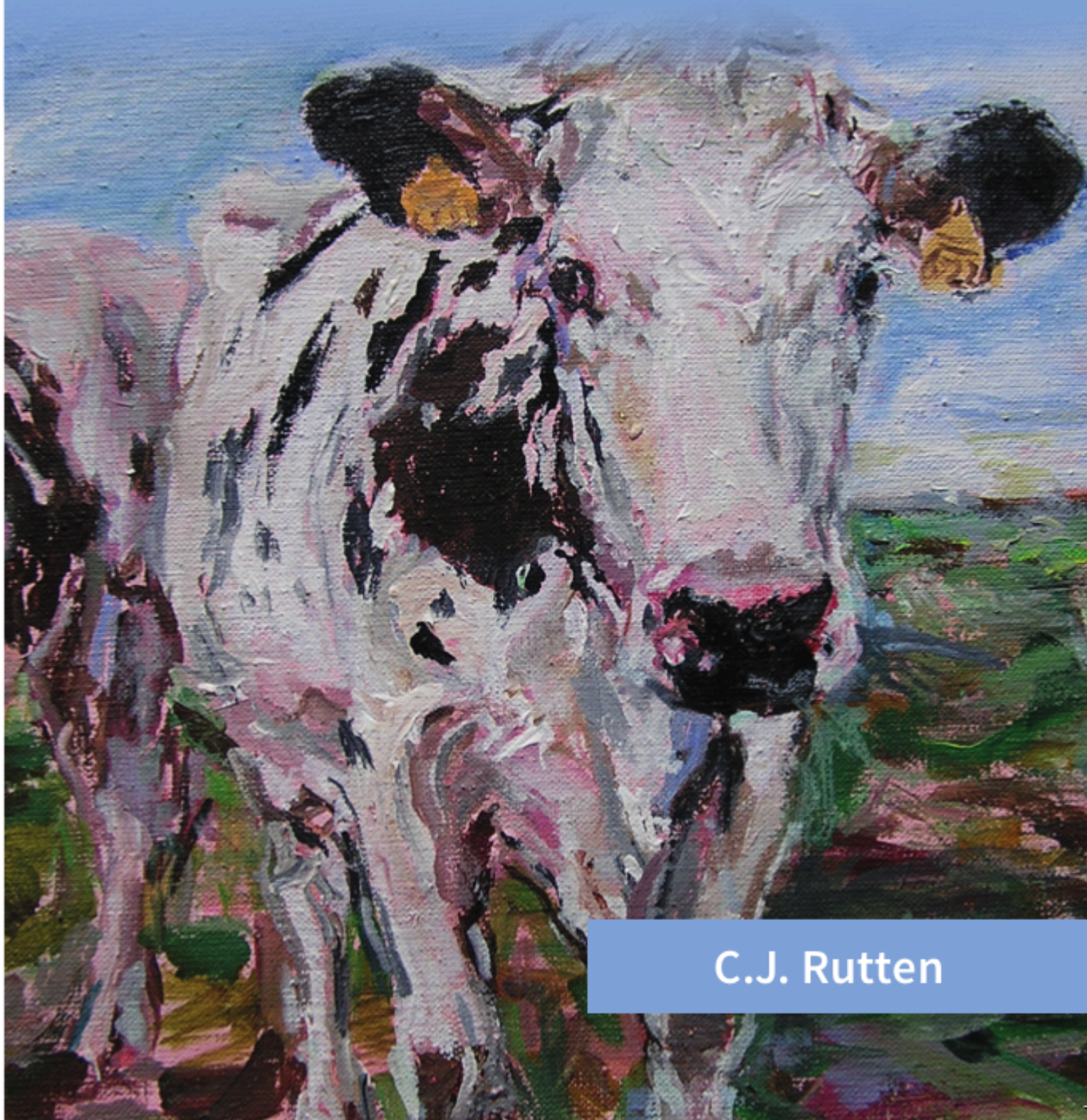
C.J. Rutten

The utility of sensor technology to support reproductive management on dairy farms.

C.J. Rutten

Thesis committee

Promotors

Prof. Dr H. Hogeveen
Personal chair at the Business Economics group
Wageningen University & Research

Prof. Dr M. Nielen
Professor of Evidence-based Veterinary Medicine
Utrecht University

Co-promotor
Dr W. Steeneveld
Assistant professor of Veterinary Economics,
Utrecht University

Other members
Prof. Dr B. Tekinerdogan, Wageningen University & Research
Dr P.L.A.M. Vos, Utrecht University
Prof. Dr L. Munksgaard, Aarhus University, Denmark
Prof. Dr N.H.G. Baken, Delft University of Technology

The utility of sensor technology to support reproductive management on dairy farms.

C.J. Rutten

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 14 June 2017
at 1:30 p.m. in the Aula.

Contents

Chapter 1

General introduction

General trends in dairy farming

Dairy farming in western countries is a dynamic industry, characterized by increases in automation, herd size, and milk yields per cow over the past years *(Barkema, et al., 2015)*. For example, the number of dairy farms in the Netherlands decreased by 38% (from over 29,000 in 2000 to around 18,000 in 2015), whereas the number of animals per farm increased by 74% (from 51 cows/farm in the year 2000 to 89 cows/farm in 2015) *(CBS, 2016)*. A further trend is the growing importance of efficient dairy production for dairy farm profitability. Achieving efficient production requires efficient use of resources (e.g., feed) and healthy animals. Technology that can replace or reduce the needed amount of labor, or improve production or animal health, is therefore important within the current trends in dairy farming. Technological progress, especially in ICT, has enabled automated applications in dairy farming.

Sensor development

The introduction of the automated milking system (**AMS**) was an important milestone in the uptake of automated applications in dairy farming. With the adoption of the AMS, precision livestock farming (**PLF**) became reality on commercial dairy farms. PLF is the optimization of production and management processes by the use of automated applications (sensors and robots) *(Berckmans, 2008)*. The use of an AMS requires the application of automated mastitis detection *(Hogeveen, et al., 2010)*, which has been implemented in AMS using electrical conductivity sensors *(Nielen, et al., 1992)*. The first studies on the potential of electrical conductivity as an automatically measured indicator of abnormal milk were published in the 1960s *(Greatrix, et al., 1968, Little, et al., 1968)*. The first sensors were introduced in the 1980s but remained unsuccessful until the uptake of AMS *(Hogeveen, et al., 2010)*. Although AMS farmers are required to invest in sensors for automatic mastitis detection, these farmers also invest in other types of sensors more than farmers with a conventional milking system *(Steeneveld and Hogeveen, 2015)*. Many farmers indicated that they have sensors on their farm because these devices came with an AMS *(Steeneveld and Hogeveen, 2015)*. The AMS can therefore be seen as an important driving force for the uptake of sensor technology by dairy farmers.

Sensors that are not part of an AMS, such as pH boluses, automated body condition scoring (BCS), and urea sensors, are used less frequently. Activity meters for estrus detection form a notable exception *(Borchers and Bewley, 2015, Steeneveld and Hogeveen, 2015)*. Activity meters are used more frequently because they are efficient tools for estrus detection and estrus detection is economically important *(Firk, et al., 2002)*. Many sensors have been developed in recent years. Several reviews of sensor research are available for estrus and mastitis detection *(Firk, et al., 2002, Hogeveen, et al., 2010, Norberg, 2005)*. A review that discusses the whole field of sensor research for animal health management, however, has

not yet been published. The lack of a framework to compare current sensor systems hampers an overall overview of the state of the art in sensor utility.

Economics of sensors

PLF research has always claimed that sensors and robots have the potential to improve dairy farm profitability, animal health and welfare, environmental impact, risk management, and reduce (manual) labor requirements *(Berckmans, 2008)*. Sensors can help farmers to manage their herd more precisely with regard to production, cow health, and environment *(Ipema, et al., 2011, Wathes, et al., 2005)*. This precision is possible because sensors and robots are capable of routinely monitoring behavioral and physiological parameters of dairy cows 24 hours a day, 365 days a year, whereas a farmer or employee could never observe these parameters in such detail. Furthermore, it has long been recognized that sensors could have economic benefits for dairy farmers because they can help improve the technical results of the farm *(Hogeveen, et al., 1994, van Asseldonk, et al., 1999a, van Asseldonk, et al., 1999b)*. Farmers have indicated that they invest in estrus detection sensors to improve estrus detection and farm profitability, and to obtain more insight in the fertility level of their herd *(Steeneveld and Hogeveen, 2015)*. Farmers thus expect better economic results and improvement in reproductive performance of the herd from the use of estrus detection sensors. Studies have theoretically quantified the potential benefits of automated BCS *(Bewley, et al., 2010a)* and estrus detection *(van Asseldonk, et al., 1999b)*. However, field studies that quantify the improvement in farm production and farm management and the subsequent economic benefits are still lacking. The utility of sensors therefore remains undetermined.

Sensors in farm management

Sensor systems have long been recognized to potentially improve decision making on dairy farms, but few sensors have been adopted by farmers *(Hogeveen, et al., 1994)*. PLF is claimed to provide unprecedented accuracy of information for managing livestock *(Wathes, et al., 2005)*. Although this potential is widely accepted, the added value is not emerging in practice, and PLF is therefore not currently living up to expectations *(Lehr, 2011)*. Farmers indicate that an important reason for not investing in new technology is that the technology produces too much output, which they do not know what to do with *(Russell and Bewley, 2013)*. Many decisions are complex on dairy farms because many factors are involved. For instance, it is difficult to make an optimal treatment decision based on all available information, treatment options, and cure-influencing cow factors *(Steeneveld, et al., 2011)*. Farmers deal with considerable uncertainty in their decision making and usually base their expectations on their herdsmanship *(Groenendaal, et al., 2004)*. To this environment, sensors add more (uncertain) data without much interpretation. The use of sensor output

therefore creates extra workload, as the farmer needs to interpret the sensor data with regard to other available information. Sensor data needs to be usable for management purposes if the potential of sensors is to be achieved.

Decision support systems have the potential to process large volumes of data to provide information that supports management decisions *(Jalvingh, 1992)*. If sensor data can be automatically integrated with other data and information, interpretation will be easier and output will be more informative. Effective support of farm management can be achieved if decision support models estimate the technical and economic consequences of various strategies, considering uncertainty and the consequences of suboptimal decisions *(Jalvingh, 1992)*. Information from a decision support system that integrates and interprets sensor data and other farm data could therefore be very useful for farmers to support their operational management decisions.

An important part of the operational management on dairy farms is the management of health and reproduction. Studies have been conducted on the economic benefits of improved reproductive performance *(Galvão, et al., 2013, Giordano, et al., 2011, Inchaisri, et al., 2010b, Shahinfar, et al., 2015)*. Firk et al. *(2002)* discussed the use of sensors for estrus detection, with a focus on detection performance and algorithms used. Sensors for estrus detection are applied on commercial dairy farms worldwide *(Borchers and Bewley, 2015, Galon, 2010, Steeneveld and Hogeveen, 2015)*. Given the importance of reproduction management and the wide adoption of sensors for estrus detection, potential exists to study the utility of sensors in reproductive management and develop decision support models that can provide farmers with information to support their reproductive management.

Reproductive management consists of decision making regarding aspects of calving, breeding and culling. To start a lactation a cow needs to calve. The farmer needs to estimate when a cow will approximately calve and thereby determine whether to move her to the close-up group and subsequently to the calving pen. Furthermore, the farmer needs to check a cow for the start of calving to determine whether *dystocia* occurs and decide whether assistance is needed. After lactation has started, the farmer needs to decide whether to breed the cow to start another lactation or to cull her at the end of the current lactation. If the decision is to breed the cow, the farmer needs to know whether postpartum anestrus is occurring and decide whether treatment is necessary. The next step for the farmer is to determine at which day in milk (DIM) to start inseminating the cow. Subsequently, the farmer needs to determine whether insemination was successful and decide when to dry off the cow.

Towards operational reproductive management

In summary, the potential of sensor technology to improve the management of individual cows has been suggested since the first developments in sensor technology for dairy

farming. Improved management of individual cows is expected to improve animal health and welfare, reduce manual labor, reduce costs, improve farm efficiency, and reduce environmental impact *(Berckmans, 2008, Ipema, et al., 2011, Steeneveld and Hogeveen, 2015)*. However, it has proven difficult to develop sensor systems that work in practice, produce meaningful output that meets the needs of farmers, and that have a clear benefit for farmers. As a consequence, adoption has generally been poor and mainly driven by the AMS, with a clear exception for estrus detection. Although the potential is widely acknowledged, only a few applications have been commercialized successfully *(Banhazi, et al., 2012)*.

The uptake of sensor systems would benefit from the development of sensor systems and decision support systems that provide farmers with information they can use in reproductive management. Furthermore, there is a lack of clear evidence on the economic benefits of using sensor systems in general, and therefore there remains a need to quantify these economic benefits. Overall, to ensure sensors live up to their full potential, there is a need for research to shift from technical development towards practical applications and integration with operational farm management. Dairy farmers operate their businesses within the broader context of modern society. This means that societal acceptance of technological solutions is also important, in addition to technical progress and economic aspects of adoption. Therefore, societal acceptance is relevant to the discussion on the utility of sensor systems for dairy farms.

Objectives of this thesis

The overall goal of this thesis is to study the utility of sensor technology to support reproductive management on dairy farms. Aspects of utility are evaluated in the five objectives. All objectives are studied and demonstrated with data and examples from Dutch dairy farms.

The first objective is to develop a theoretical framework that describes the development of sensor systems (chapter 2). This framework is demonstrated in a structured literature review.

The second objective is to explore whether sensors used for estrus detection can also be applied to predict the start of calving (chapter 3).

The third objective is to estimate the economic value of a predictive system for the start of calving (chapter 4). A normative simulation is used to quantify the economic benefit of this sensor system.

The fourth objective is to estimate the economic value of a sensor system for estrus detection (chapter 5). A normative simulation is used to quantify the economic benefits of this sensor system.

The fifth objective is to take the development of a sensor system for estrus detection further by developing and validating a prognostic model for insemination success (chapter 6).

The overall goal is discussed in the general discussion (chapter 7).

Chapter 2

Invited review: Sensors to support health management on dairy farms.

C.J. Rutten

A.G.J. Velthuis

W. Steeneveld

H. Hogeveen

ABSTRACT

Since the 1980s, efforts have been made to develop sensors that measure a parameter from an individual cow. The development started with individual cow recognition and was followed by sensors that measure the electrical conductivity of milk and pedometers that measure activity. The aim of this review is to provide a structured overview of the published sensor systems for dairy health management.

The development of sensor systems can be described by the following four levels: I. techniques that measure something about the cow (e.g., activity); II. interpretations that summarize changes in the sensor data (e.g., increase in activity) to produce information about the cow's status (e.g., estrus); III. integration of information where sensor information is supplemented with other information (e.g., economic information) to produce a advice (e.g., whether to inseminate a cow or not); and IV. the farmer makes a decision or the sensor system takes the decision autonomously (e.g., the inseminator is called).

This review has structured a total of 126 publications describing 139 sensor systems and compared them based on the four levels. The publications were published in the ISI database from January 2002 until June 2012 or in the proceedings of three conferences on precision (dairy) farming in 2009, 2010, and 2011.

Most studies concerned the detection of mastitis (25 percent), fertility (33 percent), and locomotion problems (30 percent), with fewer studies (16 percent) related to the detection of metabolic problems. Many studies presented sensor systems at levels I and II, but none did so at levels III and IV. Most of the work for mastitis (92 percent) and fertility (75 percent) is done at level II. For locomotion (53 percent) and metabolism (69 percent), more than half of the work is done at level I.

The performance of sensor systems varies based on the choice of gold standards, algorithms, and test sizes (number of farms and cows). Studies on sensor systems for mastitis and estrus have shown that sensor systems are brought to a higher level, however there is still a need to improve detection performance. Studies on sensor systems for locomotion problems have shown that the search continues for the most appropriate indicators, sensor techniques, and gold standards. Studies on metabolic problems show that it is still unclear which indicator reflects best the metabolic problems that should be detected. No systems with integrated decision support models have been found.

INTRODUCTION

There is a trend in dairy farming towards the automation of processes in order to reduce (physical) labor and labor costs *(de Koning, 2010, Svennersten-Sjaunja and Pettersson, 2008)*. This development is partly driven by the economic reality of increasing labor costs relative to capital costs. Automated systems enable dairy farmers to manage larger herds with lower labor requirements *(de Koning, 2010)*, which means that the application of automated systems fits with the trend of increasing herd sizes. Another important trend in Western Europe is the use of automated milking systems (**AMS**). The first studies on the feasibility of an AMS were done in the mid-1980s, such as the initial preliminary study by *(Rossing, et al., 1985)*. The first commercial farms introduced an AMS in the Netherlands in 1992 and by 2009 the number of farms using an AMS had increased to more than 8,000 worldwide (90 percent of which were located in Western Europe) *(de Koning, 2010)*. When using an AMS, there is no milker present who could visually judge the cows' foremilk for the presence of clots to detect mastitis. Therefore, farmers must rely on alternative methods to detect mastitis.

Since the 1980s, work has been done on devices that measure a health indicator in, up, on, or from an individual cow *(Hogeveen, et al., 2010)*. Examples of sensors include milk electrical conductivity, milk color sensors *(Espada and Vijverberg, 2002)*, acceleration sensors (attached to the cows leg) *(Chapinal, et al., 2011, Pastell, et al., 2009, Saint-Dizier and Chastant-Maillard, 2011)*, and pH sensors *(AlZahal, et al., 2007b)*. A sensor system consists of the device itself, plus the software that processes the data to produce information and/or advice.

In this review four levels are defined that describe the degree to which a sensor system informs the farmer. The sensor itself is only the first step in a sensor system. The second step is to use the sensor data in an algorithm that provides information about the health of the individual cows. In this step, it is possible to combine sensor data with non-sensor data about cow history *(Steeneveld, et al., 2010)*. The algorithm produces information about the cow's health by determining changes in the sensor data. Examples include electrical conductivity and color to measure changes in the milk in order to help the farmer detect clinical mastitis *(Hovinen and Pyörälä, 2011)*, acceleration to measure differences in walking behavior to detect lameness *(Chapinal, et al., 2011, Pastell, et al., 2009)* or acceleration to measure differences in activity to detect estrus *(Saint-Dizier and Chastant-Maillard, 2011)*, and rumen pH to measure acidity of rumen fluid to detect metabolic problems *(AlZahal, et al., 2007b)*. The third step uses this information in a decision support model that uses economic information and perhaps other information from the farmer or an advisor. The decision support model produces advice about how to act upon the detected events that have happened with the cow. One example is the breeding management tool of *(Giordano, et al., 2011)*, which advises farmers about what breeding management strategy will

maximize the net present value per cow. The fourth and final step is the decision regarding the change in the health status of the cow, as detected by the sensor (for example, mastitis detected by increased milk electrical conductivity or estrus detected by increased cow activity). In most cases, the farmer makes the decision; in some specific and advanced sensor systems, the system makes an autonomous decision. For example, the system of *(Andre, et al., 2011)* automatically adjusts the amount of concentrates fed to the cow based on her milk production and milking interval. In general sensor systems could support farmer decisions regarding treatment, insemination, and culling.

Although some reviews on sensor systems have been published, their scope has been limited to a specific application or type of sensors *(Firk, et al., 2002, Gordon, 2011, Hogeveen, et al., 2010, Norberg, 2005, Saint-Dizier and Chastant-Maillard, 2011)*. As more sensors become available and are tested more extensively, there is a need for a clear overview of what sensors have been tested, how advanced the systems are, and the quality of the produced data, information, and devices. Such a structured overview is currently lacking within the scientific literature. As the idea of sensor research is to provide farmers with tools to improve their cow health management, the central question in research regarding sensors should concern what value (meaning, the economic value, but also the usefulness for risk management and making labour easier) the sensor system adds for the farmer's decision making.

The objective of this review is to provide a structured literature review of the sensors, sensor data, data algorithms that provide information, and the corresponding decision support systems to be used in precision dairy farming. This paper describes four key points: (1) what sensor systems are developed for the detection of locomotion problems, mastitis, estrus, and metabolic problems; (2) the performance of these sensor systems; (3) the tools for decision support coupled to these sensors; and (4) the economic impact of using these sensor systems for farm management.

FRAMEWORK

This study considers a sensor to be a device that measures a physiological or behavioral parameter (related to the health or estrus) of an individual cow and enables automated, on-farm detection of changes in this condition that is related to a health event (such as disease) and requires action on the part of the farmer (such as treatment).

There are two categories of sensors: attached and non-attached. Attached sensors may be on-cow sensors that are fitted on the outside of the cow's body, or in-cow sensors that are inside the body (for example, rumen bolus or implant). Non-attached sensors are off-cow sensors that cows pass by, over, or through for measurement. Two specific forms of non-

attached sensors are in-line and on-line sensors. In-line sensors take measurements in a continuous flow of a product from the cow. The only available option for in-line measurement is in the milk line. On-line sensors automatically take a sample (milk, for example) that is analyzed by the sensor.

This review has used the scheme shown in Figure 1 to categorize sensor systems. The scheme shows a framework that describes the steps from a sensor to a decision. Sensors are categorized in the levels of this scheme according to their description in the literature. Sensors are only described if they reach at least level I, known as "technique," which means that they measure an aspect of the cows' condition or status. The two categories identified within this level are solely measuring a parameter and an assumed relation. In some sensors the produced data is processed by a data algorithm (for example a pedometer records clicks of a mercury switch, the data algorithm produces a step count per time unit from these clicks). The next step (level II) is called "data interpretation" and measures changes in the sensor data to produce information about the cows' status (e.g., estrus). The two categories identified within this level are a statistically tested relation and a validated algorithm. From a statistically tested relation, it is possible to build a predictive model (detection algorithm) that classifies the cows' status (for example, in estrus or not in estrus). For validation, a dataset (not the one used to build the detection algorithm) is used to assess the performance by comparing the classification of the algorithm with the gold standard. A further feature can be updating or resetting the detection algorithm with gold standard measurements during operation in practice, this would mean the algorithm adapts to an individual farm or changing circumstances. Level III integrates the sensor information with other information (such as economic information), to produce advice for the farmer. Furthermore, information of individual cows can be aggregated by a monitoring algorithm at the herd level. The output of this algorithm can be seen as either general information on the herds health for the farmer or additional data input for the detection algorithm. The decision is eventually made either by the farmer or autonomously by the sensor system (level IV, known as "decision making").
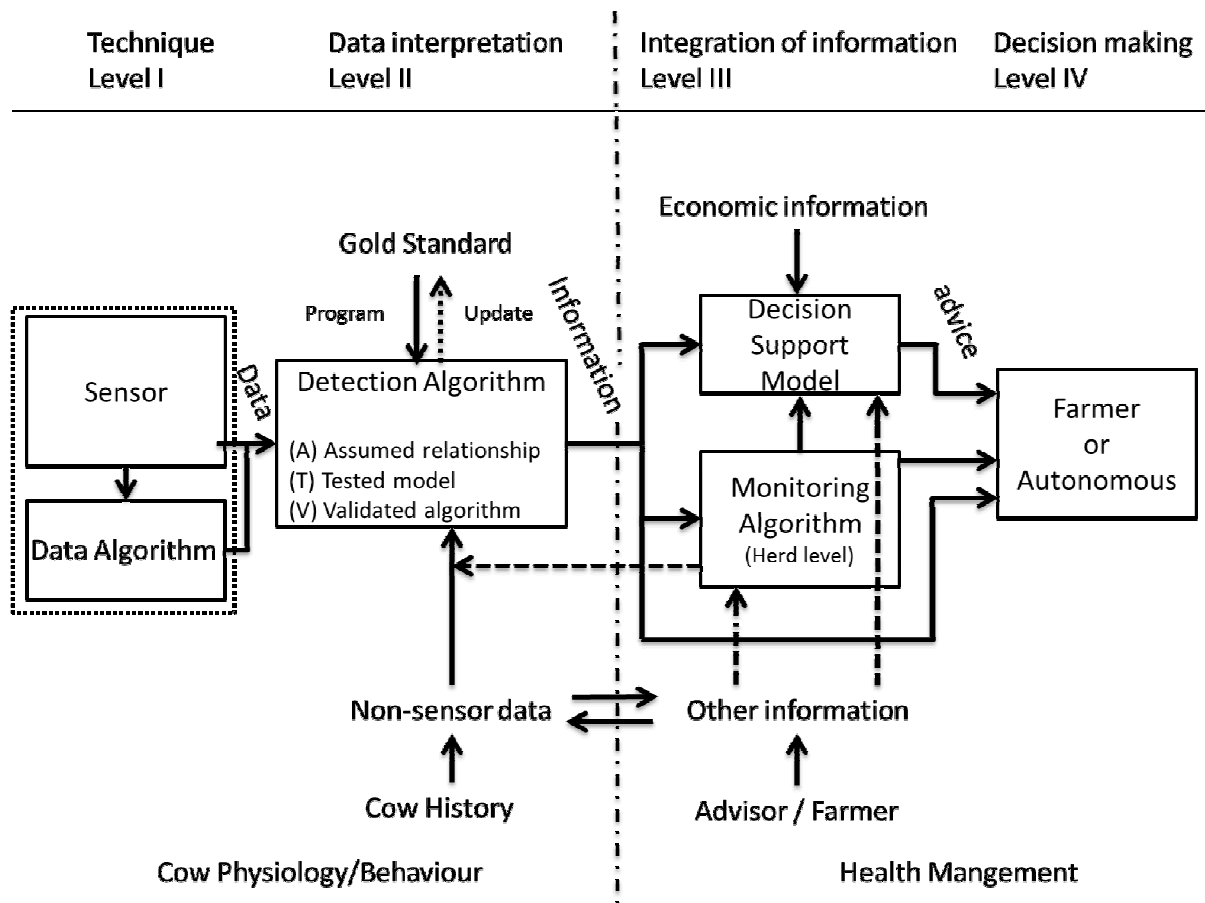
Figure 1: Framework of the use of sensor information in dairy farm management.

Each sensor system is described using thirteen categories referring to the used technique, collected data, used algorithms and performance. The first three categories provide a brief summary of the technical aspects of the device: "type of sensor (1), "sensor location" (with respect to the cow) (2), and "type of measurements" (3). Alerts given by the sensor are compared to a "gold standard" (4), which describes the occurrence of an event in reality *(Hogeveen, et al., 2010)*. The relation between the gold standard, sensor data, and possible data additional to the sensors' data that does not originate from the sensor under study ("non-sensor data" (5)) is described by an "algorithm" (6). A level III sensor system will have a "decision support model" (7). In the "economic information" (8) category, the costs and benefits associated with possible decisions are described, while "other information" (9) describes additional information, such as treatment success rate. For the sensor algorithm, the percentage of detected true events ("sensitivity" (10)) and the percentage of truly healthy animals classified as healthy ("specificity" (11)) are presented, along with the "test size" (12) upon which these are based. Note that only sensitivity and specificity are summarized here, other indicators of detection performance that were sometimes described in literature like the false alert rate are not included here. The final category – "system output" (13) – is what the sensor system presents to the farmer.

Having gained a clear overview of the available sensor systems, the data and information produced by these sensor systems was compared. The quality of the data and information (that is, whether the data or information was fit for use by farmers) was assessed. The criteria for this assessment were whether the sensor system detects a clearly defined change in the cows' health (such as disease or estrus) and presents this change clearly to the farmer (such as in the form of an alert rather than as a graph of sensor data). Also, detection sensitivity and specificity were discussed in respect to the used gold standard and the test scale. Furthermore, the added value of the sensor system for the farmer's decision making was discussed. This comparison led to a discussion of what further research would be needed to improve performance of available sensor systems, develop new sensors and make sensor systems that can be used in practice on farms and are of more practical value on-farm.

LITERATURE SELECTION

The relevant literature was searched based on keywords including sensors, dairy farming, and automated detection, in combination with words such as mastitis, locomotion, lameness, metabolic, estrus, and fertility. Literature was also identified by a forward search, using the citations and a backward search using the references of the papers found through the keyword search. Journals from the ISI database (Web of Science, Thomson Reuters, New York, USA) were used for the period from January 2002 until June 2012, and the proceedings of relevant scientific conferences held between 2007 and 2012 were searched. The conferences included the First North American Conference on Precision Dairy Farming (Toronto, 2010) and the European Conference on Precision Livestock Farming (Prague, 2011 and Wageningen, 2009).

SENSORS

All publications describing sensor systems are summarized in tables in the appendices of the review. These tables present general information about the sensor systems, more detailed information about the sensor systems can be found in the online appendices of the review.

Mastitis

Level I: 'sensor technique'. For automated detection of mastitis, 31 publications were found with 37 described sensor systems (some publications described multiple sensors, algorithms, or diseases). Four (11 percent) of these publications were proceedings papers. Electrical Conductivity (**EC**) was the main sensor system studied (in 15 studies, 48 percent), followed by a combination of EC sensors and milk color sensors (seven studies, 23 percent). Some other studies used a biosensor to detect certain enzymes, including haptoglobine, L-lactate dehydrogenase or NAGase (five studies, 16 percent), an SCC sensor (two studies, 6 percent), or a reticular bolus that measured temperature (one study, 3 percent). The EC and milk color sensors were in-line sensors, meaning that they measure in a continuous milk flow. The biosensors and SCC sensors took measurements from automatically collected milk samples, meaning that they were on-line sensors. The bolus was inserted in the cow's reticulum, which made it an in-cow sensor.

Level II: 'Data interpretation'. Figure 2 shows that 34 (92 percent) published sensor systems for automated detection of mastitis included data interpretation, which means that the (statistical) relation between gold standard (e.g., the California Mastitis Test (**CMT**)) and sensor measurements (e.g., milk EC) was studied. For 27 (73 percent) studies the algorithm was validated (that is, they used a dataset to calculate predicted mastitis classification per cow and compare these with the gold standard). For 15 (41 percent) sensor systems, the sensitivity and specificity of the detection algorithm were determined and reported; this is discussed further below.

Level III 'integration of information'. Figure 2 shows that none of the sensor systems integrated the sensor information with other information.

Level IV 'decision making'. Seventeen (46 percent) sensor systems provided the farmer with a mastitis alert, while 12 (32 percent) also provided a probability added to this mastitis alert to describe the certainty, and four (11 percent) provided a degree or classification of mastitis. Eight (22 percent) sensor systems provided the farmer with raw sensor data (such as milk NAGase concentration) and/or were unclear in describing what they provided to the farmer.
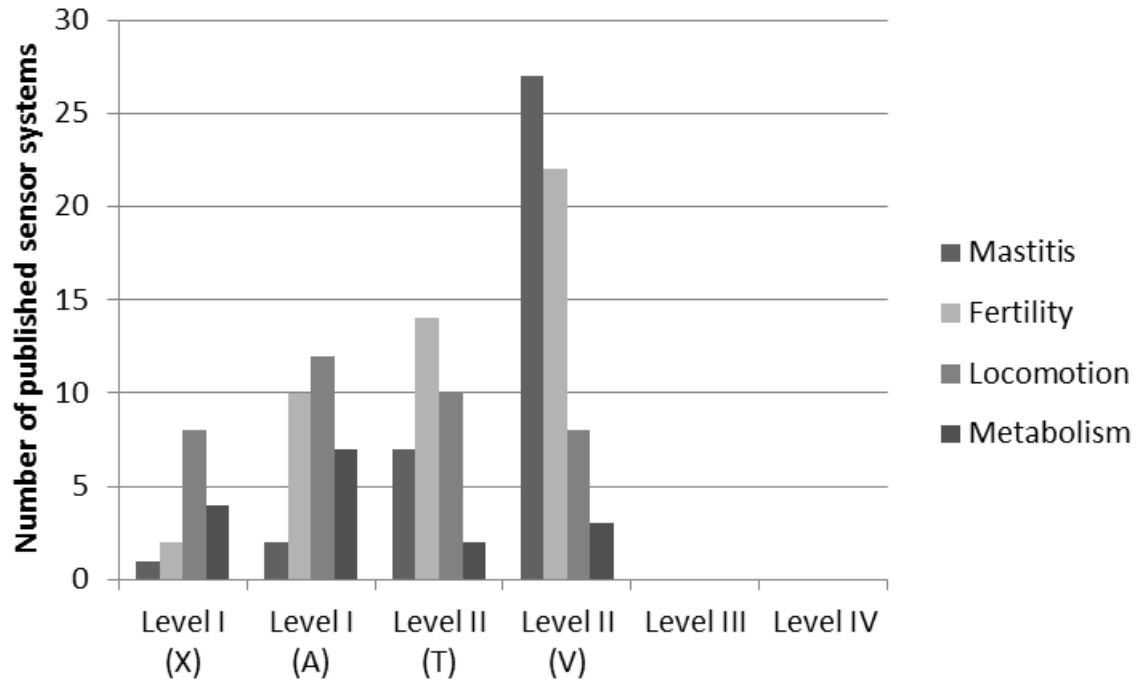
Figure 2: Number of studied sensor systems for mastitis, fertility, locomotion and metabolism per development level, Level I technique, Level II data interpretation, Level III integration of information and Level IV decision making. The levels I and II are subdivided into: solely measuring a parameter (Level I, X), an assumed relation between gold standard and sensor data (Level I, A), a statistically tested relation between gold standard and sensor data (Level II, T) or a validated algorithm for detection (Level II, V).

System quality. Detection performance, which is described by sensitivity and specificity, varies considerably. The reported sensitivities ranged from 55 percent to 89 percent, while reported specificities ranged from 56 percent to 99 percent. For the sensor systems studied, a trade-off exists between sensitivity and specificity as high sensitivity (> 80%) is combined with low specificity and vice versa. None of the studies reported a combination of high sensitivity and high specificity. Moreover, none of the studies met the ISO/FDIS 20966 limit of 80 percent sensitivity with 99 percent specificity. Although one sensor system reported a sensitivity of 100 percent and a specificity of 99 percent for a combination of SCC and EC measurements, these values were based on nine positive gold standard cases. Some variation in test scale was reported: 10 publications reported that their sensor system was tested on more than one farm, where the others tested their sensor system on only one farm.

EC is the most studied sensor technique for mastitis detection and in some cases was combined with milk color sensors. Algorithms have been built for most of these sensors, many of which have been validated. Although the biosensors and SCC sensor have been tested, it is unclear whether these sensors show a better performance than EC sensors. Infrared cameras have been tested, but only assumed relations with mastitis are available.

Therefore, it is unclear how these could perform as a mastitis sensor. Even if systems do not meet the ISO limit, the value that an automated mastitis detection system provides to a farmer is obvious, as the alternative is having no automated detection at all. Especially for those farmers using an AMS or managing a large herd the need for a sensor system is present. However, the perfect mastitis alert does not seem to have been found.

Due to the large variation in reported performance, gold standards, test scales, and algorithms used, it is difficult to compare the performance of various sensor types. For example, sensitivity in the range of 83 to 92 percent with specificity between 75 and 94 percent were reported for a fuzzy logic algorithm with SCC as a gold standard *(Cavero, et al., 2006)*. Other studies using a fuzzy logic algorithm *(Kamphuis, et al., 2008b, Liberati and Zappavigna, 2009)*, do not report a comparable performance. When visual judgment of milk is used as a gold standard, naive Bayesian networks (70 percent sensitivity and 97.8 percent specificity) *(Steeneveld, et al., 2010)* and artificial neural networks (91 percent sensitivity and 87 percent specificity) *(Sun, et al., 2010)* show the best detection performance. The detection performances under two different gold standards for different algorithms are hardly comparable. Of the used gold standards, visual judgment of milk and CMT had the advantage that they could be performed easily and on-farm, while SCC requires lab analysis. It is difficult to say which of these gold standards is the most appropriate; on one hand, SCC might be more accurate, while on the other hand, visual judgment and CMT might be done more frequently. For EC, in combination with milk color sensors, good performance has been reported (84.6 percent sensitivity and 99.4 percent specificity) *(Song and van der Tol, 2010)*, with treated CM cases as a gold standard. Because it is not clearly known how many of the occurred cases of CM have been treated, this gold standard's reliability is questionable at best. As a result it is only sure that treated cases were CM cases, how many untreated cases of CM are detected or remain undetected by the sensor system remains unknown. Therefore, the reported performance of such a sensor system should be treated with caution. The Herd Navigator® (DeLaval, Tumba, Sweden), which automatically takes and analyzes milk samples, seems to perform good (80–82 percent sensitivity and 98 percent specificity) *(Mazeris, 2010, Vreeburg, 2010)*, although this is only based on two non-peer-reviewed studies published in conference proceedings and the used gold standard is unclear. When EC is compared with EC in combination with milk color sensors, sensitivity and specificity seem to be lower for the combination of the two sensor types. Accordingly the EC sensors are tested on a smaller scale (number of cows and/or number of farms) than the combination of EC and milk color sensors. An explanation for this observation could be that variation between farms (and between cows) influences the detection performance of the sensor systems. However, it cannot be concluded whether this difference should be attributed to the sensor system or to the difference in test scale.

Fertility

Level I: 'sensor technique'. For automated detection of estrus, 41 publications were found with 48 described sensor systems, six (15 percent) of these publications were proceedings papers. Most of these studied the activity of the cow (25 studies, 61 percent). Other publications studied the progesterone level in milk (six studies, 15 percent), (mounting) behavior (six studies, 15 percent), vocalization (one study, 2 percent), and body temperature (two studies, 5 percent). The sensors used to measure activity were pedometers, activity meters (sometimes also called activometers), and 3D-accelerometers. These sensors were all attached to one of the cow's body (pedometers and 3D accelerometers usually to the cows left hind leg and activity meters to a neck collar) and are therefore classified as on-cow sensors. The sensors used to measure progesterone were biosensors and immunostrips. As progesterone was determined in automatically collected milk samples, these sensors are classified as on-line sensors. The sensors for mounting behavior were the HeatWatch sensor and a video camera. The HeatWatch sensor is a device that measures the pressure caused when another cow mounts the cow with the HeatWatch sensor. This sensor was attached to the cow's back, which classified it as an on-cow sensor. The video camera is classified as an off-cow sensor. The sensor for vocalization was a microphone attached to the cow's neck (that is, an on-cow sensor). The sensors used for body temperature were a temperature transducer implanted in the cow's body and a bolus inserted in the cow's reticulum (that is, an in-cow sensor).

Level II: 'data interpretation'. Figure 2 shows that 36 (75 percent) sensor systems included data interpretation, which means that algorithms were built in these publications, 22 (46 percent) validated their algorithm. The sensitivity and specificity were reported for 14 (29 percent) of the published algorithms.

Level III: 'integration of information'. Figure 2 shows that none of the sensor systems integrated the sensor information with other information. However, a simulation study that focuses on integration of information is available and could be related to the Herd Navigator®; this will be covered in more detail in the discussion section.

Level IV: 'decision making'. With regard to fertility, 29 (60 percent) sensor systems provided the farmer with an estrus alert, seven (15 percent) sensor systems also provided a probability with this estrus alert to describe the certainty, and three (6 percent) reported the reproduction status (for example, postpartum anestrus, estrus cycling, potentially pregnant) to the farmer. Seven sensor systems provided the farmer with raw sensor data (for example, the milk progesterone level) or were unclear in describing what they provided to the farmer.

System quality. Most sensor systems require the farmer to rely on his herdsmanship. Sixteen (35 percent) of the published sensor systems reported sensitivity above 80 percent, eight (17 percent) reported specificity above 98 percent, and six (13 percent) reported specificity somewhere between 90 and 98 percent. Two studies (4 percent) reported

sensitivity above 80 percent in combination with specificity of 99–100 percent; however, these studies used confirmed pregnancy as the gold standard. The largest test scale was three farms reported in five studies (10 percent), while six studies (13 percent) reported a test scale of two farms, and 25 studies (52 percent) reported a scale of one farm. None of the studies reported tests done on a few cows.

The pedometer was the most studied sensor system used to detect estrus, while 3D-accelerometers have been studied. For fertility, performance varies largely in terms of sensitivity and specificity, and varies in terms of the algorithm used and the gold standard. For the reported detection performance there are no recommendations (like the ISO limit for mastitis) available for estrus detection. The "good performances" mentioned in this paper should not be seen as target values, because determination of strict target values asks for in-depth discussion and experiments. Good performance has been reported for both pedometers and leg-attached 3D-accelerometers (sensitivity ~80–90 percent and specificity ~>90 percent). However, a successful insemination was used as gold standard; successful inseminations are for certain preceded by true estrus cases and also most certainly not all true estrus cases resulted in successful inseminations. Therefore, the reported performances for such sensor systems should be considered with caution. Progesterone measurements can be considered to be the gold standard for estrus sensors *(Cavalieri, et al., 2003a, Friggens, et al., 2008)*, so progesterone sensors seem to be promising sensor systems, although not much has been published on the performance of such sensor systems. Furthermore, progesterone measurements in laboratories are expensive, so an on-farm progesterone sensor system will be costly. Whether or not such a system is profitable remains unknown.

Activity meters showed combinations of high sensitivity (~80–90 percent) and specificity (~>90 percent) with milk progesterone measurements as a gold standard. However, this performance was reported in a single study, which suggests using some caution when valuing this performance. HeatWatch, microphone and temperature implant did not show better performance than pedometers or activity meters. A video camera system with automated image analysis has been tested (sensitivity 85 percent and specificity 99 percent *(Alawneh, et al., 2006)*) with successful inseminations as a gold standard. Automated video analysis requires cows to be within range of the video camera and exhibit behavior that the sensor system can recognize. Because of this, and the used gold standard, the value of this sensor system in practice would seem questionable. In some studies, the cycles are synchronized before the start of an evaluation experiment, for pedometers a sensitivity of 81 percent *(Cavalieri, et al., 2003a, Cavalieri, et al., 2003b)* and for HeatWatch a sensitivity of 88 percent *(Cavalieri, et al., 2003a, Cavalieri, et al., 2003b, Cavalieri, et al., 2003c)* were reported in these studies. Sensor systems that detect estrus have added value, as farmers are known to miss cases of true estrus by visual observation *(Firk, et al., 2002)*. However, important information about whether and/or when to inseminate a cow could be integrated in these sensor systems to improve the quality of the information provided to the farmer.

Locomotion

Level I: 'sensor technique'. Thirty-eight publications describing a sensor system were found for automated detection of locomotion problems, five (13 percent) of these publications were proceedings papers. Mostly, weight distribution between the cow's legs (17 studies, 45 percent) and walking behavior (16 studies, 42 percent) were studied. Some other publications studied (walking) activity (five studies, 13 percent). The sensors used to measure activity were pedometers and activity meters, attached to the cows' leg or to a neck collar and therefore classified as on-cow sensors. The sensors used for walking behavior were 3D-accelerometers and video cameras, which were classified as on-cow and off-cow sensors, respectively. The sensors for the weight distribution between the cows' legs were four balance weighing floors, weighing platforms, two-parallel force plates, and force distribution plates. Because these sensors require the cow to stand on them or walk over them, they are classified as off-cow sensors.

Level II: 'data interpretation'. Figure 2 shows that 18 (47 percent) sensor systems included data interpretation, eight (21 percent) of which had validated their algorithm. Sensitivity and specificity were reported for six (16 percent) of the algorithms.

Level III: 'integration of information'. None of the sensor systems integrated the sensor information with economic information (Figure 2).

Level IV: 'decision making'. Seven (18 percent) of the published sensor systems provided the farmer with an alert for abnormal locomotive behavior, while four (11 percent) also provided a probability with this alert and 14 (37 percent) provided graphs of processed sensor data (such as weight or walking activity over time). Nine (24 percent) of the published sensor systems provided farmers with raw sensor data (such as weight distribution over cows' legs) and/or were unclear in describing what they provided to the farmer.

System quality. A combination of high (above 80 percent) sensitivity and specificity was reported for four (11 percent) of the published sensor systems. The other studies reported performance of detection in the range of 22–80 percent for sensitivity, specificity, or both. Four (11 percent) of the published sensor systems were tested on two or more farms, while other studies reported a test scale of one farm. In this last group of studies, 14 (37 percent) of the published sensor systems were tested on 15 or fewer cows.

For detection of locomotion problems, pedometers and 3D-accelerometers have been studied most frequently. However, video camera systems, with automated image analysis, have also been tested. For locomotion, the performance seems to be high in an experimental setting; however, cows needed to be guided to walk in an appropriate manner in front of the camera (correct walking speed, proper distance to the camera, and one cow at a time). Consequently, application in practice seems to be difficult because it requires important adjustments in barn layout and operational management. The studies on sensor systems for locomotion problems showed an association between sensor data and lameness

(either statistical or in data patterns). It seems as though sensor systems could potentially discriminate between clinically lame and non-lame cows. Therefore, the studies seem to focus mostly on finding a good lameness indicator and a good way to assess it. Similarly, the demands for a lameness alert are not clear and subsequent actions by the farmer have not been studied, which means it is unknown whether the sensor systems provided added value compared to farmers' visual observation of the cows' gait. Furthermore, it remains unclear whether the sensor is only able to detect severe locomotion problems that are also easily detectable by visual observation.

Metabolism

Level I: 'sensor technique'. For automated detection of metabolism parameters, 16 publications describing a sensor system were found, six (38 percent) of these publications were proceedings papers. Mostly, the pH of rumen fluid (six studies, 38 percent) or the rumen temperature (three studies, 19 percent) is studied with a sensor. Some other publications studied sensors that measured the percentage of milk fat (one study, 6 percent), the level of ketone bodies in the milk (three studies, 19 percent), the cows' rumination or walking activity (two studies, 13 percent), or the cows' body temperature (one study, 6 percent). Both the pH and temperature sensors were radio telemetric rumen boluses, which made them in-cow sensors. Furthermore, the percentage of milk fat was measured with an in-line spectrophotometer. The walking activity was measured in these studies using a pedometer (measuring walking activity) that was attached to the cows' hind leg (making it an on-cow sensor). Body temperature was measured with a thermal camera (that is, an off-cow sensor) with automated image analysis to produce a body condition score. Finally, the level of ketone bodies in the milk was measured with an on-line spectroscopy sensor or the HerdNavigator® (that is, an on-line sensor).

Level II: 'data interpretation'. Figure 2 shows that five (31 percent) of the published sensor systems include data interpretation, of which three (19 percent) had validated their algorithm. Sensitivity and specificity were reported for one (6 percent) of the published sensor systems.

Level III: 'integration of information'. Figure 2 shows that none of the sensor systems integrated the sensor information with economic information.

Level IV: 'decision making'. One sensor system provided the farmer with a ketosis alert. Other sensor systems provided the farmer with milk fat percentage or raw sensor data. These forms of sensor system output require farmers to interpret results from sensor measurements and there meaning for the cows' health themselves .

System quality. Many sensor systems regarding metabolism provided the farmer with either raw sensor data (e.g., rumen pH) or indicators (e.g., milk fat percent) that are not

related to clearly defined problems or actions. One sensor system gave alerts for ketosis with a sensitivity of 70 percent, which is not considered to be high. This means that other factors like DIM are needed in order to accurately calculate the risk of a cow suffering from sub-clinical ketosis *(de Roos, et al., 2007)*. So, although specificity is 95 percent and the test scale was quite large with 217 farms included in the study, this sensor system had limitations. All reviewed articles involved testing on one farm and often included a small number of cows, part from one other study that was conducted on multiple farms *(Edwards and Tozer, 2004)*.

One study describing a rumination sensor attached to a neck collar, that records sounds of the rumination process with a microphone *(Bar and Solomon, 2010)* has been found. Although changes in rumination behavior was suggested to have a relation with cows health, the relation between rumination behavior and cow health was not studied.

For pH- and temperature boluses, it remains unclear what the added value is to the farmer, as no interpretation of the data is provided. While the added value of information seems obvious for detection of early lactation diseases and ketosis. However, these sensor systems need refinement, enhanced detection and better data presentation. It was not clear whether the chosen devices measure the most appropriate parameters for detecting early lactation diseases and ketosis. It remains somewhat unclear whether the detection of metabolic problems by sensors has added value, given that the unclear link between sensor information about the cow's metabolic status and a specific problem (hence a treatment/action). In effect, the studies seem to focus mainly on whether it is feasible to measure a parameter from the cow correctly with a sensor.

DISCUSSION

Main findings

This review has summarized a total of 126 published studies describing 139 sensor systems for animal health management. The number of publications for mastitis (25 percent), fertility (33 percent), and locomotion (30 percent) is approximately two to three times higher than the number of publications on sensor systems to detect metabolic diseases (13 percent). There are three possible explanations for this difference. The first is the fact that sensors for fertility and mastitis have been studied for a longer time. Secondly, the literature has shown that mastitis, locomotion problems, and fertility problems are currently the main health issues on dairy farms *(Bruijnis, et al., 2010, Groenendaal, et al., 2004, Halasa, et al., 2007)*. Finally, metabolic problems are complex disorders with a loss of milk production as main clinical characteristic. This drop in milk production, however, is seen with many diseases (e.g., mastitis) and therefore is not specific enough to indicate metabolic disorders. It

therefore might be more difficult to find an appropriate indicator for metabolic problems that could be measured with a (on-farm) sensor.

Much work has been done at levels I and II of Figure 1 and no studies were found for levels III and IV. For mastitis (92 percent of the publications on mastitis sensors) and fertility (75 percent of the publications on estrus detection), most work is done on level II. For locomotion (53 percent of the publications on detection of locomotion problems) and metabolism (69 percent of the publications on detection of metabolic problems), more than half of the work is done on level I. Most studies on mastitis (73 percent of the publications on mastitis sensors) and fertility (46 percent of the publications on estrus detection) described a validated algorithm, while fewer validated algorithms are available for locomotion (21 percent of the publications on detection of locomotion problems) and metabolism (19 percent of the publications on detection of metabolic problems). For locomotion and metabolism, approximately half of the studies describe a sensor with nothing more than an assumed relation to a disease, for mastitis and fertility this is 8 percent and 25 percent, respectively. Test scales of the studies vary considerably. For mastitis and fertility, test scales are generally larger than the test scales for locomotion (four studies) and metabolism (six studies). For example, 10 cows or fewer were used in four locomotion studies and in six metabolism studies, whereas studies on sensors for mastitis and fertility all included more than 10 cows. Three locomotion studies and four metabolism studies included more than one farm. Eleven mastitis studies and eleven fertility studies included more than one farm. The effect of test size on detection performance is unclear, as publications with different test sizes also vary in one or more other aspects (such as sensor type, algorithm, or gold standard).

It is unclear whether sensors that measure locomotion or standing behavior can detect leg or claw problems at an early stage. As leg and claw disorders are painful for a cow, she tries to cope with this pain by changing her walking and standing behavior *(Juarez, et al., 2003)*. Therefore, cows that already have pain will be detected by the sensor. From a welfare perspective, earlier detection would be more sensible. Whether or not the published sensor systems for locomotion problems could be useful tools on dairy farms to improve animal welfare remains the question. Accordingly, it is shown that specificity was higher for severely lame cows (higher gait score) *(Bicalho, et al., 2007)*. Furthermore, leg–weight ratio was a more accurate lameness predictor when a gait score of 3.5 was used as a lameness threshold instead of a gait score of 3 *(Pastell, et al., 2010)*. These two remarks, as well as the low detection performance that was observed for locomotion sensor systems in general, might suggest that mostly severe lameness could be detected with these sensor systems. Severely lame cows have a strongly abnormal gait and/or remain lying down more and a farmer could identify these animals easily with a quick observation. Furthermore, economic studies indicate that subclinical cases of leg or claw disorders cause economic losses as well as clinical cases *(Bruijnis, et al., 2010)*. Studies that used a scoring system (such as locomotion or gait scoring) defined a cow with a score of 3 or higher as lame. Cows with a score of 2 do

have a non-normal gait and are possibly sub-clinically lame. Hoof trimming of cows with lameness score 2 might prevent (severe) clinical lameness and prevent (more) painful lesions in the cows claws. However, the sensor systems ignore these cows completely, which means that the sensor system produces no information that can be used for early treatment of locomotion problems. A variety of mathematical models have been used to build algorithms for automated detection and these algorithms also vary greatly in reported sensitivity and specificity. No sensor systems have been found to match the ISO standard for sensitivity and specificity, which has been formulated as appropriate for automated detection of mastitis. Due to the large variation in performance, the algorithms used, and the study design, it is not possible to make general statements on the appropriateness of the algorithms.

The output (together with detection performance) of a sensor system determines the value for the farmer. Raw sensor data requires farmers to interpret the data themselves in order to determine whether something has changed in the cow's health status. When the sensor system contains an algorithm that produces an alert, the interpretation for the farmer is been made easier as it gives a clear statement about the cow's health status (the reliability of this statement is a different issue). Within the alerts, estrus and mastitis are more informative for the farmer as these alerts do regard a specific condition. Alerts for locomotion problems and metabolic problems are less informative because these alerts regard abnormal values for a locomotion or metabolism indicator, but not the specific disease or disorder causing the abnormality. Consequently, the farmer is required to examine the cow or other data on the cow (such as milk production records) to determine the cow's health status. By contrast, alerts for mastitis and estrus indicate a specific disease or event and the farmer only needs to decide what to do.

No publications have been found that mention some decision support for dairy farms, based on sensor data. Therefore, all sensor systems require the farmer to rely on his herdsmanship (that is, the combination of the farmer's intuition, experience, and knowledge as a herdsman) in interpreting the available information to make a decision. Sensor systems that provide a probability can make it easier for the farmer to interpret the available information, for instance, to distinguish between more and less important alerts.

History

The first publications on automated detection to assist animal health management can be found in the period between 1970 and 1990. The development of the responder that enables the automatic identification and data collection of individual cows can be seen as the start of this research field. Research institutes in Germany (1976), the UK (1973), the Netherlands (1976), and the US (1976) started by developing these transponders and tested them in practice *(Rossing, 1999)*. Among the first applications was the automated individual feeding of dairy cows *(Ipema and Rossing, 1987)*. Experiments were done with electrical

conductivity to detect mastitis and pedometers to detect estrus. In the mid-1970s, it was recognized that visual observation of estrus was becoming more difficult in increased herd sizes; therefore, advice was given for improving visual observation, such as detection aids and reserving 20–30 minutes for observation, twice a day *(Foote, 1975)*. As an alternative to visual observation, pedometers were tested in the early 1980s. For example, a comparison of detection methods concluded that pedometers could observe slightly fewer estrus cases than could be done with visual observation; the pedometers were not fit for remote data recording at this time *(Williams, et al., 1981)*. Pedometers are often used in Israel for fertility management in large dairy herds, while cycle synchronization is widely used in the United States *(Galon, 2010)*. The pregnancy rate of inseminated cows has been shown to be equal for estrus synchronization and estrus detection with pedometers *(Galon, 2010)*. This result implies that automated estrus detection might be an alternative for estrus synchronization in the fertility management of larger dairy herds. The first studies focusing on electrical conductivity as a measure for mastitis were performed in the late 1960s; relations between EC and mastitis indicators were found but interpretation of EC measurement caused difficulties, due to large variation in EC values *(Greatrix, et al., 1968, Little, et al., 1968)*. The first suggestions for in-line EC measurement can be found in the literature around the mid-1970s *(Linzell and Peaker, 1975, Wolfe, et al., 1972)*. Although research was done in those years, a breakthrough only occurred when the number of farms using an AMS increased around 2000. Subsequently, much research was done on refining detection models for mastitis and on alternative sensors to detect mastitis. New sensor systems were developed because of increasing herd sizes, increasing labor cost as well as the increased use and familiarity with information technology in general and sensor systems specifically, due to experience with an AMS, for example. Next to these trends in dairy farming, trends in public opinion in Western countries should also be considered when developing new sensor systems. It follows that it is important for dairy farmers to improve/maintain the welfare of their cows and to show that this is done to society. From this perspective, sensor systems could be seen as an statement of how the dairy sector is taking animal welfare seriously and is taking its responsibility in safeguarding animal welfare.

As discussed, techniques like EC and pedometers are somewhat older than other sensors techniques mentioned in this review. Studies were published on these techniques in the 1980s and 1990s, which excludes them from the current review, which only covers publications since 2002. Previous reviews have discussed older literature on EC and pedometers *(Brandt, et al., 2010, Firk, et al., 2002, Hogeveen, et al., 2010, Hovinen and Pyörälä, 2011, Norberg, 2005)*. Locomotion and fertility seem to be relatively new topics in sensor research, which means that the present review was likely to include most of the publications on these topics.

Gold standard

The choice for a gold standard is important for the detection performance of a sensor system. How well a gold standard reflects reality determines the number of 'true' cases used for algorithm development and validation. If true cases of disease or estrus are missing, or false cases are included in the gold standard data (visual observation and scoring system are sensitive for this problem), then the processes of algorithm building and validation will be affected. As cases will be missing or false cases will be included in the dataset the algorithm will be misspecified and in the validation some alert will be wrongly classified as false positive or false negative. In a more practical sense, the intended purpose of a sensor system is important when choosing a gold standard. For the substitution of labor by capital – which means that the sensor system will do a farmer's job – a gold standard that reflects a farmer's detection capabilities could be appropriate. However, for an early warning system, the gold standard should be able to correctly pick up disease or estrus at an early stage. Another example of the relation between gold standard and intended purpose is detection of clinical mastitis for which visual judgement of milk for clots is appropriate. By contrast, using SCC and CMT as gold standard would include detection of subclinical cases. It is also important to consider the frequency at which the gold standard is determined in the studies. Overly long intervals between gold standard assessments will result in missed true cases, whereas short intervals increase the workload involved, and consequently the cost of the experiments. An example of an overly long interval could be using monthly SCC estimates from milk quality controls as a gold standard. This would result in missed (sub)clinical mastitis cases or late detection of (sub)clinical mastitis cases in the reference data used to build and validate a detection algorithm. An example of a short interval would be daily progesterone measurements. It would mean collecting milk samples per cow each day and analysis in a laboratory, which means high costs and workload.

Time resolution

Time resolution (also referred to in the literature as the time window of detection) can be split up into two slightly different concepts. The first is the time resolution of the sensor, including measurements, interpretation, and detection. The second is the time resolution of validation, which means matching of alerts to gold standard measurements in a defined time period.

The frequency used to record measurements of the sensor have an influence on the minimal time of detection. Furthermore, an increased number of measurements decreases the influence that erroneous measurements have on predictions of the detection algorithm. For the detection algorithm, the frequency with which it produces information about a cows' health, based on sensor data, partly determines the time between changes in the cows' physiology and the farmer being informed of the change. Obviously, the time between

detection and informing the farmer also depends on the farmer. Technical innovations (like smartphones) might be able to help inform the farmer with sensor information more quickly. The time between the change in the cow's health in reality and informing the farmer is also important in relation to various diseases. For instance, a locomotion problem, although painful, does not require the farmer's immediate attention, whereas estrus only has a limited period during which successful insemination is possible.

As both the gold standard and sensor system information are point estimates that are measured or determined with their respective frequencies, validation studies must match these point estimates to each other. For mastitis, an in-depth analysis has been conducted of the influence that "time resolution of validation" has on sensor system performance and its use in practice *(Hogeveen, et al., 2010)*. A longer time period for matching gold standard and alerts generally results in higher sensitivity and specificity. Early detection can be accomplished by producing information regarding the cows health before clinical signs of disease occur *(Hogeveen, et al., 2010)*. However, information provided by the detection algorithm after the onset of disease postpones treatment, whereas the farmer may perceive information prior to onset as a false alert, they consider a maximum 24 hours before onset of the disease as desirable *(Mollenhorst, et al., 2012)*. For estrus detection there cannot be much debate about the fact that a timeframe of about 24 hours is needed to enable insemination after an estrus alert. For locomotion and metabolic problems the time component is less clear, as the detected conditions (diseases) of the cow remain unclear. Therefore, it in unknown what a farmer should do with the information, let alone the farmers preferences can be studied.

In the literature used for this review, it was not always clear whether the term "time window of detection" referred to the "time resolution of the sensor" or the "time resolution of validation." In addition, for the "time resolution of the sensor," often only one or two single aspects, such as the frequency the sensor used to take measurements, were mentioned. What is more, the literature only discussed aspects of the time resolution briefly, if at all. So, considering the publications in the review, the time resolution of the sensor systems remains a greatly neglected aspect in publications.

Sensor systems excluded from this review

Devices not described in (recent) scientific literature. This overview might not include all sensor systems that have been developed, patented or are available on the market. This is because the present review has only summarized the sensor systems published in scientific literature in recent years. A few examples of sensors that are on the market, but not documented in the scientific literature are the "SensOor" (Agis Automatisering, Harmelen, the Netherlands), which is a temperature, motion, and sound sensor that is attached to a cow's ear tags, and the "Vel Phone" (Amelis Medria, Carpiquet, France), a vaginal

thermometer that predicts the moment of calving. By contrast, the scientific literature has reported on certain sensors for which it is not clear if they are available on the market. For sensors like the HerdNavigator®, IceTag™, and the sensors of AMS's of Lely (Lely Industries, Maassluis, the Netherlands) and DeLaval (DeLaval, Tumba, Sweden), clear connections between scientific publications and commercially available sensor systems can be made.

Devices that are not (yet) sensors. Some relevant publications do not fit the used definition of a sensor for the support of animal health management, either because the sensor systems are not (yet) automated devices or because they are applied for optimization of production rather than health management.

Thermal cameras have been used for the detection of mastitis; however, these devices were handheld *(Berry, et al., 2003, Colak, et al., 2008, Hovinen, et al., 2008, Polat, et al., 2010)*. While these papers have described an association between mastitis and udder skin temperature, no validated algorithms have been described. These cameras could be developed into a sensor system by, for example, installing them in an AMS to take measurements during milking.

It is necessary to mention certain publications that have used laboratory tests and simulation studies as they describe the principles upon which other sensor systems like the Herd Navigator® are based. The Herd Navigator® itself has been scarcely described in the literature. The relation between the chosen indicator and a reference test has been studied for mastitis *(Chagunda, et al., 2006a)*. The models calculate a risk of disease and when a cow should be tested by the sensor system *(Chagunda, et al., 2006b, Friggens and Chagunda, 2005, Friggens, et al., 2007)*. The model for estrus detection has been tested with real cow data *(Friggens, et al., 2008)* and the economic impact of such a sensor system on the farm has been studied *(Østergaard, et al., 2005)*. Furthermore, for estrus information of the sensor, the estrus alert is integrated with economic and other information. The information used was prices for milk, livestock, feed, veterinary services and insemination using a budgeting approach *(Østergaard, et al., 2005)*. Therefore, although the complete system has been only described in proceedings papers, the underlining principles have been studied in greater depth.

An example of a sensor system that is not included in this review because it is not related to animal health management is the dynamic concentrate feeding system for dairy cows that includes all levels of Figure 1 *(Andre, et al., 2010a, Andre, et al., 2010b, Andre, et al., 2011)*. In short, this system monitors the cows' response in milk yield to an increase in concentrates. It then optimizes the concentrate supply by calculating the net merit of the increase in milk yield from extra concentrate supply by considering milk price (including milk fat and protein) and concentrate price. This work can be seen as a very specific application of sensor techniques, as the system is capable of making an automated decision for the farmer. In animal health management, this might not be feasible for many applications, as an automated treatment of the cow appears to be technically infeasible at this moment.

Decision support for animal health management

Studies on decision support and economic considerations of management decisions are available for both mastitis and fertility, but they are not integrated in a sensor system. For mastitis, decision support systems provide advice for clinical cases *(Halasa, et al., 2007, Pinzon-Sanchez, et al., 2011, Seegers, et al., 2003, Steeneveld, et al., 2011)*. However, subclinical mastitis may be detected by a sensor system and could be relevant as a sub-clinical case could develop into clinical mastitis or chronic subclinical mastitis. Literature is currently available that describes the treatment effectiveness and economic implications of treating subclinical mastitis *(Steeneveld, et al., 2007, Swinkels, et al., 2005a, Swinkels, et al., 2005b, van den Borne, et al., 2010)*. Treatment of persistent sub-clinical mastitis is economically profitable for many cows when indirect effects of cure are considered (that is, prevention of clinical flare-ups and transmission to other cows) *(Swinkels, et al., 2005b)*. However, most studies on sensor systems for mastitis still focus on detecting clinical mastitis and do not determine sub-clinical cases for which treatment would be profitable. Another limitation is that many decision support systems were developed to support cow health management when milking in a milking parlor. However, the current trend in Western Europe is the use of AMSs, which have changed the operational management on dairy farms dramatically. A few studies have focused on pathogen specific treatment of mastitis, although the economic merit of pathogen specific treatment seems to be absent *(Steeneveld, et al., 2011)*. Decision support systems are available for fertility management (e.g., Groenendaal et al., 2004; Olynk and Wolf, 2009; Inchaisri et al., 2010,;2011) and could be used to develop sensor systems with integrated decision support. Some refinement could be useful, such as translating sensor measurement into an optimal insemination moment that corresponds with the decision support system. Furthermore, the certainty associated with an alert (that is, the degree to which the sensor system is certain that this alert is true) could be a nice addition to a decision support system. Not a lot of literature is available on the effectiveness of treatments for locomotion and metabolism, or the economic impact of such treatments. For locomotion, the economic impact of lameness *(Bruijnis, et al., 2010, Ettema and Østergaard, 2006)* and the effect of preventive measures *(Ettema and Østergaard, 2006)* have been studied. For metabolism, a lot of research has been conducted on animal nutrition to find the optimal feeding strategy to prevent metabolic problems; for example, fatty liver and negative energy balance *(Bobe, et al., 2004, van Knegsel, et al., 2007)*.

Economic implications

For farmers, the decision to invest in sensor technology to support cow health management will depend largely on the profitability of such a sensor system. In addition, a farm's economic prospects and financial position (farm solvency, for example), the presence of a potential successor, and farm size are general factors that underlie investment decisions

(Aramyan, et al., 2007, Oude Lansink, et al., 2001). The economic benefits of an automated estrus detection system have been studied, such as the simulation study based on the average characteristics of a Dutch dairy herd (e.g., 7,500 kg milk, estrus detection rate of 50 percent and conception rate of 40 percent) by van Asseldonk et al.*(1999b)*. Under the assumption that estrus detection was improved from 50 percent to 90 percent, gross margin would increase with Dfl. 1.25 (€0.57) per 100 kg fat and protein corrected milk *(van Asseldonk, et al., 1999b)*. Accordingly, this increase in the gross margin of a sensor system determines the financial room (or investment potential for a sensor system) for the annual costs of a sensor system (including investment and operational costs). Investment in sensor systems can be analyzed by calculating the net present value for investment profitability, including milk or slaughter, for example, and expenses (such as feed or disease), combined with a stochastic Monte Carlo simulation to assess the uncertainty inherent to dairy farming *(Bewley, et al., 2010b)*. For example, the profitability of investment in an automated BCS system depends largely on the current BCS distribution of a herd, the impact of non-optimal BCS and the ability to make changes to achieve optimal BCS *(Bewley, et al., 2010a)*. However, these studies ignore the effect of a sensor system on labor requirement of the dairy farm, which also has economic importance. For the economic implications of sensor systems in dairy health management only the mentioned studies of Bewley et al.*(2010a, 2010b)* and van Asseldonk et al. *(1999a, 1999b)* have been found.

Towards application in practice.

Farmers should consider various factors in their operational management, such as the cows' condition, expected development of the cows' condition, the expected effect of an action, and the costs of (not) taking action. For most farmers, these considerations are a "guestimate," at best, and mostly based on their herdsmanship *(Groenendaal, et al., 2004)*. It is difficult to make an optimal treatment decision based on all available information, treatment options, and cure-influencing cow factors *(Steeneveld, et al., 2011)*. Therefore, decision support systems can help farmers make these decisions based less on farmers' herdsmanship and more on the economic value of an individual cow *(Groenendaal, et al., 2004)*.

Based on the findings of this literature review, a few suggestions, both general and specific, can be made for future research. In sensor studies, separate datasets should be used for algorithm building and validation in order to show appropriate and inter-study comparable performance. Furthermore, the time resolution of a sensor system should be described more clearly.

For mastitis, there is a need for better detection models in practice. Whether the focus needs to be on new algorithms or new sensors is open for debate. Decision support systems have been developed for mastitis, but these were not developed for detection based on

sensor data and did not consider the uncertainties (like false alerts or missed cases) associated with this detection.

Various types of sensors have been studied for estrus detection and, although performance varies, promising systems have been described. Various studies have also been published on decision support. An interesting avenue for new research could be the use of an integrated fertility system in practice. Another possibility could be to compare the economic (and technical) performance of farms using sensor technology for fertility management and farms that do not use these sensors.

For locomotion, sensors have been suggested and tested, but the performance of detection models needs further research. Definition of the gold standard for locomotion calls for further study, especially with regard to the appropriate frequency for inspection. It would be helpful to know what lameness cases (in terms of severity, number of cows, infectious or non-infectious etc.) a sensor system should find. Only a few studies have described decision support models. New research could be conducted on the effect that lameness treatment has on cow health and farm income.

For metabolic problems, more research is needed to identify the indicators for developing metabolic problems; for example, the relation between rumen pH and rumen health. This knowledge would be needed to develop detection models that clearly indicate to the farmer whether something is wrong with the cow and, if so, what.

CONCLUSIONS

Most studies concerned the detection of mastitis (25 percent), fertility (33 percent), and locomotion problems (30 percent), with fewer studies (16 percent) related to the detection of metabolic problems. Many studies presented sensor systems at levels I and II, but none did so at levels III and IV. Most of the work for mastitis (92 percent) and fertility (75 percent) is done at level II. For locomotion (53 percent) and metabolism (69 percent), more than half of the work is done at level I.

Sensor systems for mastitis and fertility have been developed to higher levels (see Figure 1) than for leg and claw problems and metabolic problems. Most published studies for mastitis and fertility clearly describe what disease they are aiming to detect, and most of these studies focus on the performance of the sensor system. For locomotion, the studies focus on finding a good method of measuring a parameter and detecting locomotion problems. For metabolic problems, the studies focus on finding good methods to measure a parameter, although the relation with cow health remains unclear for these measurements.

For sensors systems, there is no clear difference in the performance of various algorithms. Detection performance of the sensor systems varies based on the choice of gold standards, algorithms and test sizes (number of farms and cows). The most important remark for further sensor research is to have a clear aim of what information about the cows health should be produced by the sensor system under study. In respect to the aimed information an appropriate gold standard, algorithm, test size and time resolution should be chosen.

Analysis of investment in sensor systems has been scarcely published. Similarly the economic and management value of sensor information on farms remains unclear. No published sensor systems have an integrated decision support system.

ACKNOWLEDGEMENTS

Appendix 2.1: Characteristics of sensors for automated detection of mastitis in dairy cows. Level I describes the sensor and its place in respect to the cow. Level II describes the data processing with the used gold standard (GS), used algorithm (Alg), the level of algorithm validation (val), whether the algorithm uses non-sensor data (non-sensor), the sensitivity (SE(%)) and specificity (SP(%)) of the algorithm, the numbers of gold standard positive (+) and negative (-) cases in the used test (T cases) and validation (V cases) datasets and the test size in number of farms (F) and number of cows (C). A '-' indicates that the characteristic was not found in the publication, a '?' indicates that the description of the characteristic was unclear in the publication.

| Source | Level I | | Level II | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	Se(%)	SP(%)	T cases	V cases	Tested
Miekley, et al., 2012	EC[1]	In-line [2]	RCM, SSC	MA		No	Alert	75	18	-	-	1F
Liberati and Zappavigna, 2009	EC	In-line	?	FL	T	Yes	Alert, Pr	?	?	?	?	1F 40C
Cavero, et al., 2006	EC	In-line	SCC	FL	V	Yes	Alert, Pr	83-92	75-94	+126/+1612/+620	+70/+736/+322	1F 478C
Cavero, et al., 2008	EC	In-line	SCC	NN	V	Yes	Alert, Pr	78-84	51-75	?	?	1F 478C
Cavero, et al., 2007	EC	In-line	SCC	MA	V	No	Alert	85	73	?	?	1F 160C
				MA	V	No	Alert	84	73	?	?	1F 160C
				MA	V	No	Alert	88	67	?	?	1F 160C
Claycomb, et al., 2009	EC	In-line	?	Abs	V	No	Alert	68-86	?	-	-	1F 194C
Hassan, et al., 2009	ER	In-line	?	NN	T	Yes	Clas. NI min. maj.	-	-	?	?	1F 111C
Højsgaard and Friggens, 2010	EC	In-line On-line	?	FA	V	No	degree of Mas	?	?	+58 -71	-	1F 332C
Hovinen, et al., 2006	EC	In-line	SCC	Abs	V	No	Alert	?	?	?	?	4F 112C

Appendix 2.1: Continued

| Source | Level I | | Level II | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	Se(%)	SP(%)	T cases	V cases	Tested
Kamphuis, et al., 2008b	EC	In-line	RCM	Abs, MA	V	No	Alert	80[3]	-	-	?	1F 194C
Kamphuis, et al., 2008b	EC, CellSense	In-line, On-line	RCM	FL	V	No	Cl mas likelihood	80[3]	-	-	?	1F 194C
Lukas, et al., 2009	EC, Milk yield	In-line	RCM	MA	V	Yes	Health Alert	?	?	?	?	1F 587C
Mollenhorst, et al., 2010	AMS sensor	In-line, On-line	VJ, SCC	IQR, Abs	V	No	Alert	100	99	-797 +19(EC) +9 (SCC)	-	3F 191C
Norberg, et al., 2004	EC	In-line	RCM	Reg.	V	No	Clas. He Su Cl	?	?	?	?	1F 320C (5 year)
Sloth, et al., 2003	EC	In-line	RCM	Reg	T	No	?	-	-	-520 +301	-	1F 322C
Sun, et al., 2010	EC	In-line	VJ	NN	V	Yes	Alert, Pr	91	87	-2265 +627	-970 +268	1F 94C
Zecconi, et al., 2004	EC	In-line	OM	Reg	V	No	Alert	55-60	49	-1249 +165	?	3F 850C
	EC + COL	In-line	SCC	Abs	V	No	Alert	?	?	-1249 +165	?	4F 171C
Steeneveld, et al., 2010	EC+COL	In-line	VJ	Reg	V	Yes	Alert, Pr	70	97.8	-339,011 +151	-169,506 +76	1F 602C
Kamphuis, et al., 2010a	EC+COL	In-line	VJ	DTI	V	Yes	Pr Mas	5-56	93-99	-305 +87	-34 +10	6F 838C
Kamphuis, et al., 2010b	EC+ COL	In-line	VJ	DTI	V	Yes	Pr Mas	40-85	57-89	-24,717 +243	-50,000 +105	9F 1109C

	Level I		Level II									
Source	Sensor	Place	GS	Alg	Val	Non-sensor	Output	Se(%)	SP(%)	T cases	V cases	Tested
Kamphuis, et al., 2008a	EC+COL [4]	In-line	VJ	Cor	T	No	No	-	-	-2962 +270	-	1F 65C
Song and van der Tol, 2010	EC+COL	In-line	RCM	Abs	V	Yes	Alert	85	99	-	+26	5F 649C
Akerstedt, et al., 2006	Optical biosensor	On-line	ELISA	Abs	A	No	No	-	-	-43 +28	-	1F ?C
Chagunda, et al., 2006b	?	On-line	?	Reg	V	Yes	RAM, RCM, DNM	82	99	-	-19,109 +492	1F 100C
Welbeck, et al., 2011	CM5 chip	On-line	ELISA	No	A	No	No	-	-	-	-	-
Mazeris, 2010	HN	On-line	?	?	A	Yes	RV	80	-	-	-	?
Vreeburg, 2010	HN	On-line	?	?	V	Yes	RV	82	98	-	-	2F 286C
Kamphuis, et al., 2008b	CellSense	In-line	SCC (Lab)	Abs	V	No	Alert	83	-	-	?	1F 194C
Bewley and Schutz, 2010	TB	Reticulum	?	MA	T	Yes	No	-	-	-131,080 +101	-	1F 298C

[1] Used abbreviations:

Sensor: Milk color (Col); Electrical Conductivity (EC); Electrical Resistance (ER); Herd Navigator (HN)Temperature (TB);

GS: Bacteriological culturing (BC); Observation by Milker (OM); Recorded cases of Clinical Mastitis (RCM); Visual Judgement (VJ);

Alg: Absolute value (Abs); Correlation (Cor); Decision Tree Induction (DTI); Factor Analysis (FA); Fuzzy Logic (FL); Inter Quarter Ratio (IQR); Moving Average (MA); Neural Network (NN); Regression (Reg);

Val: Assumed relation between gold standard and sensor data (A); Statistically tested relation between gold standard and sensor data (T); Validated algorithm for detection (V);

Output: Clinical (Cl); Classification (Clas.); Days to Next Sampling (DNM); Healthy (He); minor pathogenIMI (min); major pathogen IMI (maj); Mastitis (Mas); No Infection (NI); Probability (Pr); Probability of Gram Status (PrGS); Risk of Acute Mastitis (RAM); Risk of Chronic Mastitis (RCM); Risk Values (RV); Subclinical (Su);

[2] Milking parlour

[3] Sensitivity was fixed in this study in order to determine the model with the lowest false alert rate.

[4] Collection bin with electrical pins and Light emitting diodes (LED)

Appendix 2.2: Characteristics of sensors for automated detection of estrus in dairy cows. Level I describes the sensor and its place in respect to the cow. Level II describes the data processing with the used gold standard (GS), used algorithm (Alg), the level of algorithm validation (val), whether the algorithm uses non-sensor data (non-sensor), the sensitivity (SE(%)) and specificity (SP(%)) of the algorithm, the numbers of gold standard positive (+) and negative (-) cases in the used test (T cases) and validation (V cases) datasets and the test size in number of farms (F), number of cows (C) and number of heifers (H). A '-' indicates that the characteristic was not found in the publication, a '?' indicates that the description of the characteristic was unclear in the publication.

| Source | Level I | | Level II | | | | | | | | | | Tested |
	Sensor	Place	GS	Alg	val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	
O'Connell, et al., 2011	Pedometer	Leg	[Prog][1]	Vit	T	No	Alert, Pr	-	-	?	+58	1F 58C
Brehme, et al., 2008	ALT-pedometer	Left	VO	No	A	No	Graph	-	-	-	-	1F 10C
Brunassi, et al., 2010	Pedometer	Leg	[Prog] CP	FL	V	Yes	Alert[2]	84	98	+98	?	1F 98C
Edwards and Tozer, 2004	Pedometer	Leg	?	No	T	No	Alert	-	-	-	-	3F 1448C
Cavalieri, et al., 2003a[3]	Pedometer	Leg	[Prog]	MA	V	No	Alert	81	-	-[4]	-52 +43	3F 98C
Firk, et al., 2003b	Pedometer	Leg	SAI	FL	V	No	Alert, Pr	71-96	96-99	-	-	1F 660C
Firk, et al., 2003a	Pedometer	Leg	RAI	FL	V	Yes	Alert, Pr	88	99	-	-	1F 373C
Cavalieri, et al., 2003b[3]	Pedometer	Leg	[Prog]	MA	V	No	Alert	81	-	?	?	3F 98C
Holman, et al., 2011	Pedometer	Leg	[Prog]	Dev	V	No	Alert	59	-	-[4]	+158	1F 67C
Liberati and Zappavigna, 2009	Pedometer	Leg	?	FL	T	Yes	Alert, Pr	?	?	?	?	1F 107C
Roelofs, et al., 2005	Pedometer	Leg	VO	Dev	T	No	Alert	-	-	+63	-	1F 49C
Sakaguchi, et al., 2007	Pedometer	Leg NC	REO	TS	T	No	Alert	-	-	-	-	1F 15H

Appendix 2.2: Continued

| Source | Level I | | Level II | | | | | | | T cases | V cases | Tested |
	Sensor	Place	GS	Alg	val	Non-sensor	Output	SE(%)	SP(%)			
Gillis, et al., 2002	Biosensor	On-line ME	ELISA	No	A	No	No	-	-	-	-	?
Kappel, et al., 2007	Biosensor	On-line	?	No	A	No	No	-	-	-	-	1C
Sananikone, et al., 2004	Immunostrips	On-line	SC	No	A	No	No	-	-	-	-	-
Tschmelak, et al., 2005	RIANA biosensor	On-line	SC	No	A	No	No	-	-	-	-	-
Mazeris, 2010	HN	On-line	?	?	?	Yes	RV	-	-	-	-	?
Vreeburg, 2010	HN	On-line	?	?	V	Yes	RV	98-99	?	-	-	2F 286C
Hockey, et al., 2010a	Activity meter	NC	[Prog]	Dev	V	No	Alert	80-90	98	-[4]	+141	2F 199C
	Activity meter	NC	[Prog]	Dev	V	No	Alert	90-94	90	-[4]	+141	2F 199C
	Activity meter	NC	[Prog]	Dev	V	No	Alert	87-93	93	-[4]	+141	2F 199C
	Activity meter	NC	[Prog]	Dev	V	No	Alert	82-88	95	-[4]	+141	2F 199C
	Activity meter	NC	[Prog]	Dev	V	No	Alert	79-87	96	-[4]	+141	2F 199C
Hockey, et al., 2010b	Activity meter	NC	[Prog]	Dev	T	No	Alert	-	-	-[4]	?	2F 246C
Holman, et al., 2011	Activity meter	NC	[Prog]	Dev	V	No	Alert	63	-	-[4]	+146	1F 67C
Liberati and Zappavigna, 2009	Activity meter	NC	?	FL	T	Yes	Alert, Pr	?	?	?	?	1F 40C
Løvendahl and Chagunda, 2010	Activity meter	NC	SAI	Dev	V	No	Alert	?	?	?	?	1F ?C

Appendix 2.2: Continued

	Level I		Level II									
Source	Sensor	Place	GS	Alg	val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
Peralta, et al., 2005	Activity meter	NC	RAI	?	A	No	Alert	-	-	-	-	1F 1075C
Kamphuis, et al., 2012	Activity meter	NC	[Prog]	MA	V	No	Alert	77	99	-[4]	-22.660 +835	1F 635C
	Activity meter	NC	[Prog]	MA	V	No	Alert	62	99	-[4]	-22.660 +835	1F 635C
Balzer, et al., 2009	SmardWatch	NC	?	?	T	No	No	-	-	-	-	1F ?C
Kultus, et al., 2011	SmardWatch	NC	[Prog]	No	A	No	Graph	-	-	?	?	1F 21C
Jonsson, et al., 2011	Activity meter	Leg	SAI	GLR	V	No	Alert	89	99	?	+18	1F 18C
	Activity meter	Leg	SAI	GLR	V	No	Alert	50	100	?	+18	1F 18C
	Activity meter	Leg	SAI	GLR	V	No	Alert	89	100	?	+18	1F 18C
Muller and Schrader, 2003	Accelerometer	Leg	No	No	A	No	No	-	-	-	-	1F 12C
de Mol, et al., 2009a	2D Accelerometers	NC, Leg	MT	?	A	No	L/S	-	-	-	-	1F 6C
Palmer, et al., 2010	HeatWatch	CS	[Prog]	Dev	T	No	Alert	-	-	?	+95	1F 46C
Peralta, et al., 2005	HeatWatch	CS	RAI	?	A	No	Alert	-	-	-	-	1F 1075C
Cavalieri, et al., 2003a[3]	HeatWatch	CR	[Prog]	?	V	No	Alert	88	-	-[4]	-59 +43	3F ?C

Appendix 2.2: Continued

	Level I		Level II							T cases	V cases	
Source	Sensor	Place	GS	Alg	val	Non-sensor	Output	SE(%)	SP(%)			Tested
Cavalieri, et al., 2003b[3]	HeatWatch	CR	[Prog]	?	V	No	Alert	88	-	?	?	3F ?C
Cavalieri, et al., 2003c[3]	HeatWatch	CR	VO	Reg	T	No	Alert	-	-	?	?	1F 20C
Saumande, 2002	DEC-system	CS	B[Prog]	?	T	No	Alert	-	-	+96	-	1F 30C
Alawneh, et al., 2006	Video camera	ME	SAI	?	V	No	Alert	85	99	-[4]	-11257 +406	1F 240C
Schön, et al., 2007	Microphone	NC	VO	Reg	T	No	?	-	-	-	-	1F 10H
Miranda, et al., 2009	Temp transducer	I	No	No	A	No	No	-	-	-	-	?C
Bewley and Schutz, 2010	Temp Bolus	Ret	?	MA	T	Yes	No	-	-	-	-	1F 298C

[1] Used abbreviations:

Sensor: Herd Navigator® (HN);

Place: Cows Rump (CR); Cow Sacrum (CS); Implant (I); Milking Equipment (ME); Neck Collar (NC); Reticulum (Ret);

GS: Blood Progesterone level (B[Prog]); Confirmed Pregnancy (CP); Milking Times (MT); Progesterone level ([Prog]); Registered Artificial Insemination (RAI); Rectal Examination of Ovaries (REO); Successful Artificial Insemination (SAI); Standard Concentrations (SC); Visual Observation (VO);

Alg: Absolute Deviation from average or mean of previous time period (Dev); Fuzzy Logic (FL); Generalised Likelihood Ratio (GLR); Moving Average (MA); Regression (Reg); Time Series (TS); Viterbi algorithm (Vit);

Val: Assumed relation between gold standard and sensor data(A); statistically Tested relation between gold standard and sensor data (T); Validated algorithm for detection (V);

Output: Lying or Standing (L/S); Probability (Pr); Risk Values (RV);

[2] Alert categories: in estrus, maybe in estrus and not in estrus;

[3] Oestrus synchronization before start of experiment with sensor

[4] An available algorithm was used therefore the test dataset was not applicable.

Appendix 2.3: Characteristics of sensors for automated detection of locomotion problems in dairy cows. Level I describes the sensor and its place in respect to the cow. Level II describes the data processing with the used gold standard (GS), used algorithm (Alg), the level of algorithm validation (val), whether the algorithm uses non-sensor data (non-sensor), the sensitivity (SE(%)) and specificity (SP(%)) of the algorithm, the numbers of gold standard positive (+) and negative (-) cases in the used test (T cases) and validation (V cases) datasets and the test size in number of farms (F) and number of cows (C). A '-' indicates that the characteristic was not found in the publication, a '?' indicates that the description of the characteristic was unclear in the publication.

| Source | Level I | | Level II | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
Chapinal, et al., 2011	3D Accelerometer	Leg, CT[1]	Ga	Cor	T	No	?	-	-	-	-	1F 12C
Chapinal, et al., 2011	3D Accelerometer (IceTag3D™)	Leg	Ga + CHE	Reg	T	No	Alert	-	-	-	-	2F
Darr and Epperson, 2009	3D Accelerometer	Leg	VO	No	-	No	No	-	-	-	-	3C
Martiskainen, et al., 2009	3D Accelerometer	NC	VO	Reg	V	No	Alert, Pr	65	-	-	-94 +101	1F 30C
Pastell, et al., 2009	3D Accelerometer	Leg	LS	Stat	T	No	Graphs	-	-	-6 +6	-	11C
Scheibe and Gromann, 2006	3D Accelerometer	Leg, NC	?	No	A	No	Graphs	-	-	-2 +2	-	1F 4C
Nielsen, et al., 2010	3D Accelerometer (IceTag3D™)	Leg	VO	No	-	No	No	-	-	-	-	1F 10C
Tolkamp, et al., 2010	3D Accelerometer (IceTag3D™)	Leg	No	No	-	No	Graphs LF	-	-	-	-	1F
Ledgerwood, et al., 2010	3D Data loggers	Leg	VO	No	-	No	No	-	-	-	-	2F
Mottram and Bell, 2010	3D Accelerometer	NC	VO	No	A	No	Graph	-	-	-	-	1F 20C
de Mol, et al., 2009b	3D Accelerometer	Leg	?	MA	A	No	StC	-	-	-2 +1	-	1F 3C
de Mol, et al., 2011	3D Accelerometer	Leg	VO[2]	Abs	A	No	LB SB	-	-	-	-	1F 3C

| Source | Level I | | | Level II | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
Ito, et al., 2010	Data loggers	?	Ga	Reg	V	No	Alert	39-56	72-96	?	-503 +23	28F 1319C
Liberati and Zappavigna, 2009	Pedometer	Leg	?	FL	T	Yes	Alert, Pr	?	?	?	?	1F 107C
Mazrier, et al., 2006	Pedometer	Leg	CHE	Abs	T	No	Graph	?	?	+38	-	1F 12C
Miekley, et al., 2012	Pedometer	Leg	Tr	CU	V	No	Alert	50	80	-	-	1F
Higginson, et al., 2010	Pedometer	Leg	?	Cor	T	No	?	-	-	-	-	1F 16C
Kujala, et al., 2008	Four balance weighting floor	FLA	LS, HT	Stat	T	No	Alert	?	?	-211 +104	-	1F
Pastell and Kujala, 2007	Four balance system (four plates)	FLA	LS, HT	NN	V	No	Alert	100	57	-36 +9	-36 +9	1F
Pastell and Madsen, 2008	Four balance weighting floor	FLA	VO, HT	CU	-	No	Drift in data	?	?	-	-	1F
Pastell, et al., 2008a	Four strain gauge balances	FLA	VO		A	No	Graphs	?	?	-	-	1F
Pastell, et al., 2008b	Four strain gauge balances	FLA	LS	Abs	A	No	Graphs	-	-	-	-	1F
van der Tol and van der Kamp, 2010	Weigthing floor	FLA	EV	MA	V	No	Alert	51-74	94-99	?	?	1F
				MA				85	77			191C
Chapinal, et al., 2009	Weighting platform	Fl	No	No	-	No	Graph	-	-	-	-	1F 12C

Appendix 2.3: Continued

| Source | Level I | | | Level II | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
Chapinal, et al., 2010	Weighting platform	Fl	Ga, CHE	Reg	T	No	Alert	-	-	?	?	1F
Neveux, et al., 2006	Weighting platform	Fl	Ga, HT	No	A	No	No	-	-	-	-	1F
Pastell, et al., 2010	Weighting platform	Fl	Ga, HT	Reg	T	No	Ga	-	-	[3]	-	1F
Rajkondawar, et al., 2006	2 Parallel force plats	Fl	GLV	Reg	V	No	Ga, Pr	?	?	-	-	1F
Rajkondawar, et al., 2002	2 Parallel force plats	Fl	EV	No	A	No	Graph	-	-	?	?	6C
Tasch and Rajkondawar, 2004	2 Parallel force plats	Fl	?	No	A	No	Graphs	-	-	?	-	?
van der Tol, et al., 2002	Pressure distribution plate	Fl	No	No	-	No	Images	-	-	-	-	8C
van der Tol, et al., 2003	Pressure distribution plate and a force plate	Fl	No	No	-	No	Images	-	-	-	-	9C
Pastell, et al., 2008b	Elektromagnetic film	?	LS	Abs	A	No	Graphs	-	-	-	-	1F
Maertens, et al., 2011	Pressure sensitive mat	Fl	Ga	Reg	V	No	Ga	76-90	86-100	-	[4]	1F 12C

Appendix 2.3: Continued

	Level I		Level II									
Source	Sensor	Place	GS	Alg	Val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
Pluk, et al., 2010	Video camera	NTC	LS	Stat	T	No	?	-	-	-12 +3	-	2F 79C
Poursaberi, et al., 2010	Video camera	NTC	SLI	Reg	V	No	Ga	100	97.6	-	-130 +54	1F 94C
Song, et al., 2008	Video camera	NTC	LS	No	A	No	Graph	-	-	[5]	-	15C
Liberati and Zappavigna, 2009	Activometers	NC	?	FL	T	Yes	Alert, Pr	?	?	?	?	1F 40C

[1] Used abbreviations:

Sensor:

Place: Cow Torso (CT); Floor / passage way (Fl); Floor AMS (FLA); Neck Collar (NC); Next To Corridor (NTC);

GS: Clinical Hoof Examination (CHE); Examination by Veterinarian (EV); Hoof Trimming (HT); Locomotion Score (LS); Gait score (Ga); Gait and Lesion scoring by Veterinarian (GLV); Scoring of Lameness Indicators (SLI); Treatment (Tr);Visual Observation (VO);

Alg: Absolute value (Abs); Correlation (Cor); CUMSUM (CU); Fuzzy Logic (FL); Moving Average (MA); Neural Network (NN); Regression (Reg); Statistical test (Stat);

Val: Assumed relation between gold standard and sensor data (A); statistically Tested relation between gold standard and sensor data (T); Validated algorithm for detection (V);

Output: Gait score (Ga); Lying Bout (LB); Lying Frequency (LF); Probability (Pr); Standing Bout (SB); Step Count (StC);

[2] Visual Judgement of video recordings

[3] Gold Standard cases are gait score 2: 11, gait score 2.5: 9, gait score 3: 23,gait score 3.5: 7, gait score 4: 5.

[4] Gold Standard cases are gait score 1: 54, gait score 2: 55, gait score 3: 50.

[5] Gold Standard cases are gait score 1: 4, gait score 2: 8, gait score 3: 1, gait score 4: 2.

Appendix 2.4: Characteristics of sensors for automated detection of metabolic problems in dairy cows. Level I describes the sensor and its place in respect to the cow. Level II describes the data processing with the used gold standard (GS), used algorithm (Alg), the level of algorithm validation (val), whether the algorithm uses non-sensor data (non-sensor), the sensitivity (SE(%)) and specificity (SP(%)) of the algorithm, the numbers of gold standard positive (+) and negative (-) cases in the used test (T cases) and validation (V cases) datasets and the test size in number of farms (F) and number of cows (C). A '-' indicates that the characteristic was not found in the publication, a '?' indicates that the description of the characteristic was unclear in the publication.

| Source | Level I | | Level II | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
AlZahal, et al., 2011	Electrode	Rumen[1]	VT[2]	Stat	T	No	No	-	-	-	-	1F 8C
AlZahal, et al., 2009	Telemetric bolus	Rumen	No	No	A	No	No	-	-	-	-	1F 4C
Ipema, et al., 2008[3]	Temperature sensor	Rumen	No	No	-	No	No	-	-	-	-	1F 1C
AlZahal, et al., 2007a	?	Rumen	?	Reg	V	No	[4]	?	?	-	-	1F ?C
AlZahal, et al., 2007b	pH-electrode	Rumen[1]	No	No	A	No	No	-	-	-	-	1F 1C
AlZahal, et al., 2009	pH-electrode	Rumen[1]	No	No	A	No	No	-	-	-	-	1F 4C
Sato, et al 2012.	Glass electrode	Rumen (bolus)	RFS	Cor	A	No	No	-	-	-	-	1F 4C
Goense, et al., 2009[3]	pH-electrode	Rumen	No	No	-	No	No	-	-	-	-	No
Gasteiner, et al., 2009	pH-electrode	Rumen	SPS	Stat	A	No	No	-	-	-	-	1F 5C[5]
Ohtani, et al., 2005	Fiber-optic spectrophotometer	In-line	ORM	Reg	T	Yes	F%	-	-	-	-	?F 92C

Appendix 2.4: Continued

| Source | Level I | | Level II | | | | | | | | | | |
	Sensor	Place	GS	Alg	Val	Non-sensor	Output	SE(%)	SP(%)	T cases	V cases	Tested
de Roos, et al., 2007	Infrared spectroscopy	On-line[6]	CA	Cor	V	No	Alert	70	95	?	-1724 +356	217F 1085C
Mazeris, 2010	HN	On-line	?	?	?	Yes	RV	-	-	-	-	?
Vreeburg, 2010	HN	On-line	?	?	V	Yes	RV	?	?	-	-	2F 286C
Bar and Solomon, 2010	Rumination sensor	NC	VO	Reg	T	No	No	-	-	-	-	1F 75C
Halachmi, et al., 2009	TCM	APW	VO, US	Stat	T	No	BCS	-	-	-	-	1F 186C
Edwards and Tozer, 2004	Pedometer	Leg	JV	?	T	Yes	?	-	-	-947 +498	-	3F 1445C

[1] Rumen fistulated cows

[2] Used abbreviations:

Sensor: Herd Navigator® (HN); Thermal Camera with MatLab automated image analysis (TCM);

Place: Above Passage Way after milking parlour exit (APW); Neck Collar (NC);

GS: Chemical Analysis (CA); Judgement Veterinarian (JV); Official Reference Method in Japan (ORM); Rumen Fluid spot Sampling, handheld pH meter (RFS); Standard pH solutions (SPS); Ultra Sound (tissue thickness) (US); Vaginal Temperature (VT); Visual Observation (VO);

Alg: Statistical test (Stat);

Val: Assumed relation between gold standard and sensor data (A); statistically Tested relation between gold standard and sensor data (T); Validated algorithm for detection (V);

Output: Risk Values (RV);

[3] Test of technical feasibility of measuring pH In-Cow.

[4] Accumulated time below pH cut of points

[5] Tested in five steers.

[6] In-line could be a future possibility.

Chapter 3

Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows.

C.J. Rutten

C. Kamphuis

H. Hogeveen

K. Huijps

M. Nielen

W. Steeneveld

ABSTRACT

Management during calving is important for the health and survival of dairy cows and their calves. Although the expected calving date is known, this information is imprecise and farmers still have to check a cow regularly to identify when it starts calving. A sensor system that predicts the moment of calving could help farmers efficiently check cows for calving. Observation of a cow prior to calving is important because *dystocia* can occur, which requires timely intervention to mitigate adverse effects on both cow and calf. In this study, 400 cows on a Dutch dairy farm were equipped with sensors. The sensor was a single device in an ear tag, which synthesised cumulative activity, rumination activity, feeding activity, and temperature on an hourly basis. Data were collected during a one-year period. During this period, the exact moment of 417 calvings was recorded using camera images of the calving pen taken every five minutes. In total, 114 calving moments could be linked with sensor data. The moment at which calving started was defined as the first camera snapshot with visible evidence that the cow was having contractions or had started labour. Two logit models were developed: a model with the expected calving date as independent variable and a model with additional independent variables based on sensor data. The areas under the curves of the Receiver Operating Characteristic were 0.885 and 0.929 for these models, respectively. The model with expected calving date only had a sensitivity of 9.1%, whereas the model with additional sensor data has a sensitivity of 36.4%, both with a fixed false positive rate of 1%. Results indicate that the inclusion of sensor data improves the prediction of the start of calving; therefore the sensor data has value for the prediction of the moment of calving. The model with the expected calving date and sensor data had a sensitivity of 21.2% at an one-hour time window and 42.4% at a three-hour time window, both with a false positive rate of 1%. This indicates that prediction of the specific hour in which calving started was not possible with a high accuracy. The inclusion of sensor data improves the accuracy of a prediction of the start of calving, compared to a prediction based only on the expected calving date. Farmers can use the alerts of the predictive model as an indication that cows should be supervised more closely in the next hours.

INTRODUCTION

Up to one-third of calves born on dairy farms are born after *dystocia* and have increased risks of disease and mortality *(Barrier, et al., 2013)*. Severe *dystocia* causes stillbirth in 49% of cases and calves born after *dystocia* are 1.5 times more likely to develop a disease during the first 120 days of age *(Lombard, et al., 2007)*. For cows, the likelihood of conception decreases as the number of days open increases *(Fourichon, et al., 2000)*, and culling risk is higher *(Rajala-Schultz and Grohn, 1999)* within a lactation that starts with a *dystotic* calving. Moreover, *dystocia* increases the risk of damage to the uterus and infections, which increases the risk of metritis *(Rajala-Schultz and Grohn, 1999, Schuenemann, et al., 2013, Sheldon, et al., 2009)*. *Dystocia* is therefore a health and welfare problem for both cows and calves. High calf mortality can also be seen as an image problem for the whole dairy sector.

Risk factors for *dystocia* include biology of the cow (e.g. breed and parity), calf gender *(Norman, et al., 2010)* calf weight and management (e.g., housing and pre-calving movement) *(Mee, et al., 2014, Piwczynski, et al., 2013)*. Farmers can influence these risk factors through management, for instance, by changing their breeding strategy but also by human supervision during the calving process. Lombard et al. *(2007)* observed that 24% of stillbirths occurred with unassisted calvings. Supervision during the calving process, which enables appropriate intervention, is therefore likely to reduce the number of stillbirths and other health and welfare effects that *dystocia* has on calves and dairy cows *(Barrier, et al., 2013, Mee, et al., 2014)*.

Farmers currently only have the expected calving date on which to base the decision to supervise cows more intensively. The true calving date varies between 267-295 days after a successful insemination *(Inchaisri, et al., 2010a)*, whereas the expected calving date is on average 280 days post insemination. Hence it is challenging for farmers to correctly determine which cows should be supervised more often or more intensively, and when appropriate interventions are needed. Farmers thus have to visually check pregnant cows that approach their expected calving date and this increases the work load for a farmer.

There are several behavioural and physiological parameters are associated with the start of calving, that can be monitored automatically by sensors. Feeding and ruminating behaviour of dairy cows decreases gradually in the last two weeks before calving and drops suddenly at calving *(Bar and Solomon, 2010)*. Sensors seem capable of detecting these changes *(Bar and Solomon, 2010, Bucher and Sundrum, 2014, Schirmann, et al., 2013)*. Time spent on feeding also decreases, dry matter intake tends to decrease slightly *(Schirmann, et al., 2013);(Bucher and Sundrum, 2014)*, and activity changes in the 24 hours before calving *(Clark, et al., 2015, Miedema, et al., 2011b, Saint-Dizier and Chastant-Maillard, 2015)*. Titler et al. *(Titler, et al., 2015)* demonstrated that an activity index could be used to predict whether a cow would calve in the six hours following an increase in the activity index. Previous studies have shown

that temperature (measured at the vulva, rectum, and rumen) decreases during the 24 hours prior to calving *(Saint-Dizier and Chastant-Maillard, 2015)*. Oeullet et al. *(2016)* have shown that all these parameters, which can be measured by sensors have value for the prediction of calving.

A more accurate prediction of the start of calving than the expected calving date would enable farmers to identify when a cow requires intensive supervision. This will help ensure appropriate intervention when needed and reduce the workload for the farmer from unnecessarily checking cows. Although studies have shown that sensor data has value for the prediction of calving, an independent validation of the accuracy of such a prediction has not been studied yet. Furthermore, an evaluation of the additional value of sensor data compared to the expected calving date is also missing in the literature. In this study, rumination, activity, and temperature measured automatically by a single sensor are used to predict the start of calving in dairy cows by (1) evaluating at which moment, relative to the start of calving, sensor data has predictive value, (2) exploring the potential value of sensor data in addition to the expected calving date in estimating the start of calving, and (3) developing an independently validated model that predicts the start of calving.

MATERIAL AND METHODS

Gold standard definition

The definition of the start of the calving process is essential for the development of a model that predicts the calving moment. The moment of actual calving is not informative for a farmer, as potential *dystocia* should be detected and resolved shortly after the start of calving. The start of the calving process is therefore a better moment to generate an alert for calving. This study defined the start of the calving process as the first camera snapshot with visible evidence that the cow was having contractions or had started labour. When a born calf was seen on camera the start of calving could be deduced by scrolling back in time. The moment as defined in the current study refers to the start of the second stage of parturition were the foetus is expelled *(Parkinson, et al., 2001b)*. The most notable signs are visible abdominal muscle contraction and movement of ears and head that indicate pressure to expel the foetus. Typically the cow is lying down on her side (lateral recumbency), but standing upright is possible. Date and time of this camera snapshot were used as the gold standard for the start of the calving process, defined at the respective hour.

Data collection

On a commercial Dutch dairy farm, 400 cows were equipped with Agis SensOor sensors (Agis Automatisering B.V., Harmelen, the Netherlands). These sensors are 3D-accelerometers attached to the ear tag of the cow and report rumination, feeding, activity, and temperature on an hourly basis *(Bikker, et al., 2014)*. Data were collected from September 1, 2013 until November 1, 2014 from late gestation dairy cows housed in a straw bedded pen.

The dairy farmer was asked to record the date and time at which he had noticed a cow had calved. The start of the calving process as defined for this study was assigned by manual evaluation of snapshot images taken by a video camera every five minutes. The farmer-recorded estimates of the calving moment were used to reduce the amount of images that were screened. Animal husbandry students (BSc, van Hall-Larenstein, Leeuwarden, the Netherlands) were instructed to use the camera images to determine the exact start of the calving process for each cow. In total, 414 cows calved; exact calving moments were determined for 240 of these cows by screening images. Of these 240 calving moments, 90 belonged to heifers. The farmer only equipped these heifers with sensors post-partum as part of normal management procedure. Consequently, these 90 calving moments had no sensor data available. The remaining 150 calving moments had sensor data available and were used for further analysis.

Expected calving date

Insemination records were used to calculate the expected calving date at 280 days post insemination for each cow. Expected calving dates were required to fall within a period from three weeks before to three weeks after the actual calving date. This method was based on the generally accepted average gestation length of 280 days *(Parkinson, et al., 2001c)* in combination with the three week interval for ovulation *(Parkinson, et al., 2001a)*. If an expected calving date did not fall within this six-week period, it was assumed that the insemination did not lead to a calving and the expected calving date was therefore assumed missing. If an expected calving date fell within this six-week period, the expected calving date was used to estimate the number of days to expected calving date (DTC). This variable is negative in the days prior to the expected calving date and zero at the expected calving date.

Sensor data

For each hour of the day, the SensOor system assigns the minutes within that hour to one of the five following sensor parameters (Var_i): ruminating ($i=1$), eating ($i=2$), active ($i=3$), highly active ($i=4$), or not active ($i=5$). These five sensor parameters are measured by a

single sensor. The sum of the five sensor parameters adds up to a total of one hour. This means, for instance, that one minute spent on rumination cannot be spent on being active. Therefore, these five sensor parameters are not fully independent. In addition to these five sensor parameters, ear temperature ($i = 6$) was measured and recorded for each hour of the day.

Missing sensor data. The dataset contained hourly blocks with missing values for some of the five sensor parameters or ear temperature, and some hourly blocks were missing entirely. The analysis required a continues series of hourly blocks over time. Therefore, missing hourly blocks were added and missing values within hourly blocks were imputed. For practical application of a sensor system, a straightforward imputation algorithm that only uses data from preceding hourly blocks was considered most appropriate. The behaviour of cows (i.e., ruminating, walking around, and lying down) was assumed to show a diurnal pattern as described previously *(Roelofs, et al., 2005)*. Therefore, it was assumed that a reasonable imputation could be achieved by substituting the missing data with the average of data for the same hourly block from the previous three days. A weighted average was calculated; data closer to the hourly block with missing data received more weight than older data. If data were unavailable for any of the three days, only the available days were used. In total 470 hourly blocks for 45 calvings were imputed by the described methodology, the number of imputations ranged from 1 to 73 hourly blocks per calving. If no data were available for all of the three days, the calving was excluded entirely from further analysis (n=36). The final dataset for further analysis contained 114 calvings.

Independent variables from sensor data. The behaviour of the cow was assumed to change before the start of calving. Therefore, the change over time (t, where $t = 0$ represents the start of calving) in sensor parameters was estimated and used as an independent variable for the development of a model. The first step to develop independent variables from sensor data was the calculation of a rolling mean. For all sensor parameters, a rolling mean was calculated over the 72 hours preceding the start of calving ($rollVar_i^t$) (equation 1).

$$rollVar_i^t = \frac{\sum_{t=1}^{-72} Var_i^t}{72}$$

(1)

The second step was the estimation of the change over time in the sensor data (Δ in sensor data). The i independent variables at moment t (X_i^t) were calculated by estimating the deviation of the observation at moment t from the rolling mean of the preceding 72 hours ($t - 72$) (equation 2).

$$X_i^t = Var_i^t - rollVar_i^{t-72}$$
(2)

For example, if a cow is ruminating 23 minutes on the current hourly block and the 72 hr rolling mean of 3 days earlier was 30 minutes the Δ in sensor data is -7 min/hr and the independent variable was assigned the value -7.

Data selection

The starting moments of calving were combined with sensor data so that the start of calving was connected to the hourly block of sensor data in which the calving started. For each calving moment, data from the 24-hour period ending with the hour in which calving started were selected as a case dataset. Each case dataset was matched randomly to three control datasets. Control datasets were selected from the same cow and ended at the same hour of the day as the case dataset. The control datasets were sampled from the period between three weeks and one week before the start of calving. The case and control datasets resulted in the case-control dataset, which contained four records (one case and three controls) for each calving moment. Each record contained 24 variables for each sensor parameter, which correspond to the Δ in sensor data (X_i^t) for each hour of the 24-hour period prior to the start of calving. These case and control datasets contained 34 hourly blocks for 9 calvings which were imputed as described in section 2.4.1, the number of imputations ranged from 2 to 10 hourly blocks per calving.

The mean values per hour across all cows in the dataset and their 95% confidence intervals were calculated for all five sensor parameters in both the case and control datasets. These values were plotted and used to visually examine whether the case dataset differed on average from the control dataset. Sensor parameters for which the confidence intervals around the average value over all cows of the case and control datasets did not overlap were included for further analysis.

The available calvings were then split in a training dataset (two thirds of the available calvings, i.e. 79 calvings) and a testing dataset (35 calvings). The training dataset contained 316 records (i.e. four records for each of the 79 calvings). The training dataset was selected by randomly sampling cows from the case-control dataset; for each randomly selected case dataset, the corresponding control datasets were also included in the training dataset. It was not possible for a cow to be present in both the testing and training dataset because each cow calved only once during the period of data collection.

Model development

Two models were developed, a model with "DTC" as the only independent variable (model DTC) and a model with "DTC" and sensor variables as independent variables (model DTC + sensor). For the model "DTC + sensor", sensor variables with predictive value for the start of calving were selected using a stepwise selection procedure.

Logistic regression models. Logistic regression was chosen as the method to estimate a model that predicts the start of calving. The dependent variable was the binary variable "start of calving" (1 = calving started and 0 = calving did not start). The prediction of the resulting logit model ranges between 0 and 1 and can be interpreted as the probability that calving will start. The general logit model is described by Equation 3. In Equation 3 $\beta_{\text{int } ercept}$ represents the estimate of the model intercept, $\beta_{\text{int } ercept}$ the ith parameter estimate for $\beta_{\text{int } ercept}$ the ith model parameter and moment t.

$$p_t = \frac{1}{1 + e^{-(\beta_{\text{int} ercept} + \beta_1 * X_1^t + \beta_2 * X_2^t + ... + \beta_i * X_i^t)}} \tag{3}$$

Independent variable selection using individual sensor parameters. First each of the five sensor parameters and ear temperature were used individually in the variable selection using individual sensor parameters (**ISP**). The Δ in sensor data (X_i^t) for each of the 24 hourly blocks in the training dataset was used as an independent variable for the logit model in the ISP. For each sensor parameter, the model contained the independent variable "DTC" and an additional 24 independent variables for each Δ in sensor data, ranging from 23 hours before the hour in which calving started up to and including the hour in which calving started. A logit model was fitted for each sensor parameter and the combination of independent variables (each independent variable corresponded to an hourly block ranging from -23 to 0 hours antepartum) with the lowest Akaike's Information Criterion, corrected for small sample sizes (**AICc**), was selected in a stepwise selection procedure. Stepwise selection uses a combination of forward addition and backward elimination to select independent variables *(Calcagno, 2013)*. In each step the respective AICc was determined and the model with the lowest AICc was selected, the AICc of the selected model is reported.

Independent variable selection using combined sensor parameters (CSP). The independent variables selected in the ISP selection for all Δ in sensor data, were combined in a single model for further selection using combined sensor parameters (**CSP***)*, applying the same stepwise selection procedure as used in the ISP selection. The CSP selection resulted in the model "DTC + sensor". The performance of this model was evaluated further. The model "DTC + sensor" was used to predict the probability that calving will start for the test dataset. The AICc's for both the models selected in ISP and CSP were reported.

Model evaluation

The test dataset used for model evaluation contained all hourly blocks from the last three weeks before calving. For the evaluation of the predictive performance of the two models (model "DTC" and model "DTC + sensor"), the binary variable "start of calving" and the model predictions from the test dataset were used to generate curves of the Receiver Operator Characteristic (ROC curves) and to estimate the area under the ROC curve (AUC). Based on the ROC curves, a threshold for the probability that calving started was chosen that resembled a false positive rate of 1%. Thresholds were subsequently used to generate an alert for an hourly block if the probability for that hourly block exceeded the chosen threshold. These alerts were then classified as true positive, false positive, true negative, or false negative relative to the gold standard. Three different evaluation schemes were used. A graphical description of these evaluation schemes is presented in Figure 1.

The first evaluation scheme (Scheme 1, Figure 1) was based on a day. This scheme was used to compare the models "DTC" and "DTC + sensor" and to explore the additional value of sensor data. Calving alerts were generated for each hourly block. Hourly blocks were then aggregated into 24-hour blocks, defined so that the hour in which calving started was always the last hour of a 24-hour block. For each day and for each individual calving moment in the dataset, it was determined whether the model generated an alert for a cow. Each day was then classified as either true positive, false positive, true negative, or false negative, assuming the day of calving as gold standard. Hence, for each cow only one alert per 24 hour period was evaluated. Based on this classification the sensitivity and specificity at daily level were estimated.

The second and third evaluation schemes (Scheme 2 and 3, Figure 1) used an hourly basis. Alerts were generated for each hourly block. Hence for each cow 24 alerts were evaluated per 24 hour period. The second evaluation scheme used a strict time window of one hour, which means that generated alerts were classified as true positives only for the hour in which calving started. All other alerts were classified as false positives. This scheme was used to evaluate how accurately the start of calving could be predicted. For the third evaluation scheme a broader, three-hour time window was used for evaluation; generated alerts were classified as true positives for the hour in which calving started and for the preceding two hours. If more than one alert was given in this three-hour block, the alerts were considered to be a single true positive alert. False alerts were not merged, so each hourly alert outside this three-hour time window was considered to be a single false positive alert. In addition to the third scheme also time windows of 6 (scheme 4) and 12 hours (scheme 5) were used based on a similar approach to the third scheme. When a broader time window was used the number of alerts changed. For instance, two alerts within the broader time window are considered as one true positive alert. Alerts classified as true positive in a broader time window, were classified as false positive in a smaller time window. The schemes 3, 4 and 5 were used to evaluate by how much detection performance would increase when a less

precise prediction was accepted and to discuss what precision would be achievable and desirable in practice.
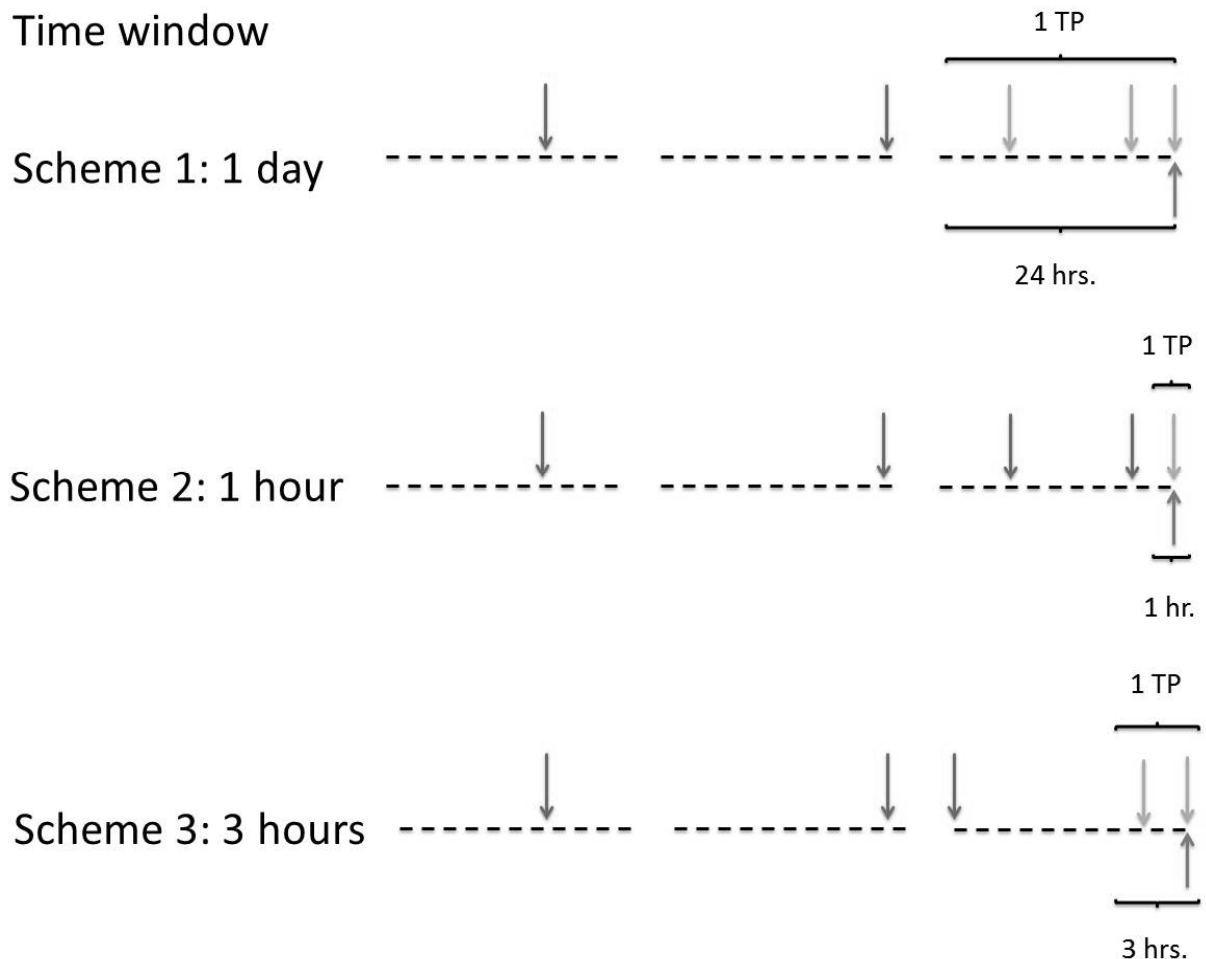

Figure 1: Schematic description of the evaluation schemes used. The schemes are on a daily, hourly, and three-hour basis. Dashed lines represent a period of 24 hours, arrows below the line represent a calving and arrows above the line represent alerts. Alerts classified as true positive are indicated with "TP", all other alerts were classified as false positive "FP" alerts. In this example 5 alerts were generated and classified using different schemes, that results in different numbers of TP and FP.

Statistical package

All data editing and analyses were done in R 3.0.2 *(R Development Core Team, 2008)* with the add-on packages dplyr 0.2 *(Wickham and Francois, 2014)*, Zoo 1.7-10 *(Zeileis and Grothendieck, 2005)*, Glmulti 1.0.7 *(Calcagno, 2013)*, and ROCR 1.0-5 *(Sing, et al., 2005)*.

The recommendations on transparent reporting of predictive models in human medicine were considered in this study and the so called TRIPOD statement *(Moons, et al., 2015)* is provided in an appendix (3.1).

RESULTS

Figure 2 shows the mean sensor values for the 24 hours before the hour in which calving started for the case (n=114, solid line) and control datasets (n=342; dashed line), with their respective 95% confidence intervals. The confidence intervals for the case and control datasets did not overlap for the sensor parameters "activity" and "highly active"; whereas the confidence intervals partially overlapped for "ruminating" and "temperature". Confidence intervals overlapped during the entire 24-hour period for "not active" and "feeding". Overlapping confidence intervals indicate that mean values between the case and control datasets were not different given the sample size. Therefore, only "activity", "highly active", "ruminating", and "temperature" were reparametrized in Δ and used in the ISP and CSP selection procedures.

For the models in the ISP selection, the AICc's were: 127.44 for "activity", 136.88 for "temperature", 137.32 for "ruminating", 163.31 for "highly active", and 165.70 for "DTC". The independent variable "DTC" was selected in all models in the ISP analyses. In addition, independent variables with the Δ in sensor data were selected for "activity" (-18, -7, -6, and -2 hours relative to the start of calving), "temperature" (-23, -17, and 0 hours relative to the start of calving), "ruminating" (-13, -12, -7, -4, -3, -2, and -1 hours relative to the start of calving), and "highly active" (-18, -14, -13, -10, -6, and -2 hours relative to the start of calving).

The final 'DTC + sensor' model resulting from the CSP selection procedure had an AICc of 70.40, lower than those found in the ISP selections. Table 1 summarizes the parameter estimates for this model. The remaining independent variables after CSP were "activity" (-6 and -2 hours relative to the start of calving), "temperature" (-17 and 0 relative to the start of calving), "ruminating" (-12, -7, -3, -2, and -1 hours relative to the start of calving), and "highly active" (-18, -14, -13, -10, -6, and -2 hours relative to the start of calving).
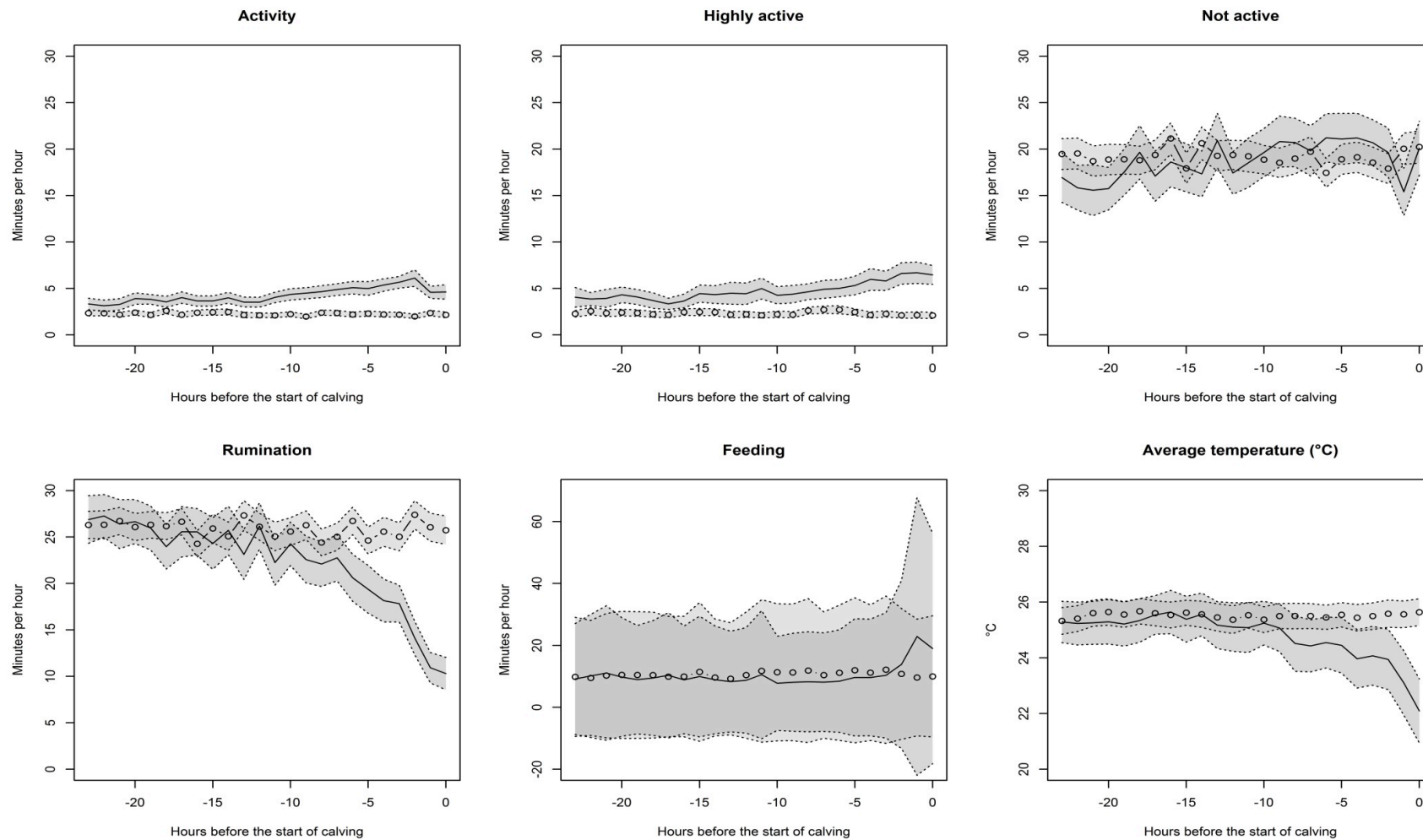
Figure 2: Graphical description of mean values over all cows and corresponding 95% confidence intervals (dashed line) for the sensor data of the six parameters in the sensor dataset over a period of 24 hours. The case datasets, n=114 (solid line), end with the start of calving at hour 0. The corresponding control datasets, n=342 (dots contected by a solid line) contain data from a random day of the dry-off period of the same cow, ending at the same hour of the day in which calving started in the case dataset.

Figure 3 plots the ROC curves for the "DTC" (AUC = 0.885, dashed line) and the "DTC + sensor" (AUC = 0.929, solid line) models. The AUC increased by 0.044 when Δ in sensor data was included in addition to the expected calving date ("DTC"), which is the current information available to dairy farmers.

Table 2 summarizes performance indicators for the "DTC" and "DTC + sensor" models. Adding sensor data to the model increased sensitivity from 9.1% (model "DTC") to 36.4% (model "DTC + sensor") when evaluated on a daily basis. Evaluating the "DTC + sensor" model on an hourly basis, resulted in a sensitivity of 21.2% (Scheme 2; Figure 1). Note that in the daily basis for each cow one alert in 24 hours and in the hourly basis 24 alerts in 24 hours were evaluated. Extending the time window for the hourly basis to three hours increased sensitivity to 42.3% (Scheme 3; Figure 1). Broader time windows of 6 and 12 hours increased the sensitivity even more. The number of false positive alerts ranged from 110-148 alerts within the last 3 weeks antepartum depending on the used time window.

Table 1. The multivariable logistic regression model for the prediction of the start of calving, including sensor data and days to expected calving date (DCT), model "DTC + sensor": parameter estimates with their respective standard errors (S.E.), test statistics (Z-value), and p-values. The variable "days to expected calving date" is the number of days until the expected calving date. The other variables are derived from sensor data as the Δ (min/hr or °C) relative to a 72 hr rolling mean from 3 days ago.

| Parameter	Hours before calving started	Estimate	S.E.	Z-value	P-value
Intercept		-2.400	1.054	-2.276	0.024
DCT (days)		0.554	0.062	8.922	0.000
Rumination (min/hr)	-13	-0.522	0.127	-4.104	0.000
	-12	0.449	0.108	4.172	0.000
	-7	0.288	0.098	2.934	0.004
	-3	-0.276	0.106	-2.607	0.010
	-2	-0.096	0.110	-0.877	0.381
	-1	-0.676	0.154	-4.395	0.000
Temperature (°C)	-17	0.508	0.211	2.415	0.016
	0	-1.341	0.247	-5.422	0.000
Activity (min/hr)	-6	1.466	0.280	5.228	0.000
	-2	1.247	0.271	4.595	0.000
Highly active (min/hr)	-18	-0.764	0.358	-2.136	0.033
	-14	-0.151	0.294	-0.515	0.607
	-10	1.512	0.304	4.972	0.000
	-6	-2.031	0.545	-3.728	0.000
	-2	0.575	0.347	1.659	0.098
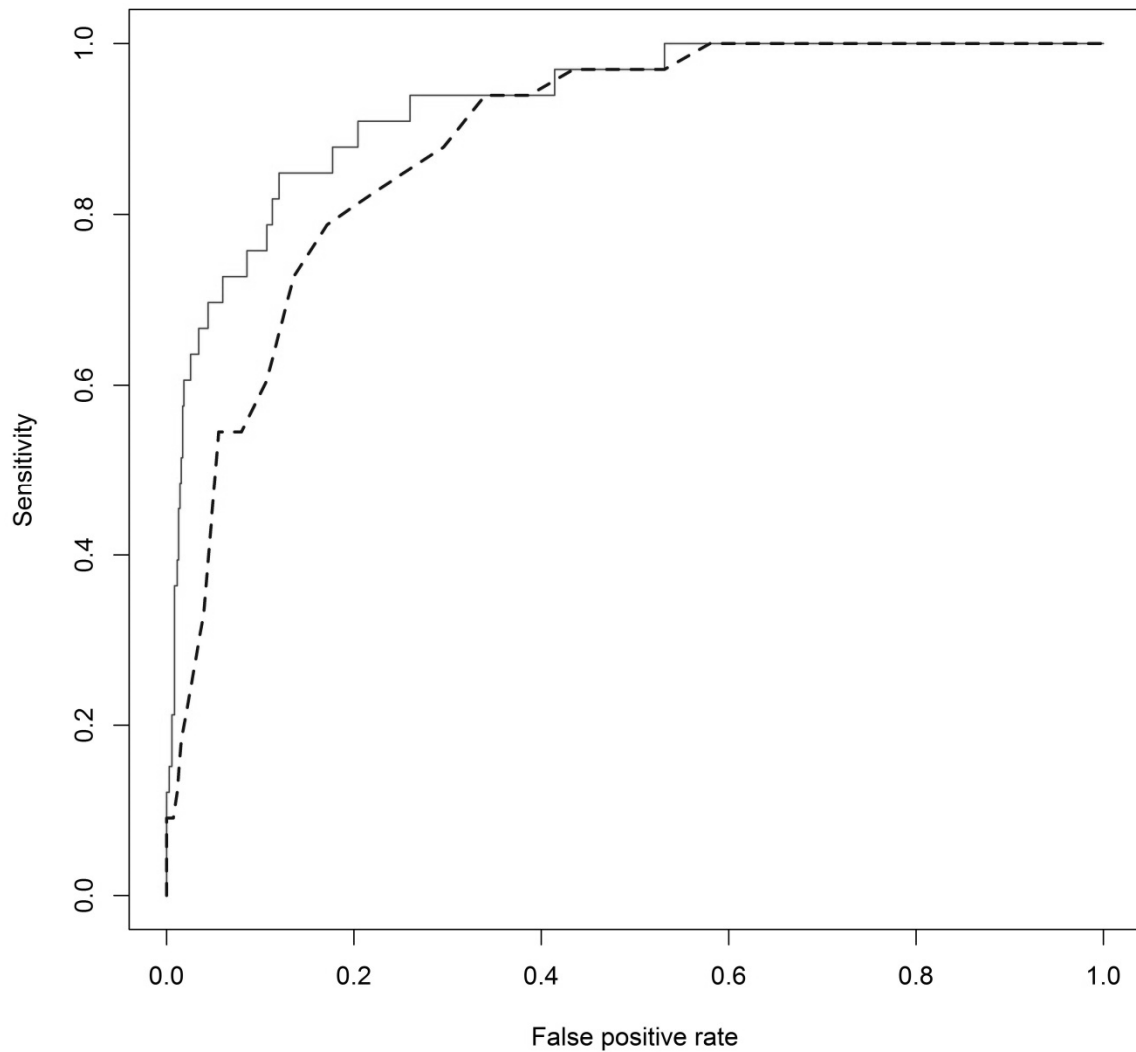
Figure 3: Curves of the receiver operating characteristic for model "DTC", which has expected calving date as independent variable (dashed line), and model "DTC + sensor", which has sensor data and the expected calving date as independent variables (solid line). Both models were used to predict the day on which calving started. The respective areas under the curve (AUC) were 0.885 for "DTC" and 0.929 for "DTC + sensor".

Table 2. Predictive performance of the models with days to the expected calving date only (DTC) and with sensor data in addition to "DTC" (DTC + sensor): sensitivity, specificity and area under the curve (AUC) of the receiver operating characteristic. The corresponding numbers of true positive (TP) and false positive (FP) alerts are presented as well. Both models were evaluated on a daily basis. Model "DTC + sensor" was further evaluated on an hourly basis with different time windows. Model evaluation were conducted on an independent test dataset with the start of calving and data of 3 weeks prior to the start of calving.

	AUC	Sensitivity (%)	TP	Specificity (%)	FP
Evaluation on daily basis					
DTC	0.885	9.1	3	99.3	5
DTC + Sensor	0.929	36.4	12	98.9	8
Evaluation of model DTC + sensor on hourly basis					
1-hour time window	0.901	21.2	7	99.1	148
3-hour time window	0.901	42.4	14	99.2	135
6-hour time window	0.901	48.5	16	99.3	124
12 hour time window	0.901	51.5	17	99.4	110

The highest number of alerts were generated in the last 12 hours before calving started; 28% of all 134 alerts generated during the last week before calving were generated in the last 12 hours. Alerts were generated throughout the entire week, but less than during the last 12 hours before calving.

DISCUSSION

In this study, sensor parameters "activity", "temperature", "rumination" and "highly active" changed in the 24 hours prior to the start of calving. From a model selection process using the AICc, the independent variable "DTC" was selected together with independent variables with the Δ in sensor data primarily from the last 10 hours prior to the start of calving. "Activity" contributed the most to the model and "highly active" contributed the least to the model, in terms of AICc. Combining sensor parameters in the CSP analysis resulted in the lowest AICc, indicating that the "DTC + sensor" model has more predictive value than models in the ISP analysis, which used single sensor parameters. Model "DTC + sensor" has a lower AICc, a higher AUC, and a higher sensitivity (at a comparable specificity) than the "DTC" model, when evaluated on a daily basis (scheme 1; Figure 1). The sensitivity of the "DTC + sensor" model increased around 20 percentage points at a specificity of approximately 99%, when evaluated on an hourly basis using a three-hour time window. The model performance in this study, as measured in AUC values on a separate test dataset (Table 2), was higher

than previous findings of Ouellet et al. *(2016)*. This was unexpected because the latter study lacked an independent validation and had a smaller dataset (32 cows) *(Ouellet, et al., 2016)*.

An increase in "activity" and "highly active" (Figure 1) was observed in this study. Two studies that cross validated the detection performance for calving detection report sensitivities around 80% and specificities around 90% *(Borchers, et al., 2015, Rutten, et al., 2015)*. These findings seem to indicate a better performance than the current findings. It should be noted that these studies generated alerts every two hours instead of every hour and that reduces the total number of alerts including the false positive alerts. Other studies have also reported activity to increase during the 24 hours before calving *(Clark, et al., 2015, Miedema, et al., 2011b, Saint-Dizier and Chastant-Maillard, 2015, Titler, et al., 2015)*. For "temperature", a decrease of about 3°C was observed during the hour in which calving started, whereas other studies have reported decreases in the range of 0.2-0.5 °C *(Burfeind, et al., 2011, Ouellet, et al., 2016)*. The sensor in the current study measured ear temperature, which is more sensitive to environmental influences than core body temperature *(Gonzalezjimenez and Blaxter, 1962)*. This might explain the differences between this study and values found in literature. However, "temperature" had the second highest impact (based on AICc) in the current study, indicating that "temperature" had predictive value relative to the other sensor parameters, regardless of possible environmental influences on the temperature measurements.

A decrease in "rumination" prior to calving was observed in this study, consistent with other studies *(Bar and Solomon, 2010, Bucher and Sundrum, 2014, Clark, et al., 2015, Ouellet, et al., 2016, Pahl, et al., Saint-Dizier and Chastant-Maillard, 2015, Schirmann, et al., 2013)*. In the current study, this decrease was about 15 minutes per hour (the difference between the case and control dataset during the hour in which calving started). Clark et al. *(2015)* reported a reduction in rumination of five minutes per hour, whereas Schirmann et al. *(2013)* reported a reduction of 30 minutes per two hours during the last 24 hours before calving. These studies reported their findings on a daily or two-hour basis, which makes comparison difficult. Pahl et al. *(2014)* observed that the decrease in rumination was most notable in the last two hours before calving and that rumination time varied considerably amongst cows, these results seem consistent with the observations of the current study.

Previous studies have observed changes in feeding and resting behaviour determined by sensors in the last 24 hours before calving *(Bucher and Sundrum, 2014, Clark, et al., 2015, Schirmann, et al., 2013, Titler, et al., 2015)*. No such effects were observed in the data of the sensor parameters "feeding" and "not active" in the current study. A possible explanation for the difference with previous studies is that the SensOor sensor used ear movements to determine feeding and resting behaviour. Resting behaviour and activity were described as minutes per hour spent on this behaviour. However, in late gestation cows restlessness and alteration between standing and lying are an important indication that calving is to start, i.e. she is searching for a place to lie down *(Miedema, et al., 2011b)*. The current algorithm of

SenOoor does not pick up the change between lying and standing. Furthermore, when a cow is lying down this time could be assigned to "ruminating" or "not active". On the other hand cows can also be ruminating or not active while standing upright. If lying time or the change between lying and standing could be determined from the raw data of SensOor it could be possible to improve the prediction of calving.

The current study focussed on the specific hour at which calving started. The choice to develop a predictive model for this moment was based on the idea that it is important that a farmer is present shortly after calving has started. A farmer`s presence is important because farmers need to be aware of *dystocia* as soon as possible to mitigate adverse effects *(Barrier, et al., 2013, Mee, et al., 2014)*. It is also important that a calf is fed colostrum shortly after birth, preferably within one to two hours *(Conneely, et al., 2014, Klein-Jobstl, et al., 2014)*. As only 21% of the calvings could be detected exactly, a less precise alert that requires the farmer to check pregnant cows more frequently might be of more practical use. With a broader time window, it was possible to predict 43.5% of the calvings within a three-hour time window at a specificity of 99%. This suggests that for practical applications, missing the start of calvings should be balanced against accepting more (false) alerts.

Alerts were mainly generated during the 12 hours before calving. In this period, changes in the sensor data were visible when the case datasets were compared to the control datasets (Figure 2). The predictive models appear capable of detecting behavioural changes associated with calving in the hours before calving starts. Therefore, it seems that sensors pick up behavioural changes that gradually develop in the hours before calving. Due to this gradual process, however, there is no sudden change in behaviour in the hour in which calving starts. This means that a specific alert for the start of calving is not yet feasible.

A more relaxed time window in which alerts are considered true positive improved predictive performance (evaluation based on Scheme 3 versus Scheme 2; Figure 1), because most false positive hourly alerts were generated in the 12 hours before the start of calving. Classifying alerts one to three hours before the start of calving as true positive could be reasonable, because these alerts can be seen as an indicator that calving is about to start. Alerts given three or more hours before the start of calving may be too early. Whether alerts are regarded as false or true positive alerts will depend on the preferences and attitude of the farmer.

Another way to increase sensitivity is to use a lower threshold for the probability that calving starts. As more alerts will then be generated, sensitivity will increase. This increase however, will be accompanied by more false positive alerts. Figure 3 shows a steep line in the ROC curve for a false positive rate of up to 10% for model "DTC + sensor". This indicates that a small increase in the false positive rate is associated with a large increase in sensitivity. Previous studies indicated that farmers prefer alerts close to an event and that false positive alerts reduced farmers' faith in automated mastitis detection systems of automated milking systems *(Hogeveen, et al., 2010, Mollenhorst, et al., 2012)*. Although exact preferences will

differ for the automated prediction of the start of calving, farmers may also prefer the fewest false positive alerts possible, similar to automated mastitis detection systems.

Although sensors detect behavioural and physiological changes that are related to calving and can be used to generate alerts that could have practical relevance, a predictive model that specifically predicts the exact moment at which calving starts was not feasible in the current study. Mainly because sensitivity was only 21.2% for exact hourly prediction. An alert that indicates which cows a farmer should supervise more closely in the next few hours seems feasible. The sensitivity was higher than 50% for a time window of 6 or 12 hours. For practical application, such an alert could be useful to detect cows that should be checked visually in the coming hours. This application is valuable for farmers, as it provides a reminder of cows that are close to parturition and also indicates cows that might otherwise calve unexpectedly, i.e. based only on the expected calving date. The most optimal time window is debatable. It is important to consider how a farmer could use the alerts to organise his labor around calving management. Twelve hours could be too long, as calving may start at the end of this twelve hour period and the other hours could be used for other tasks on the farm. Six hours may be a reasonable compromise between sensitivity and practical use for organizing labor. Furthermore, six hours would cover most of the night period in which a farmer would be absent from the barn. So, a model that predicts whether a cow will start calving in the coming six hours seems feasible although some refinement would be needed. Therefore, future research could focus on how the number of alerts and the time period in which these are generated can be used to identify cows that will calve in the next few hours. Other sensors have been shown to measure comparable changes associated with calving, so using different sensors to measure activity, rumination, or temperature is unlikely to greatly improve the specific prediction of the start of calving. Sensors that measure behavioural or physiological changes more directly related to calving might have additional value for the specific prediction of the start of calving. Such non-invasive sensors include: heart rate monitors (currently used in respiration studies *(Machado, et al., 2016)*), sensors that measure muscle contractions, sensors that monitor the standing and lying pattern *(Nielsen, et al., 2010)*, and biosensors that measure hormone (residue) levels (currently used for measurements in milk *(Brandt, et al., 2010)* or detection of pathogens *(Casalinuovo, et al., 2006)*). Future research should focus on a combination of more sensor variables than the current study did, as adding other variables from different sensors may improve the prediction of calving. It might be interesting to study whether a prediction model for calving could also distinguish cows who will have *dystocia* from cows who will not. However possibilities for such an application based on behavioural parameters may be limited, as no behavioural differences have been observed between cows who experience *dystocia* and cows who do not *(Miedema, et al., 2011a)*. On the other hand heifers are known to have a higher risk of dystocia *(Mee, 2008)*, therefore future research should include heifers.

CONCLUSIONS

This study shows that sensor data can be of added value for a more accurate prediction of the start of calving than the expected calving date alone. However, the number of false positive alerts was relatively high and at best, the moment at which calving started was correctly predicted for fewer than half the calvings. False positive alerts were mainly observed in the last 12 hours before calving started. In this period, sensor data differed between the case and control datasets, but these data were not specific enough to be used for prediction of the exact hour in which calving starts. In practice, sensor data may still have merit for the prediction of calving as most alerts are generated within 12 hours before calving started. A cow receiving multiple alerts within a few hours provides an indication to the farmer that the cow should be supervised more closely. Therefore, a model that predicts whether a cow will calve within the next six hours seems feasible and reasonable for application in practice.

ACKNOWLEDGEMENTS

TRIPOD

APPENDIX 3.1

Section/Topic			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Page 1, line 1-2
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Page 1-2, line 23-49
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Page 3-4, line 72-78
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	Page 4-5, line 98-104
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation datasets, if applicable.	Page 5-6, line 109-162
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	Page 6, line 127-128
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	Page 5-6, line 116-120
	5b	D;V	Describe eligibility criteria for participants.	Page 6, line 124-128
	5c	D;V	Give details of treatments received, if relevant.	N.A.
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Page 5, line 112-114
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N.A. data analysed ex post
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Page 6-9, line 142-195
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N.A.
Sample size	8	D;V	Explain how the study size was arrived at.	Page 6, 135-140 and page 7-8, line 165-180
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Page 7-8, line 164-180
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	Page 10, line 230-237
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	Page 9, line 210-215 and Page 10-11, line 224-260
	10c	V	For validation, describe how the predictions were calculated.	Page 12, line 263-267
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	Page 12-13, line 264-301

	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N.A.
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N.A.
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Page 9-10, line 216-222
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	N.A.
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	N.A.
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	N.A.
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Page 9-10, line 216-218
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N.A.
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Table 1
	15b	D	Explain how to the use the prediction model.	Formula 3
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	Table 2
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N.A.
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	Page 18-19, line 417-435
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	N.A.
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Page 19-20, line 437-463
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	Page 20, line 461-482
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and datasets.	N.A.
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	Page 21-22, line 504-510

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Chapter 4

Prediction of the start of calving: what's in it for mid-size Holstein herds in the United States?

L. Leenaerts

C.J. Rutten

C. Kamphuis

H. Hogeveen

J. Bewley

ABSTRACT

This study's objective was to determine the economic benefits of a sensor system that predicts the start of calving in dairy cattle on typical mid-sized (100-500 cows) dairy farms in the United States. To do so a specialized calving sensor already used in the equine sector was compared to a estrus detection system with an additional algorithm for calving detection. Dynamic discrete event Monte Carlo simulation was used to estimate the economic benefits. Stochastic information for input variables was derived from scientific literature, survey results, and the authors' expertise. Effects on insemination costs, time spent observing close-up cows, assisting cows during calving, days open, treatment, culling and lost milk production due to metritis, stillbirth rate, and the costs, lifetime, time to apply the sensor, and subscription costs related to the sensor systems were considered. Marginal profit of the equine sensor was on average $0.97 per cow on mid-sized dairy farms with a range from - $21.70 to $963.67. This profit mainly consisted of a reduction in labor costs, and a reduction in metritis incidence and stillbirth rates. The alternative sensor was already used for estrus detection, therefore no investment cost were incurred. This caused profit for the estrus sensor the be higher that the profit of the equine sensor. The most influential input was the labor costs regarding calving management.

INTRODUCTION

In recent years, many sensor systems have been proposed and evaluated for dairy cows *(Rutten, et al., 2013)*. Such sensors could potentially be used as an alternative to visual observation for onset and progression of calving *(Borchers, et al., 2015, Ouellet, et al., 2016, Rutten, et al., 2017, Titler, et al., 2015)*. For instance, the walking behaviour and postures of cows prior to calving have been analyzed with video camera's *(Cangar, et al., 2008)*. Another example is temperature, which has been shown to decrease before the start of calving and this can be monitored with sensors *(Burfeind, et al., 2011, Saint-Dizier and Chastant-Maillard, 2015)*. Lastly, changes in rumination, lying time and lying bouts determined by sensors have been associated with the start of calving *(Bar and Solomon, 2010, Borchers, et al., 2015, Bucher and Sundrum, 2014, Ouellet, et al., 2016, Rutten, et al., 2017, Titler, et al., 2015)*.

Calving is a stressful moment for both cow and calf. In the United States, 19.0% of heifers and 6.0% of multiparous Holstein cows experience a severe form of dystocia at calving *(Meyer, et al., 2000)*. Another 9.7% of heifers and 4.7% of multiparous cows experience a milder form of dystocia *(Meyer, et al., 2000)*. Stillbirth rates in Holstein herds are 7% *(Meyer, et al., 2000)*. Calves born dead account for 78% of the stillbirths, whereas calves born alive but died within 48 hours after birth account for 22% *(USDA, 2007)*. Providing proper and timely assistance at calving can significantly reduce calf morbidity and mortality, especially when dystocia occurs *(Lombard, et al., 2007)*. Next to the effect on the calf, timely calving assistance reduces parturition stress and improves the cow's reproductive performance *(Barrier, et al., 2013, Mee, et al., 2014, Palombi, et al., 2013)*. Timely detection of the start of calving is therefore important to ensure timely assistance. Farmers on average conduct observations on cows that ate about to calf with 3.6 hour time intervals during the day while at night this interval increases to 5.2 hours *(USDA, 2007)*. So farmers observe late gestation cows more frequently during the day than at night. It has been suggested that more frequent observation of late gestation cows at night, on Sundays and bank holidays would reduce still birth rates *(Mee, et al., 2014)*. Visual observation is a labor- and time-consuming practice *(Palombi, et al., 2013)*, moreover the presence of an observer has been associated with an increased stress level resulting in calving problems *(Dufty, 1981)*. So, the current practice could be improved by introducing an automated monitoring system that observes late gestation cows more frequently than farmers do. Such a system should induce minimal stress and give the farmer a timely alert for the start of calving.

In large U.S. herds (>500 cows), the number of calvings is most likely large enough to have an herdsman dedicated to calving management. On smaller dairies (100-500 cows), however, this dedicated calving management is not feasible and time will have to be divided between calving management and other farm tasks. Studies that evaluated the reasons of farmers to adopt sensor technology found that economic value is an important factor *(Borchers and*

Bewley, 2015, Steeneveld and Hogeveen, 2015). Economic evaluations of sensor systems are therefore important. The adoption of sensors for calving detection on commercial dairy farms is still low *(Palombi, et al., 2013)*. However, a system that helps farmers by predicting the moment of calving could be beneficial. Such a system would reduce time spend on visual observation of the calving process. For sensor systems that detect calving two approaches can be used. First of all a specialized sensors system for calving detection could be used. The advantage could be a high detection performance because the system is specifically designed to detect calving, but such a system will require additional investments. An example of such a specialized sensor system can be found in the equine industry *(Palombi, et al., 2013)*. This specific sensor is an intravaginal device from which one module is expelled during the second stage of parturition. This expelled module transmits a signal when it is outside the body *(Palombi, et al., 2013)*. When this sensor system was applied on dairy farms studies reported a significant decrease on calf mortality, uterine infections and number of days open *(Palombi, et al., 2013, Paolucci, et al., 2010)*. Secondly an existing sensor could be extended with an additional algorithm to detect calving, next to its original purpose. Such a sensor could have a lower detection performance because it was not specifically designed to detect calving. But, the system was assumed to be already in use, so no additional investment or labor was required to use the sensor system. An example of such a sensor is a sensor used for estrus detection to which calving detection was added as additional feature *(Rutten, et al., 2017)*.

The objective of this study was to determine the economic feasibility of two sensor systems that predicts the start of calving on mid-sized (100-500 cows) dairy farms. One specialized sensor, which is currently adopted in the equine sector and one currently used for estrus detection with an additional calving algorithm will be appraised.

MATERIALS AND METHODS

Study design

A dynamic discrete event Monte Carlo simulation model was developed to evaluate the benefits and costs of a sensor for the start of calving. With such a model, it is possible to encompass for the variation that naturally occurs in a biological system *(Bruijnis, et al., 2010)*. The Monte Carlo simulation model was developed in Microsoft Excel 2013® using add-in software @Risk 7.0® *(Palisade-Corporation, 2010)*. Model inputs were based on scientific literature, survey results and the authors' expertise. A sensitivity analysis was conducted to identify the variables that had the largest influence on the profitability of a sensor system for calving.

Model structure

General. A schematic description of the simulation model is given by Figure 1. Each iteration of the simulation model equals one calving. First it is simulated whether a heifer or a multiparous cow calved. In the next step it was simulated whether *dystocia* occurred. If dystocia occurred, subsequent stillbirth and metritis were simulated with their effects on postpartum insemination success. It was assumed that when a sensor gives an alert for calving a herdsman will check the cow and provide the needed assistance if necessary and as a consequence could reduce the adverse effects of *dystocia*. The number of calvings was simulated per year and split up for heifers and multiparous cows. For this herd 50,000 calvings were simulated in order to reach model convergence. A U.S. mid-size Holstein herd of 300 cows was assumed for the economic evaluation. For this herd, the herd parameters from Table 1 were used and the herd size was assumed to be constant over time. Application of a sensor system for calving was assumed to have an effect on insemination costs, time spent observing close-up cows, assisting cows during calving, days open, treatment, culling, stillbirth rate and lost milk production due to metritis. The marginal profit of a sensor system for the start of calving was defined by the financial difference between visual calving detection and detection by the sensor system.
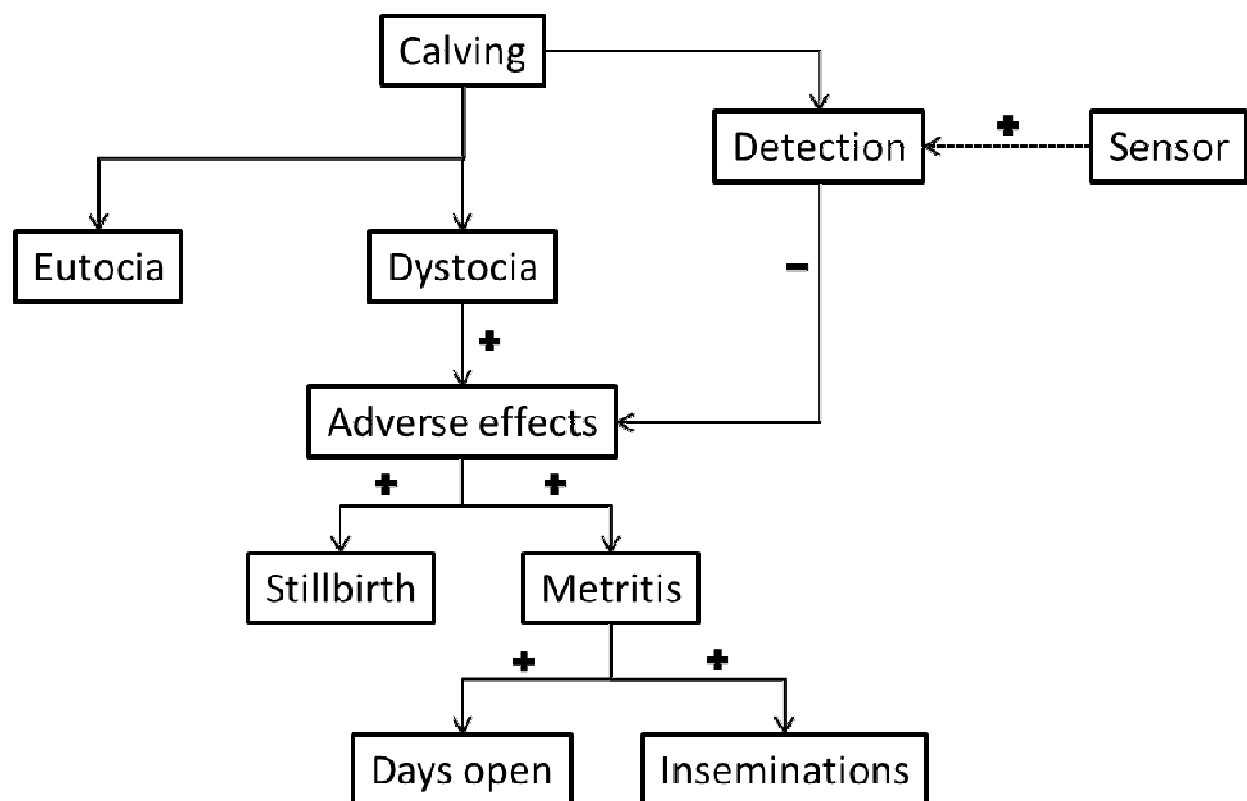

Figure 1: Schematic description of the discrete event Monte Carlo simulation model used to estimate the economic benefits of a sensor for calving detection.

Table 1. Herd Parameters

Parameter	Value	Reference
Calving interval (months)	13.0	*USDA (2016)*
Number of milking cows	300	Authors
Percentage of heifer calves (%)	46.7	*del Rio, et al., 2007*
Dystocia incidence (%) (mild, severe)	Heifers: (9.7, 19.0) Cows: (4.7, 6.0)	*Meyer, et al., 2000*
Stillbirth rate due to dystocia (%) (no dystocia, mild dystocia, severe dystocia)	Heifers: (6.1, 14.3, 27.7) Cows: (3.9, 12.6, 26.5)	*Meyer, et al., 2000*
Mortality (48h – weaning) (%)	9.1	*USDA, 2007*
Mortality (weaning – calving) (%)	2.0	*USDA, 2007*
Pre-calving culling rate (%)	1.0	*USDA, 2007*
Twinning rate (%)	4.15	*del Rio, et al., 2007*
% of herd in herd by lactation number	1 (36.10%) 2 (26.00%) 3 (17.70%) 4 (11.00%) ≥5 (9.20%)	*Dhuyvetter, et al., 2007*
Rolling herd average milk production	10236 kg	*Food and Agricultural Policy Research Institue (FAPRE), 2015*
Age at first calving (months)	24.8	*USDA (2016)*
Mature cow live weight (kg)	624	*Dhuyvetter, et al., 2007*
Days in milk to not breed an open cow (days)	300	*Caraviello, et al., 2006b*
Dystocia rate (%)	Heifers: 28.7 Cows: 10.7	*Meyer, et al., 2000*
Dystocia odds ratio of metritis	2.12	*Dubuc, et al., 2010*
Voluntary waiting period (days)	56	*DeJarnette, et al., 2007*
Estrus detection rate (%)	85	*Firk, et al., 2002*
Target days open (days in milk)	115	*Bartolome and Archbald, 2011*
Reduction in dystocia related metritis (%) when a sensor system is used	20	Authors
Reduction in metritis related metritis (%) when a sensor system is used	20	Authors
Effect of metritis on inseminations	0.33	*Fourichon, et al., 2000*
Effect of metritis on days open (days)	18.6	*Fourichon, et al., 2000*

The total costs of days open were estimated with the methodology of Dolecheck and Bewley *(2016)*. To calculate the costs of days open parity specific regression equations were used, which included all returns forgone per cow by extra days open. To estimate the returns forgone for reduced milk production "milk price" and "feed price" (Table 2) were used and these prices were estimated by calculating average prices for the years 2011-2015 *(Food and Agricultural Policy Research Institue (FAPRE), 2015)*. The costs per extra day open after 115 DIM were estimated at $0.66 per day excluding insemination costs. Up to 115 DIM the costs

of an extra day open were assumed to be near zero based on the estimations of Dolecheck et al. *(2016a)*.

Table 2. Financial Parameters

Parameter	Value	Reference
Labor cost during day ($/hr)	11.6	*Hertz, 2016*
Labor cost during night ($/hr)	15	Authors
Heifer calf value ($)	400	*Dhuyvetter, et al., 2007*
Bull calf value ($)	100	*Dhuyvetter, et al., 2007*
Insemination costs ($/insemination)	18	*Overton, 2005*
Non reproduction related metritis costs ($/case)	Heifer: 161.91 Cow: 187.04	*Liang, 2013*
Price of sensor ($)	300	*Moocall, 2016*
Annual subscription cost ($/sensor)	150	Authors
Discount rate (%)	8	*Dolecheck and Bewley, 2016*
Veterinarian costs ($/cow/year)	62	*Dolecheck and Bewley, 2016*
Milk price ($/kg)	0.39	*Food and Agricultural Policy Research Institue (FAPRE), 2015*
Feed price ($/kg dry matter)	0.20	*Food and Agricultural Policy Research Institue (FAPRE), 2015*
Replacement price ($)	1800	*Dhuyvetter, et al., 2007*
Cull cow price ($/kg)	1.87	*Dolecheck and Bewley, 2016*

Visual calving detection. For visual calving detection the assumption was used that a herdsman visually and manually checks late gestation cows every morning and evening for signs that calving will start, e.g., weakening of pelvic ligaments and teat filling. These visual checks were used for the situation without a sensor system, which served as the reference situation in the analysis. The time spent on visual checks by farmers for the start of calving were based on the response obtained from a questionnaire. This questionnaire was distributed to 7 mid-size (100-500 cows) Holstein dairy farmers from Kentucky during the Center of Kentucky Dairy Short Course in Green County on February 11, 2016. Farm labor was assumed to be more expensive ($15 / hr) at night (author expertise), than during the day ($11.60 / hr) *(Hertz, 2016)*.

Calving detection with a specific sensor system. For the baseline scenario it was assumed that the specific calving sensor developed in the equine sector (equine sensor) as described by Palombi et al. *(2013)* was applied. In their study Palombi et al. *(2013)* found that the "equine sensor", when applied on dairy cows, detected all calvings and did not generate false alerts. As that specific study included only 60 calvings, the current study adjusted the performance slightly, but still assumed a good detection performance for this "equine sensor", namely 95% sensitivity with 99% specificity. When the "equine sensor" was used to predict the start of calving, the assumption was that a herdsman checks the alarm and intervenes when necessary and thereby reducing the adverse effects of *dystocia*. In the situation with the "equine sensor" it was assumed that for 50% of the cows expected to

calve within 12 hours, a fast close-up check was done during the day (Table 3). Furthermore, costs for inserting the sensor in the cow, annual costs of the system and telephone costs for sending an alarm were estimated.

Table 3. Calving observation and assistance related parameters

Parameter	Value	Reference
Duration of night (hrs)	8	Authors
Daytime Calving observation interval (hrs)	0.5 (1.40%)	*USDA, 2007*
	2 (45.80%)	
	4 (35.20%)	
	6 (8.70%)	
	8 (8.90%)	
Night Calving observation interval (hrs)	0.5 (3.60%)	*USDA, 2007*
	2 (15.10%)	
	4 (27.40%)	
	6 (27.70%)	
	8 (26.20%)	
Fast close-up check	3	Authors
Thorough observation duration (min)	13.3	Questionnaire
Assistance duration (min)	22.0	Questionnaire
Lifetime of sensor (years)	4	Authors
Amount of sensors needed (100, 300, 500 cows)	(4,8,12)	Authors
Duration of applying sensor (min)	Heifer: 5.4	*Palombi, et al., 2013*
	Cow: 4.5	
Heifer calvings assisted (%)	31	*USDA, 2007*
Cow calvings assisted (%)	20.6	*USDA, 2007*
Time spent checking alarms (min/calving)	10	Authors
Increase in calving assistance when a sensor system is used (%)	10	Authors

Calving detection with a multipurpose sensor system. The second sensor system that this study evaluates is a sensor for multiple purposes. This sensor measures activity and rumination activity and is already used for estrus detection (estrus sensor). Two time windows were assumed with 21.2% or 54.4% sensitivity and specificity was assumed to be 99% *(Rutten, et al., 2017)*. The "estrus sensor" was already in use so no additional investment or labor for attachment of the sensors was assumed. Furthermore, the multipurpose sensor did not require extra labor for attachment of the sensor or subscription for telecommunication. It was assumed that the multipurpose sensor was used in the same way as the "equine sensor".

Dystocia reduction through calving sensors. The use of a calving sensor (equine or estrus sensor) was assumed to mitigate the effects of *dystocia,* where *dystocia* was assumed to only affect metritis and stillbirth rates. Metritis and stillbirth rates due to *dystocia*, were obtained by estimating the difference between the rates for *eutotic* calvings and *dystocic* calvings (Table 1). In a case control study Palombi et al. *(2013)* reported reductions in stillbirth rates and incidence of metritis. That specific study was small with only 60 calvings. Therefore, the minimum and maximum incidence rates for metritis and stillbirth in the

sensor group and the control group of the study of Palombi et al. *(2013)* were estimated with Win Epi *(de Blas, 2010)*. Based on these estimations, a baseline scenario was chosen with a 7.24 percent point lower still birth rate and a 12.63 percent point lower metritis incidence. The average time spent on checking alarms, going to the calving pen and other extra labor related to monitoring the sensor system was assumed to be 10 minutes per calving (Table 3).

Some parameters were assumed to be correlated. The observation intervals (Table 3) during the day and at night were assumed to be correlated, because it a farmer who infrequently observes his cows during the day is likely to also observe infrequently during the night. Stillbirth rates were assumed to be correlated with metritis rates, the effect of metritis on respectively days open and the required inseminations until conception. These assumed correlations were based on the idea that a stillbirth increases the risk of an infection of the endometrium causing metritis and subsequently lowering the probability of insemination success. All correlations were assumed to be 0.5.

Parameter distributions

Table 1 summarizes the herd parameters used in this study. Herd parameters included amongst others herd size, disease incidences and mortality rates. The effects of metritis on the number of days open were assumed to be normally distributed with mean 18.6 and standard deviation 0.612 *(Fourichon, et al., 2000)*. The number of inseminations required until conception was modeled with a geometric distribution assuming an effect of metritis on insemination success of 0.33. The geometric distribution is a probability distribution of the number of Bernoulli trials that are required to get the first success in a sequence of independent Bernoulli trails *(Derman, 1973)*.

Table 3 summarizes the parameters regarding calving observation and assistance. These parameters contained labor inputs, inputs regarding the sensor system and calving assistance. The variables "observation duration" and "assistance duration" (Table 3) were assumed to be distributed normally, the mean and sample variance were obtained from the questionnaire conducted during the Center of Kentucky Dairy Short Course. The normal distribution for the variable "duration of applying sensor" (Table 3) was based on Palombi et al. *(2013)*. A discrete distribution was used to model the variable "close-up observation interval" (Table 3). The variable "time spent checking alarms" (Table 3) was modeled with a beta-PERT distribution, which uses a most likely value with a minimum and maximum value that were assumed to be 25% lower and higher than the most likely value, respectively.

Table III summarizes the financial parameters. The financial parameters included costs of disease, prices of inputs and costs of the sensor system. The "metritis costs" (Table 2) were assumed to be normally distributed based on *(Liang, 2013)*, and included veterinary costs, treatment costs, labor costs, discarded milk, decreased milk production, higher culling risk and mortality risk. For the variables "prices of bull and heifer calf", "insemination costs" and

"time for fast close-up checks" (Table 2), beta-PERT distributions were used similar to the one used for "time spent checking alarms".

Sensitivity analysis

The sensitivity analysis was conducted to simulate three different scenarios. In the sensitivity analysis the influence of a selection of parameters was estimated. Firstly, the magnitude of the impact of the assumed metritis and stillbirth rates on the profitability of the sensor system was assessed. For stillbirth and metritis rates, two scenarios were simulated in which the minimal and maximal estimated incidence rates (determined with Win Epi) were used to determine a reduction in stillbirth and metritis when a sensor is used. This resulted in two scenarios with either a 2.44 or a 21.30 percent point lower stillbirth rate and with either a 4.75 or a 26.16 percent point lower metritis incidence in the "Low" and "High" scenarios, respectively. Secondly, the effect of farm size on profitability was estimated by simulating herd sizes of 100 and 500 cows. Thirdly, the effect of the number of visual checks was simulated in the sensitivity analysis with 20% and 80% of the cows expected to calve being checked. And finally, the fourth sensitivity analysis was conducted on the effect of labor costs during day and night time, both were changed positively and negatively by $ 1.-.

RESULTS

Parameter calculations

Heifers and multiparous cows accounted for respectively 39.2% and 60.8% of the simulated calvings on an average mid-size U.S. dairy farm. For a simulated herd of 300 cows with a calving interval of 13 months this is equivalent to 109 heifers and 168 multiparous cows calving each year. For heifers, 44.5% of the simulated stillbirths resulted from *dystocia*, and 55.5% from another cause. For multiparous cows, 31.2% of the simulated stillbirths resulted from *dystocia*, and 68.8% from another cause. As the incidence of *dystocia* was 28.7% for heifers and 10.7% for multiparous cows, it follows that 4.9% and 1.8% of the simulated calves were born after *dystocia* in heifers and multiparous cows, respectively. With the assumption of 95% sensitivity and a 7.24% decrease in stillbirth caused by *dystocia* when a sensor system is used, total stillbirth rate would decrease 0.34% for heifers and 0.12% for multiparous cows. With the assumption of 95% sensitivity and a 12.63% decrease the incidence of metritis caused by *dystocia* would decrease with 1.99%.

Profitability

For a herd size of 300 cows (Table 4), profit increased on average $0.97 per calving when the "equine sensor" was used (baseline scenario). The benefits resulted mostly from a decrease in time used to observe close-up cows and from a decrease in metritis and stillbirth incidence. Increased fertility added less to the benefits of a calving sensor. With 277 calvings per year, the total marginal profit was on average $269 per herd per year.

For the "estrus sensor", profit was on average $1,582 and $1,632 per herd per year (Table 4) for sensitivities of 21.2 and 54.4%, respectively. The annual costs of the sensor system were zero. With the "estrus sensor", money was saved on metritis, stillbirth and days open, and savings increased with higher sensitivity.

Sensitivity analysis

The results of the sensitivity analysis are summarized in Table 5. In the "Low" scenario with the "equine sensor" having lower effects on stillbirths, marginal profit from the use of the "equine sensor" was - $0.02 per calving lower than the baseline. This result caused the "equine sensor" to be on average unprofitable for a farm size of 300 cows. In the high scenario, with the "equine sensor" having higher effects on stillbirth rate, marginal profit from the use of the "equine sensor" was $773 per herd per year. When the sensitivity of the "equine sensor" was reduced from 95% in the baseline scenario to 80%, profit was on average $252 herd per year.

The number of visual checks a farmer conducts did influence labor costs most strongly, but also effected the costs of stillbirth and metritis. On average the profit of the "equine sensor" was - $366 and $886 per herd per year when in 20% and 80% of the calvings, respectively, the cows were checked visually.

For a herd size of 100 cows profit was on average $13.82 per calving when the "equine sensor" was used. The benefits resulted mostly from a decrease in time used to observe close-up cows and a reduction in metritis and stillbirth incidences. Increased fertility did add less to the benefits of a calving sensor. With 92 calvings per year (for a herd of 100 cows) the total marginal profit was on average $1,271 per herd per year. For a herd size of 500 cows profit was on average - $1.67 per calving when a sensor system was used. So on average the calving sensor was not profitable on a farm with a herd size of 500 cows.

A sensitivity analysis for labor costs was conducted. When labor costs decreased with $1, the profitability of the "equine" was $161 per herd per year. When labor costs increased with $1, the profitability of the system was $368 per herd per year.

Table 4. Marginal effect of using a sensor system that predicts the start of calving on total profit in $ per calving for dairy farms of 300 cows with 20,000 simulated calvings per farm. Three systems were analyzed, a sensor originally from the equine sector , an activity and rumination activity measuring sensor with a 1 hour time window (TW 1) sensitivity 21.2% and with a 6 hour time window (TW 2) with a sensitivity of 54.4%, the specificity was in both cases 99%.

	Baseline (equine sensor)			Alternative (estrus sensor)					
				TW 1			TW 2		
	Mean (sd)	Min	Max	Mean (sd)	Min	Max	Mean (sd)	Min	Max
Insemination costs	0.02 (0.80)	0	80.14	0.02 (0.77)	0	63.51	0.02 (0.79)	0	63.28
Days open	0.04 (0.74)	0	13.78	0.05 (0.77)	0	13.76	0.05 (0.79)	0	13.46
Labor[1]	4.86 (3.71)	-10.08	34.74	4.86 (3.71)	-9.48	31.77	4.86 (3.73)	-7.61	30.73
Costs of metritis[2]	0.63 (10.98)	0	344.85	0.65 (10.96)	0	294.44	0.66 (11.00)	0	313.93
Stillbirth	0.79 (16.46)	0	495.21	0.12 (6.60)	0	457.36	0.30 (9.91)	0	451.59
Annual costs of sensor system[3]	-5.38 (0.48)	-11.63	-5.04	0.00 (0.00)	0	0	0.00 (0.00)	0	0
TOTAL	0.97 (33.17)	-21.7	963.67	5.71 (22.81)	-9.48	860.84	5.89 (26.22)	-7.61	873

[1] Labor costs included labor for observing cows and assisting cows when *dystocia* is suspected.
[2] Metritis cost included costs for treatment, culling and reduced milk production.
[3] Costs of the sensor system included investment costs, telecommunication subscription and labor to attach sensors to the cows.

Table 5. Marginal effect of using a sensor system that predicts the start of calving on total profit in $ per calving for dairy farms of 300 cows with 20,000 simulated calvings per farm. A sensor originally from the equine sector was analyzed under different scenarios. Namely, low effects on metritis and stillbirth (Low), high effects on metritis and stillbirth (high), detection sensitivity of 80% (SN 80%), visual check by the farmer on 20% of the calvings (Check 20%), visual check by the farmer on 80% of the calvings (Check 80%), herdsizes of 100 and 500 cows and labor costs 1$/hr lower (Labor - $1) and higher (labor + $1) than the baseline scenario.

	Scenario								
	Low	High	SN 80%	Check 20%	Check 80%	Herdsize 100	Herdsize 500	Labor - $1	Labor + $1
Insemination costs	0.01 (0.50)	0.05 (1.18)	0.01 (0.59)	0.02 (0.78)	0.03 (0.94)	0.02 (0.77)	0.02 (0.81)	0.02 (0.78)	0.02 (0.69)
Days open	0.01 (0.41)	0.10 (1.09)	0.05 (0.76)	0.05 (0.76)	0.05 (0.76)	0.05 (0.77)	0.04 (0.75)	0.05 (0.76)	0.04 (0.75)
Labor1	4.86 (3.72)	4.86 (3.73)	4.86 (3.72)	2.68 (2.90)	7.05 (4.87)	26.29 (10.67)	0.58 (2.40)	4.42 (3.44)	5.31 (3.98)
Costs of metritis2	0.18 (5.79)	1.31 (15.53)	0.63 (10.71)	0.63 (10.70)	0.61 (10.62)	0.64 (10.87)	0.60 (10.61)	0.64 (10.83)	0.61 (10.71)
Stillbirth	0.29 (10.11)	1.86 (24.80)	0.73 (16.00)	0.68 (15.22)	0.84 (17.30)	0.69 (15.58)	0.76 (16.34)	0.74 (15.98)	0.82 (17.12)
Annual costs of sensor system3	-5.38 (0.47)	-5.38 (0.47)	-5.38 (0.47)	-5.38 (0.47)	-5.38 (0.47)	-13.88 (0.47)	-3.68 (0.47)	-5.28 (0.43)	-5.47 (0.51)
TOTAL	-0.02 (12.59)	2.79 (31.37)	0.91 (20.50)	-1.32 (19.72)	3.20 (21.83)	13.82 (22.56)	-1.67 (20.31)	0.59 (20.75)	1.33 (21.47)

[1] Labor costs included labor for observing cows and assisting cows when *dystocia* is suspected.

[2] Metritis cost included costs for treatment, culling and reduced milk production.

[3] Costs of the sensor system included investment costs, telecommunication subscription and labor to attach sensors to the cows.

DISCUSSION

A reduction in sensitivity from 95% to 80% caused the average profit to be $0.06 lower than the baseline scenario. This indicates that the reduction in metritis and stillbirth due to the use of the sensor do not have a large financial impact. The reduction of labor costs affected the profitability of the "equine sensor" most. So, if farmers rely more on a sensor system to predict the start of calving, the system will be more profitable. However, when farmers do rely more on such a calving sensor, detection performance becomes very important. A low sensitivity will cause farmer to miss calvings when relying on the sensor and require farmers to do more visual checks, which in turn increases labor costs. Whereas low specificity would result in many false alerts that the farmer will check unnecessarily, which will also increase labor costs.

The "equine sensor" used in this study was specialized and was assumed to have a high detection performance. One disadvantage of this "equine sensor" is that it requires a farmer to insert the sensor into each late gestation cow. This imposes extra labor costs for the sensor system. According to the sensitivity analysis the "equine sensor" was not profitable when effects on stillbirth and metritis incidence were low, when the farmer performed few visual checks and for a herd size of 500 cows. This indicates that for profitability, it is important that the use of a calving sensors results in lower *dystocia* related problems. Furthermore, the amount of labor that can be saved is a very important factor in the profitability of a calving sensor. By contrast, the "estrus sensor" was already in use for estrus detection and therefore, no additional costs for the sensor system were incurred. The conducted simulations indicated that these lower costs for the "estrus sensor" can outweigh the lower number of calvings detected by the system due to the lower sensitivity of this "estrus sensor".

The "equine sensor" evaluated in this study detects a temperature drop caused by expulsion. The expulsion of the sensor is therefore directly related to calving. In the equine industry this system is used *(Palombi, et al., 2013)*. A study on dairy cattle included 14 calvings, under research conditions concluded that the sensor predicted all calvings correctly and no false alerts were generated *(Palombi, et al., 2013)*. A disadvantage of this system is likely to be the invasive procedure to insert the sensor. Hygiene would be very important for such a procedure especially if the sensor is present for multiple days otherwise the risk of infections of the urinary tract could be high. A second potential problem is that an invasive procedure may induce stress for a cow. Stress in the last days before calving increases the risk on *dystocia (Mee, 2008)*..

The "estrus sensor" measures changes in activity and rumination activity which are approximations of behavioral changes occurring around calving. However, there are also other reasons for cows to be less active or ruminate less. This could explain the lower

detection performance of the multipurpose sensor. The "equine sensor" is expected to have a much higher sensitivity, than the "estrus sensor". On the other hand the "estrus sensor" was assumed to be already in use for estrus detection and thus no additional investments are required. Moreover, the "estrus sensor" does not require invasive handling of the cow, which makes application much easier in practice.

Studies have regarded the potential of improving calving management by using specialized calving sensors *(Palombi, et al., 2013, Paolucci, et al., 2010)*. The sensitivity and specificity were not studied explicitly. Therefore, assumptions were needed regarding these inputs. In principle, expulsion of the sensor should be a phenomenon that is specific to calving. However, the sensitivity and the specificity of this system remain uncertain. Another main assumption was that the farmer inserts the sensors timely, which means a few days prior to calving. A long period may increase the risk on infections because it is an intravaginal sensor. However, difference occurs in practice, between the actual and the expected calving date. This difference may result from normal variation in gestation length of plus or minus two weeks *(Inchaisri, et al., 2010a)* or because the cow calves from a previous insemination. So, for application on dairy farms the performance will be influenced by the farmers accuracy registering insemination dates and expected calving dates.

The economic evaluation of a sensor system for calving detection is complicated as many factors are highly uncertain. Diseases like retained placenta, metritis and other inflammations in and around the uterus are related to the calving process *(Fourichon, et al., 2000, Rajala-Schultz and Grohn, 1999, Schuenemann, et al., 2013, Sheldon, et al., 2009)*, but it is hard to estimate the effect of the use of a sensor system on such diseases. Since retained placenta by itself was not assumed to be a problem, only the costs of metritis were taken into account while the effect on reproduction was simulated separately. For many variables no information could be found in literature, so for these inputs values were chosen based the combination of the best knowledge of the authors and expert opinion.

The potential profitability of a sensor system for calving detection was estimated in this study. Results varied in ranges of roughly $3,000 per year, indicating a high amount of uncertainty in the current analysis. The uncertainty is caused by the combination of variation in costs (e.g., costs of metritis for a multiparous cow varied had a mean of $186.04 per cow with a standard deviation of 51.74) and uncertain technical effects (e.g. a chance of 10.7% on *dystocia* for a multiparous cow and a 12.63% chance that subsequent metritis could be prevented). Furthermore, many effects are in fact unknown at the moment. The reduction of stillbirth and metritis due to intervention when *dystocia* occurs was an assumption made by the authors. On the other hand it is also unknown what effect unnecessary interventions will have on cow health and welfare. Furthermore the amount of *dystocia* and subsequent diseases will be farm specific, depending on breeding strategy, feeding and management of the dry cows (especially transfer to a different group in late gestation). Variation in profitability of a sensor system for calving is therefore expected amongst farms.

CONCLUSIONS

The main advantages of a sensor for calving detection were lower labor costs, lower stillbirth rates and a lower incidence of metritis. This can be expected to outweigh the investment and annual subscription costs for the "equine sensor"on most farms. Therefore, the specific sensor was just profitable. The "estrus sensor" was also profitable. The profitability was highest for the "estrus sensor" because it required much lower costs for investment and application.

ACKNOWLEDGEMENTS

Chapter 5

An ex ante analysis on the use of activity meters for automated estrus detection, to invest or not to invest?

C.J. Rutten

W. Steeneveld

C. Inchaisri

H. Hogeveen

ABSTRACT

The technical performance of activity meters for automated detection of estrus in dairy farming has been studied, and such meters are already used in practice. However, information on the economic consequences of using activity meters is lacking. The current study analyses the economic benefits of a sensor system for detection of estrus and appraises the feasibility of an investment in such a system.

A stochastic dynamic simulation model was used to simulate reproductive performance of a dairy herd. The number of cow places in this herd was fixed at 130. The model started with 130 randomly drawn cows (in a Monte Carlo process) and simulated calvings and replacement of these cows in subsequent years. Default herd characteristics were: a conception rate of 50%, an 8 week dry-off period, and an average milk production level of 8,310 kg per cow per 305 days. Model inputs were derived from real farm data and expertise. For the analysis, visual detection by the farmer ("without" situation) was compared to automated detection with activity meters ("with" situation). For visual estrus detection, an estrus detection rate of 50% and a specificity of 100% was assumed. For automated estrus detection, an estrus detection rate of 80% and a specificity of 95% was assumed. The results of the cow simulation model were used to estimate the difference between the annual net cash flows in the with and without situations (marginal financial effect) and the internal rate of return (IRR) as profitability indicators.

The use of activity meters led to improved estrus detection and therefore to a decrease in the average calving interval and subsequent increase in annual milk production. For visual estrus detection , the average calving interval was 419 days and average annual milk production was 1,032,278 kg. For activity meters, the average calving interval was 403 days, and the average annual milk production was 1,043,398 kg.

It was estimated that the initial investment in activity meters would cost €17,728 for a herd of 130 cows, with an additional cost of €90 per year for the replacement of malfunctioning activity meters. Changes in annual net cash flows arising from using an activity meter included: extra revenues from increased milk production and number of calves sold, increased costs from more inseminations, calvings, and feed consumption, and reduced costs from fewer culled cows and less labor for estrus detection. These changes in cash flows were caused mainly by changes in the technical results of the simulated dairy herds, which arose from differences in the estrus detection rate and specificity between the with and without situations. The average marginal financial effect in the with and without situations was € 2,827 for the baseline scenario, with an average IRR of 11%. The IRR is a measure of the return on invested capital. Investment in activity meters was generally profitable. The most influential assumptions on the profitability of this investment were the assumed culling rules and the increase in sensitivity of estrus detection between the without and the with situation.

INTRODUCTION

Economic losses caused by the non-optimal reproductive performance of dairy cows have been explored in several studies *(Galvão, et al., 2013, Giordano, et al., 2012, Groenendaal, et al., 2004, Inchaisri, et al., 2010b)*. Inchaisri et al. (*2010b*) estimated that a longer average calving interval of 407 or 507 days, when compared to a short average calving interval of 362 days, caused an average net economic loss of respectively €34 or €231 per cow per year. The longer calving intervals led to lower milk production, fewer calves and lower costs for the management of calving. As a prolonged calving interval is associated with economic losses, optimizing reproduction management at dairy farms may be profitable.

One aspect of reproduction management is estrus detection. Studies on the efficiency of visual estrus detection have reported values around 40% *(Roelofs, et al., 2010)* and 55% *(Firk, et al., 2002)*. When optimal visual estrus detection is applied, this efficiency can be increased to a value in the range of 60% to 90% *(Firk, et al., 2002, Roelofs, et al., 2010)*. However, this increase in efficiency requires an increase in labor input, varying from 1-3 periods of 30 minutes per day *(Firk, et al., 2002, Roelofs, et al., 2010)*. When a farmer relies on visual estrus detection either the detection rate is lower or the labor input tends to be higher.

The estrus detection rate could be improved without increased labor input by using sensors for automated estrus detection. The activity meter has been the most widely studied *(Rutten, et al., 2013)* and might improve estrus detection compared to visual detection done by the farmer. Improved estrus detection increases the likelihood that a cow in estrus becomes pregnant and thereby shortens the calving interval, which is economically beneficial.

Although benefits of automated estrus detection are expected, farmers need to invest in such a system. Literature on investment in sensor technology on dairy farms is scarce. In general, the decision of a farmer to invest depends on the profitability of the investment, the farm's financial position, economic prospects, farm size and presence of a potential successor *(Aramyan, et al., 2007, Oude Lansink, et al., 2001)*. A study among dairy farmers in Kentucky (USA) found that the actual or perceived economic benefit of using sensors influences a farmer's decision to adopt sensor technology *(Russell and Bewley, 2013)*. Only limited knowledge is available about the economic benefit of sensor technology. Hence, there is a need for economic models to quantify the economic benefit of activity meters, so that farmers are able to make informed investment decisions.

Some studies on the economic effect of using sensor systems for automated estrus detection have been conducted. *(Østergaard, et al., 2005)* estimated the breakeven point for investing in a system for online progesterone measurements. *(Bewley, et al., 2010b, van Asseldonk, et al., 1999a)* described models for analyzing an investment in sensor systems.

None of these studies considered an investment in activity meters for automated estrus detection, nor appraised any system that farmers apply in practice.

Although improved estrus detection is expected to decrease calving intervals, more factors are involved. The effect on calving interval and the subsequent effects on milk production, calves sold and costs associated with calving are the end result of a complex system of interacting factors (e.g. the effect of milk yield on conception rate). This complex system results in uncertainty and variation in the effects of improved estrus detection, which is difficult to capture with an analytical modelling approach. Moreover, it is difficult to explore the effect of differences in the sensitivity of estrus detection using an analytical modelling approach. Therefore, a stochastic simulation study was chosen to analyze the investment in activity meters. This approach did allow for all possible reproductive events.

The aim of this study was to analyze the investment in a sensor system that uses activity meters for automated detection of estrus. For this investment analysis, a stochastic simulation model was developed, which modeled the reproductive performance of a dairy herd based on variation between individual cows. This model was parameterized with Dutch data and used to quantify the financial effects of improved estrus detection over multiple years.

MATERIALS AND METHODS

Cow simulation model

The cow simulation model was developed in Microsoft Excel. For the stochastic properties of the model, the add-in software @Risk 6 for Excel *(Palisade-Corporation, 2010)* was used. The basic specifications of the cow simulation model were described by Inchaisri et al. *(Inchaisri, et al., 2010b)* and are therefore only described briefly here. New components of the cow simulation model are described in full detail. The new components of the model included: the reproductive cycle, simulation of a series of lactations over time, the culling and replacement of cows, the estimation of feed requirements, detection performance for estrus detection, and the use of calendar weeks. An overview of all model inputs is given in Table 1, and a formal description of the implementation in @Risk of the new components in the model is given in Appendix 5.1.

The model simulated a cow place in weekly time steps for a period of 10 years, 520 weekly time steps in total. Cow places were filled by cows for which all relevant fertility events were simulated (Figure 1). Multiple lactations were simulated, and at the first time step each cow started with an assigned parity. The simulation model consisted of a number of parallel and

interlinked processes: the reproductive cycle, the lactation stage of a cow, i.e. the number of weeks in milk (**WIM**), and milk production.
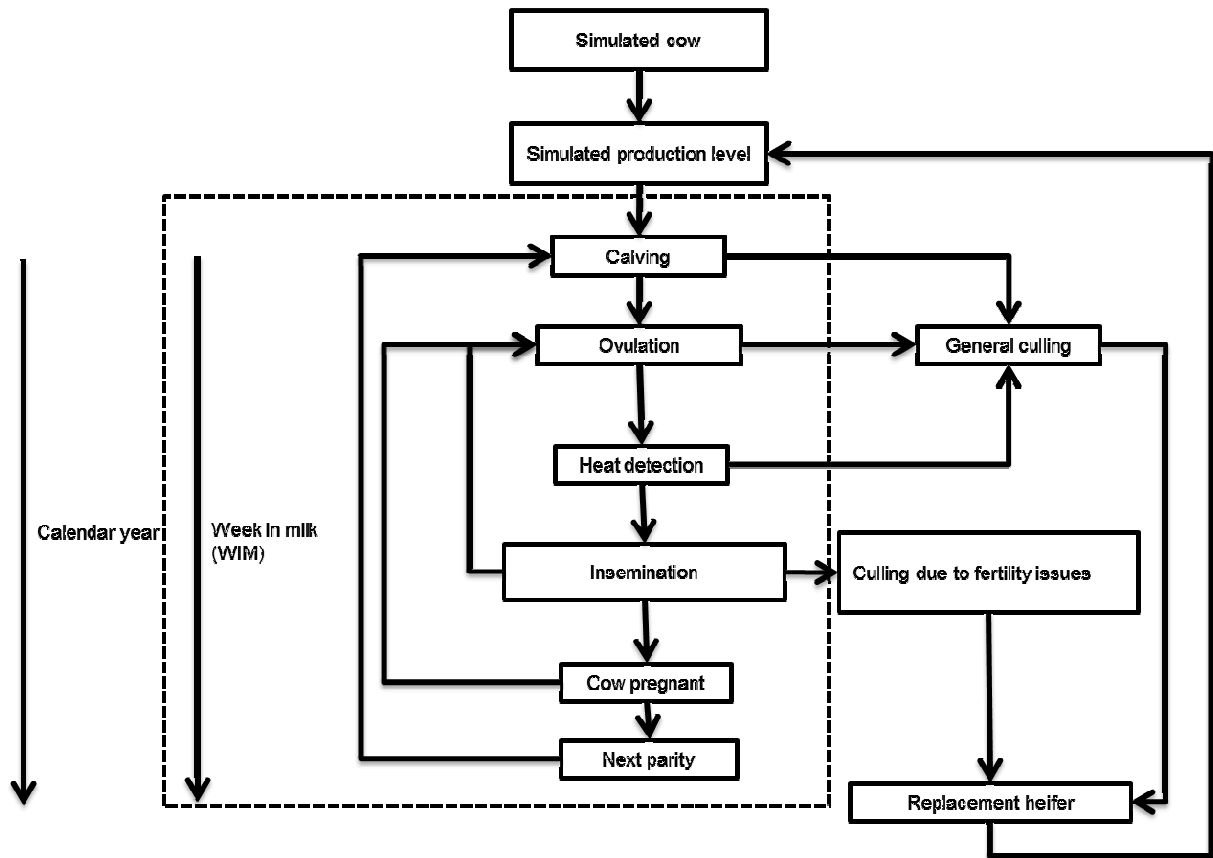

Figure 1: Schematic description of the cow simulation model, which looped through multiple lactations (shown inside the dashed frame) over simulated years. Each simulation starts in the beginning of a lactation in which several events can occur (e.g. ovulation, conception and culling), during this lactation the cow either conceived or was culled and replaced. The lactations and fertility events were simulated over multiple calendar years.

Table 1. Input values, and their relevant units and sources of origin, for variables in the cow simulation model for the baseline scenario

Variable	Value	Unit	Source
Visual estrus detection			
Sensitivity	50	%	Authors' expertise
Specificity	100	%	Authors' expertise
Activity meter			
Sensitivity	80	%	Based on: *Hockey, et al., 2010a,*
Specificity	95	%	*Kamphuis, et al., 2012*
Voluntary waiting period	12	weeks	*Inchaisri, et al., 2010b*
Dry-off period (DOP)	8	weeks	*Inchaisri, et al., 2010b*
Average pregnancy length (APL)	40	weeks	*Inchaisri, et al., 2010b*
Variation in calving moment of replacement heifer (Var_{hef})	2	weeks	Authors' expertise
Milk			
Fat percentage (F%)	4.25	%	*CRV, 2011*
Protein percentage (P%)	3.44	%	*CRV, 2011*
Culling rules			
CUL(wim)[1]	35	week	Authors' expertise
CUL(ins)[2]	6	Inseminations per lactation	Authors' expertise
Adj[3]	0.8		Based on: *(rickell and Wathes, 2011, Dechow and Goodling, 2008, Personal Communication, 2013b*
Daily milk yield at culling	15	kg	Authors' expertise
Feed[4]			
adj_par			
Parity 1	660	VEM[5]	*Remmelink, et al., 2012*
Parity 2	330	VEM	*Remmelink, et al., 2012*
adj_preg			
Week 21-26 of pregnancy	450	VEM	*Remmelink, et al., 2012*
Week 26-31 of pregnancy	1300	VEM	*Remmelink, et al., 2012*
Week 31-66 of pregnancy	2800	VEM	*Remmelink, et al., 2012*
Week >36 of pregnancy	5500	VEM	*Remmelink, et al., 2012*

[1] Cow not pregnant in week postpartum.
[2] Maximum number of unsuccessful inseminations.
[3] Proportion of culled cows that was not culled for fertility reasons.
[4] Adjustment factors for parity and pregnancy, which were used in the calculation of feed requirements.
[5] Feed requirements estimated as energy requirements in feed units lactation (VEM) as defined by van Es *(1978)*.

Reproductive cycle. A reproductive cycle started at calving, when the cow came into milk. Then the start of the cow's estrus cycle was simulated, which, on average, was the fourth WIM. Fertility disorders, such as metritis, were modeled as causing a delay in the onset of the cow's estrus cycle. Each estrus case had a probability to be detected. For this process a sensitivity and specificity were defined. The specificity was needed to account for false alerts when using an automated estrus detection system, i.e. the cow was thought to be in estrus when in fact she was not. Estrus and thereby estrus detection were simulated per week, specificity resulting in false positive alerts was simulated per day as farmers will check for estrus alerts on a daily basis rather than a weekly basis. For visual estrus detection by the farmer, a sensitivity of 50% and a specificity of 100% (no false alerts) was assumed. For activity meters, a sensitivity of 80% and a specificity of 95% was assumed. These assumptions were based on recent intervention studies with activity meters and discussion with fertility experts *(Hockey, et al., 2010a, Kamphuis, et al., 2012, Personal Communication, 2013a)*. A voluntary waiting period (**VWP**) of 12 weeks postpartum was assumed. Therefore, detected estrus events only led to insemination from week 12 onwards. False positive alerts were only simulated in periods when insemination could occur, as this was when false alerts led to additional costs for insemination. These periods were when a cow was both open and in a WIM that exceeded the VWP. In the simulation of insemination success, a conception rate and early embryonic death were taken into account. Simulation of estrus, estrus detection, and insemination continued until the cow was successfully inseminated or until a decision was made not to inseminate any longer. In the case of a successful insemination, the cow was eventually dried off. An average length of pregnancy of 280 days was assumed.

Weeks in Milk. The basic process for WIM was a straightforward counting of the WIM after calving (Inchaisri et al., 2010). In the first time step of the model, a cow with a randomly assigned parity started at WIM=1. To facilitate the calculation of annual costs and revenues, this first time step was randomly assigned to a calendar week (week of the year). From that first simulated week onwards, calendar weeks were recorded. Calendar weeks were used to calculate the annual levels for milk production, feed consumption, inseminations, calves born, and culled cows. When a cow calved the WIM was set to zero while the calendar weeks counted on. Another important determinant of WIM was culling. Culling was divided into culling due to fertility problems and general culling, which included all other culling reasons. A average culling rate of 30% *(Mohd Nor, et al., 2014)* was assumed, and 20% of the culled cows were assumed to be culled due to fertility problems *(Brickell and Wathes, 2011, Dechow and Goodling, 2008, Personal Communication, 2013b)*. In the simulation model, a set of culling rules for fertility problems was defined that was representative for the Dutch farming system. A cow was culled due to fertility problems when she failed to conceive before WIM 35 or within 6 inseminations. These rules accounted for a share of 20% in the total culling rate. In contrast to culling due to fertility, general culling was modeled as a chance process. For general culling, parity specific probabilities were used. These probabilities were derived from a frequency distribution of Dutch dairy cows, using the relative difference in the number of cows between one parity and the next.

The WIM in which the cow was culled was simulated using a chance process with a higher probability early in lactation. This distribution accounted for problems arising from calving and diseases in early lactation *(Rajala-Schultz and Grohn, 1999, Seegers, et al., 1998)*.

Milk production and feed requirements. For each individual cow, a lactation curve was simulated based on farm level milk production (input), parity, and pregnancy (Inchaisri et al., 2010). Based on the simulated lactation curve, feed requirements as energy requirements expressed in feed units lactation (VEM) van Es *(1978)* for each WIM were calculated. Higher feed requirements were simulated for first and second parity cows, and four pregnancy stages were defined to capture the different feed requirements during pregnancy (see Table 1).

Simulation and model outputs

A herd was assumed to consist of 130 cows in the baseline scenario. A herd size of 130 cows was chosen due to the trend of increasing herd sizes and adoption of automatic milking systems in dairy farming *(de Koning, 2010)*. A farm of 130 cows is representative for a future Dutch dairy farm with 2 automatic milking systems. For each herd, the model was run for 130 iterations to reflect the 130 cow places in the herd. For each analysis, a total of 50,050 iterations were run, equivalent to 385 herds. Stable simulation results were achieved at 50,000 iterations, and the extra 50 iterations were needed to ensure an integer number of herds. Each iteration ran for 10 years for each cow place. The first 4 years were needed to stabilize the culling rules, the last 6 years were used for the investment analysis.

The cow simulation model simulated 2 situations: farms with activity meters and farms without activity meters. Farms without activity meters relied on visual estrus detection. These situations are called the "*with*" and "*without*" situations. For each situation, the model was run for 50,050 iterations. In the simulation model, the 2 situations differed only by the estrus detection rate (sensitivity) and the estrus detection specificity. For each iteration, the model outputs included: total annual milk production (kg/year), annual number of inseminations, annual number of calves, annual number of culled cows, annual number of false alerts from the activity meters, and annual feed requirements (VEM/year).

Economic model

To analyze the profitability of an investment in activity meters for estrus detection, the technical outputs of the cow simulation model were used as input for a capital investment analysis.

The annual net cash flow on a farm was calculated as the sum of the cash flows for sold milk, feed costs, insemination costs (including false positive alerts), calf revenues, labor costs (including visual estrus detection, viewing estrus alerts, and confirmation of estrus alerts), and annual costs of activity meters. The annual net cash flow was then averaged over all farms in the simulation for each situation, resulting in an annual net cash flow for the with and without situations. The difference in the average annual net cash flows between the with and without situations was calculated. This difference was then averaged over the last 6 simulated years. In this study, this output is called the "*marginal financial effect*" of adopting activity meters for estrus detection. It is a measure of the average marginal benefit (or loss) from using activity meters for estrus detection. The time value of money was taken into account, where appropriate, by calculating the present values of future cash flows by using a discount factor (DF). In this study a discount rate (DR) of 5% was chosen. The calculation of the *DF* in year *k* is given by equation 1. In this study, the purchase costs of activity meters were assumed to be incurred and paid in year zero. Cash flows that arose from using the activity meter were received in subsequent years and were discounted accordingly.

$$DF_k = \frac{1}{(1+DR)^k} \tag{1}$$

Finally, the marginal financial effect of activity meters minus the investment costs was used to calculate four profitability indicators: the net present value (NPV), the benefit cost ratio (B/C ratio), the internal rate of return (IRR) and the discounted payback period (DPBP). The NPV indicates whether the present value of the net revenues exceeds the initial outlay of the investment. The NPV is calculated as the sum of all present values of revenues and all present values of expenses. The B/C ratio gives a measure of the return on invested capital. This measure is calculated by dividing the sum of present values of benefits (returns) by the sum of present values of costs. The IRR is the discount rate at which the NPV is zero. This percentage is an indicator of the annual return on invested capital over the life time of the investment. The DPBP is the number of years needed to pay back the initial purchase costs, which is calculated using discounted cash flows. The DPBP is defined as the year in which the sum of present cash flows equals or exceeds the initial purchase costs. As the technical lifetime of the sensor system was assumed to be 10 years, this time period was used in the calculation of all profitability measures. For each year of the technical lifetime, the marginal financial effect of activity meters was used as the annual cash flow arising from an investment in activity meters.

Economic input

Insemination costs. Insemination costs included the cost of sperm, the cost of the insemination service, and labor of the farmer. Based on the CRV pricelist *(CRV, 2012)*, insemination costs were estimated to be € 30 per insemination, assuming that on average 1.4 cows were inseminated per service.

Feed costs. For the estimation of feed costs, information on VEM content was based on Dutch standard values *(Remmelink, et al., 2012, Vermeij, 2012)*. Feed costs were expressed in euro per VEM. These costs were calculated based on a standard ration in which the VEM content came from concentrates (10%), concentrate replacing byproducts (10%), grass silage (48%), and corn silage (32%). For concentrates, it was assumed that standard and protein rich concentrates were used in a 1:1 ratio. Prices averaged over the period 2005 – 2012 were used. For corn and grass silage, sale prices were used as an estimate of the opportunity costs of feeding these roughages to the farm's dairy cows instead of selling roughage *(Vermeij, 2012)*.

Slaughter values and replacement heifers. The cash flows associated with culling were the slaughter value of the culled cow (revenue) and the opportunity cost of the replacement heifer (cost). For the estimation of slaughter values, live weights of 600 kg, 675 kg, and 700 kg were used for cows in their first, second, and third or higher parity, respectively *(Personal Communication, 2013c)*. Slaughter weights were calculated assuming a dressing percentage of 60%. The slaughter price, € 2.01 per kilogram slaughter weight, was an average of historical prices from the period 2003 to 2012 *(LEI, 2013)*. It was assumed that 6.7% of all culled cows were destructed at a cost of € 29 per destructed cow *(Rendac, 2013)*. The opportunity costs of a replacement heifer were represented by the sale price, as the alternative for introducing a heifer into the herd is selling the heifer. For the sale price, an average of prices during the period 2003 to 2012 was used (data for 2009 were lacking) *(LEI, 2013)*. The estimated average cash flow for culling and replacing a dairy cow was calculated as the sum of the revenues from slaughter and the opportunity costs of a replacement heifer, this was -€ 242, -€ 158 and -€ 130 for a cow in first, second, and third or higher parity, respectively.

Labor costs. The opportunity costs of labor were assumed to be € 18 per hour *(Huijps, et al., 2008)*. For visual estrus detection, it was assumed that a farmer spent 10 minutes per day on estrus detection. For automated detection, it was assumed that it would take a farmer 2.5 minutes to view the alert list of the activity meters (Table 2). Two options were analyzed for dealing with the false alerts produced by activity meters. Firstly, it was assumed that a farmer could visually confirm every alert of the activity meters, this confirmation took 5 minutes per alert. This assumption was implemented in the baseline scenario. Secondly, it was assumed that a farmer could blindly inseminate the cows after an alert, so without knowing whether the alert was a true positive or false positive alert. This last option was analyzed in a separate scenario.

Purchase cost of activity meters. The purchase cost of activity meters was estimated based on expert information *(Personal Communication, 2013a)* and the expertise of the authors on commercially available sensor systems. An activity meter was assumed to cost € 90 per device, and the neck collar to attach the sensor to the cow was assumed to cost € 18 per collar. In addition to the sensor, a system for transmitting data and software for processing data is needed, at assumed costs of € 3,600 per herd.

Annual costs of activity meters. The costs of maintaining the system were assumed to be 0.5% of the purchase price. In addition to maintenance, it was assumed that on average 1 activity meter needed to be replaced annually (Table 2).

Validation

The model was validated internally, as no suitable data were available for external validation. The following steps in the validation process were undertaken: outputs were checked for plausibility, the effect of extreme input values was evaluated, and all model outputs and assumptions were critically reviewed and discussed with experts.

Sensitivity analysis

In the baseline scenario, the situation with activity meters was compared to the without situation. Assumptions and input values have an effect on the results of the investment analysis. A sensitivity analysis was conducted to assess the effect of the assumptions and inputs in the following areas: culling, voluntary waiting period, herd size, estrus detection, and labor costs. The values and variables used in the sensitivity analysis are shown in the first column of Table 5.

Culling. Three alternative scenarios for the culling rules (culling due to fertility problems) were analyzed. In these scenarios the share of culling due to fertility in total culling was adjusted with an adjustment factor. The adjustment factors ensured that an average culling rate of 30% was maintained. A detailed description of the adjustment factor is provided in Appendix 5.1. Firstly, fully random culling (Random) was assumed in which no culling rules were used and the adjustment factor (ADJ) was set to 0%. Secondly, the culling rules were set to not pregnant in week 40 or 6 unsuccessful inseminations (Relaxed), with ADJ = 10%. Lastly, the culling rules were set to not pregnant in week 30 or 6 unsuccessful inseminations (Strict), with ADJ = 30%.

Voluntary waiting period. The baseline scenario assumed a VWP of 12 weeks, resembling the current practice on Dutch dairy farms. Inchaisri et al. *(2011b)* noted that a VWP of 9 weeks is a more economically optimal choice. Therefore, a scenario with a VWP of 9 weeks was also analyzed.

Table 2. Input values, and their relevant units and sources of origin, for variables in the economic analysis for the baseline scenario

Variable	Value	Unit	Source
Prices and costs			
Milk	0.32	€/kg	*Inchaisri, et al., 2010b*
Feed	0.16188	€/kVEM[1]	*Remmelink, et al., 2012, Vermeij, 2012*
Cost of insemination	30	€/insemination	Authors' expertise
Cost of calving management	152	€/calving	*Inchaisri, et al., 2010b*
Calf	100	€/calf	*Inchaisri, et al., 2010b*
Labor			
Hourly rate	18	€/hour	*Huijps, et al., 2008*
Confirmation alerts	5	min/alert	Authors' expertise
Reduced labor estrus detection	10	min/day	Authors' expertise
Labor viewing alerts	2.5	min/day	Authors' expertise
Slaughter values			
Parity 1	677	€/cow	*LEI, 2013, Personal Communication, 2013c, Rendac, 2013*
Parity 2	762	€/cow	*LEI, 2013, Personal Communication, 2013c, Rendac, 2013*
Parity 3+	790	€/cow	*LEI, 2013, Personal Communication, 2013c, Rendac, 2013*
Opportunity cost replacement heifer	919	€/heifer	*LEI, 2013*
Discount rate (DR)	5	%	Authors' expertise
Technical life time	10	years	Authors' expertise
Costs for activity meters			
Activity meter	90	€/cow	Experts' and authors' expertise
Neck collar	18	€/cow	Experts' and authors' expertise
Data transmission system and software	3,600	€/herd	Experts' and authors' expertise
Annual maintenance	0.5	% of purchase price	Experts' and authors' expertise
Replacement broken sensors	1	number/year	Experts' and authors' expertise

[1] Feed requirements estimated as energy requirements in feed units lactation (VEM) as defined by van Es *(1978)*.

Herd size. In the baseline scenario a herd size of 130 cows was assumed. The investment was also analyzed for a herd of 65 cows or 195 cows, these herd sizes are representative of small and large Dutch dairy farms.

Estrus detection. In practice, the visual estrus detection rate achieved by farmers may deviate between farms. It was assumed that the estrus detection rate was correlated with the time a farmer spent on estrus detection. In the sensitivity analysis the detection rate was set at 30%, 40%, 60%, or 70%, and the corresponding labor input (minutes per day) was set at 0, 5, 20, or 40, respectively. These values were based on the findings of Roelofs et al. *(2010)*. In this study, labor input ranged from 60 to 120 minutes per day, separated in several blocks of 30 or 60 minutes, whereas estrus detection rates ranged from 61% to 94% *(Roelofs, et al., 2010)*. Based on these findings, levels of labor input were assumed which would not overestimate labor costs.

Similar to visual estrus detection, the estrus detection rate of activity meters may vary between farms. It was assumed that the estrus detection rate of activity meters is correlated with the specificity of estrus detection. The detection rate for the sensitivity analysis was set at 65%, 75%, 85%, or 95%, and the corresponding specificity was set at 99%, 97.5%, 90%, or 85%, respectively. These detection performances were based on available literature *(Hockey, et al., 2010a, Kamphuis, et al., 2012)* and accounted for the trade-off between sensitivity and specificity.

Labor costs. Price fluctuations will influence the results of the investment analysis. Higher and lower values for the price of labor were used to account for farmers with a high (€ 36 /hour) or low (€ 9 /hour) opportunity cost of labor. Furthermore, some farmers might argue that they do not have an opportunity cost for their labor. Therefore, the investment was also analyzed without labor costs (€ 0 /hour).

RESULTS

Simulation results

The simulation results for the baseline scenario are shown in Table 3. The use of activity meters led to a decrease in the average calving interval per herd of 16 days, from 419 (without) to 403 days (with activity meters). This caused an increase in average annual milk production of approximately 11,000 kg/year, from 1,032,278 kg/year (without) to 1,043,389 kg/year (with). The 5th percentile results for average annual milk production were 1,032,278 kg/year (without) and 1,020,524 kg/year (with activity meters). The 95th percentile results for average annual milk production were 1,056,148 kg/year (without) and 1,066,806 kg/year (with).

Table 3. Technical simulation results and associated financial results for farms with visual estrus detection (without situation) and for farms with activity meters (with situation), and the difference in financial results between the with and without situations (Δ)

| | Technical results Mean (5th percentile / 95th percentile)[1] | | | | Financial results (€ / year) Mean (5th percentile / 95th percentile)[1] | | | | |
	Visual (Without)		Activity meters (With)		Visual (Without)		Activity meters (With)		Δ[2]
Calving interval (days)	419	(367 / 507)	403	(365 / 478)	-	-	-	-	-
Milk (*1000 kg)	1,032	(1,009 / 1,056)	1,044	(1,021 / 1,067)	330,329	(322,867 / 337,967)	334,000	(326,964 / 341,743)	3,558
Feed (*1000 kVEM[3])	792	(781 / 805)	799	(788 / 810)	-128,275	(-130,233 / -126,398)	-129,331	(-131,223 / -127,570)	-1,036
Calves (#)	144	(133 / 155)	147	(136 / 157)	-7,483	(-8,060 / -6,916)	-7,624	(-8,164 / -7,072)	-144
Inseminations (#)	223	(201 / 245)	245	(221 / 270)	-6,730	(-7,410 / -6,079)	-7,409	(-8,166 / -6,684)	-667
Inseminations / calf	1.55	(1.51 / 1.58)	1.67	(1.63 / 1.72)	-	-	-	-	-
Culling (#)	45	(34 / 56)	41	(31 / 51)	-7,400	(-9,494 / -5,481)	-6,462	(-8,449 / -6,682)	943
Labor (hour)	61	(61 / 61)	46	(50 / 43)	-1,096	(-1,096 / -1,096)	-834	(-902 / -767)	262
Net cash flow	-	-	-	-	179,345	(168,766 / 194,189)	182,262	(169,487 / 194,378)	2,917

[1] Mean annual values are presented, with 5th and 95th percentiles in parentheses.
[2] The difference between the mean financial results for the with (activity meters) and without (visual estrus detection) situations.
[3] Feed requirements estimated as energy requirements in feed units lactation (VEM) as defined by van Es *(1978)*.

In addition to the increase in milk production, the annual number of inseminations and annual number of calves born both increased with the use of activity meters, with a larger increase in inseminations than in calves born. The number of inseminations per calf was therefore higher for the situation with activity meters. An explanation for this increase is that a higher estrus detection rate leads to more detected estrus cases and therefore, on average, to earlier insemination. Insemination early in lactation has a lower chance of success, especially around peak production, and therefore more inseminations are likely to be unsuccessful.

The use of activity meters also led to a decrease in the annual number of culled cows, an average decrease of 4 culled cows per year. This decrease was a result of the culling rule that cows were culled if they were not pregnant by week 35 of the lactation. When the estrus detection rate increased, cows became pregnant earlier (hence the shorter calving interval), and fewer cows were culled for not becoming pregnant by week 35.

Financial results

The effects of the technical simulation results on the annual financial results are shown in Table 3. The increased milk production due to the use of activity meters had the largest effect on the financial results. The increased milk production led to an increase in the annual revenue of € 3,558 per year. The reduced number of culled cows and reduced labor input decreased the annual costs by € 942 and € 261, respectively. In the situation with activity meters, higher feed consumption increased costs by € 1,035 per year. The larger numbers of inseminations and calves born also increased the annual costs, by € 667 and € 143 per year, respectively. Overall, the difference in average annual net cash flow between the with and without situations, termed the marginal financial effect, was € 2,827. At the 5th percentile, the marginal financial effect was € 2,968. This was representative of low cash flow farms. At the 95th percentile, the marginal financial effect was € 2,633, representative of high cash flow farms. The marginal financial effect for low and high cash flow farms was used further in the investment analysis.

Investment analysis

The results for the investment analysis are presented in Table 4, for the different options for dealing with false estrus alerts from the activity meter. For farms that first visually confirmed estrus alerts, the average NPV of investing in activity meters values was € 4,098, with an NPV of € 5,887 for low cash flow farms and € 2,604 for high cash flow farms. The average B/C ratio was 1.23, with a ratio of 1.33 for low cash flow farms and 1.15 for high cash flow farms. As the NPV was positive and the B/C ratio greater than one, the investment was profitable. On average, the return on invested capital (the IRR) was 10%, with 11% for low cash flow

farms and 8% for high cash flow farms. The investment was paid back in 8 years for an average farm, with 8 years for low cash flow farms and 9 years for high cash flow farms. When a farmer inseminated blindly upon every alert, the investment was not profitable.

Table 4. Investment results for investing in activity meters on a dairy farm with 130 cows for the different options for dealing with false positive estrus alerts: the blind and confirmation scenarios

	Blind[1]		Confirmation[2]	
	Mean (low cash flow / high cash flow)[3]			
Cash flow (€/year)	-324	(-35 / -532)	2,287	(2,968 / 2,633)
Purchase (€/herd)	17,728	-	17,728	-
NPV[4] (€/herd)	-21,618	(-19,389 / -23,229)	4,098	(5,887 / 2,604)
B/C[5] ratio	-0.22	(-0.09 / -0.31)	1.23	(1.33 / 1.15)
IRR[6] (%)	NA	(NA / NA)	10	(11 / 8)
DPBP[7] (years)	NA	(NA / NA)	8	(8 /98)

[1] Blind refers to insemination after an alert without visual confirmation.
[2] Confirmation refers to visual confirmation of all alerts before insemination.
[3] Mean values are presented with results for low and high cash flow farms in parentheses, these indicators were calculated using the 5th and 95th percentile results for the marginal financial effect of activity meters (difference in annual net cash flow between the with and without situations).
[4] NPV: Net present value, the sum of all present revenues and present expenses.
[5] B/C ratio: Benefit / cost ratio, the ratio between the sum of present benefits (returns) and the sum of present costs.
[6] IRR: Internal rate of return, the discount rate at which the NPV is zero.
[7] DPBP: Discounted payback period, the number of years needed to payback the initial purchase costs using discounted cash flows.

Sensitivity analysis

Culling. Different culling rules for fertility problems caused changes in the number of cows culled, annual culling costs, and annual milk production. The IRR for the three different scenarios are shown in Table 5. These were 18% (Random), 18% (Relaxed), and -12% (Strict). For these scenarios, the share of cows culled for fertility problems in the total number of cows culled differed. The effect of this share on the marginal financial effect and IRR is shown in Figure 2.

Voluntary waiting period. When the VWP was shortened from 12 to 9 weeks, average calving intervals were shorter (data not shown) for both the without situation (403 days) and the situation with activity meters (385 days). However, with a VWP of 9 weeks, the IRR for investing in a sensor system was 10% (see Table 5), which was similar to the IRR in the baseline scenario. The marginal financial effect was € 2,844 for a VWP of 9 weeks (Table 5), which was slightly higher than in the baseline scenario.

Table 5. Marginal financial effect (difference in annual net cash flow between the with and without situations) and internal rate of return (IRR) for different scenarios in the sensitivity analysis.

Scenario		Marginal financial effect (€/year)		IRR (%)	
		Mean (low cash flow / high cash flow)[1]			
Baseline[2]		2,827	(2,968 / 2,633)	10	(11 / 8)
Random[3]		3,953	(4,547 / 3,473)	18	(22 / 15)
Relaxed[4]		3,972	(4,459 / 3,657)	18	(22 / 16)
Strict[5]		819	(949 / 472)	-12	(-10 / -19)
VWP[6] 9		2,844	(2,929 / 2,748)	10	(10 / 9)
Herd of 65 cows		1,779	(1,936 / 1,655)	11	(13/ 9)
Herd of 195 cows		3,870	(4,034 / 3,063)	9	(10 / 4)
Activity meters (with)[7]					
SN 65	SP 99	1,620	(1,529 / 1,587)	-2	(-3 / -2)
SN 75	SP 97.5	2,428	(2,420 / 2,251)	6	(6 / 8)
SN 85	SP 90	3,030	(3,181 / 3,127)	11	(12 / 12)
SN 95	SP 85	3,321	(3,404 / 2,913)	13	(14 / 10)
Visual (without)[8]					
SN 30	Labor 0	3,591	(4,057 / 3,291)	15	(19 / 13)
SN 40	Labor 5	3,102	(3,355 / 2,893)	12	(14 / 10)
SN 60	Labor 20	2,921	(3,038 / 2,814)	10	(11 / 9)
SN 70	Labor 40	4,417	(4,487 / 4,382)	21	(22 / 20)
Labor costs					
No labor costs	(€ 0 /hr)	2,631	(2,735 / 2,491)	9	(8 / 10)
Low labor costs	(€ 9 /hr)	2,729	(2,858 / 2,547)	9	(10 / 7)
High labor costs	(€ 36 /hr)	3,022	(3,196 / 3,806)	11	(12 / 9)

[1] Mean results are presented with results for low and high cash flow farms in parentheses, these indicators were calculated using the 5th and 95th percentile results for the marginal financial effect of activity meters (difference in annual net cash flow between the with and without situations).
[2] The baseline scenario assumed a sensitivity of 80%,with a specificity of 95% in the with situation, a sensitivity of 50% with a specificity of 100% in the without situation , a Voluntary waiting period (VWP) of 12 weeks, 20% of the culling is due to fertility problems, and labor costs of € 18 /hour.
[3] 0% of culled cows were culled for fertility problems.
[4] 10% of culled cows were culled for fertility problems.
[5] 30% of culled cows were culled for fertility problems.
[6] Voluntary waiting period (VWP) of 9 weeks.
[7] Scenarios under 'Activity meters' varied in the sensitivity (SN %) and specificity (SP %) of estrus detection.
[8] Scenarios under 'Visual' varied in the sensitivity (SN%) and labor input (in minutes) for estrus detection.
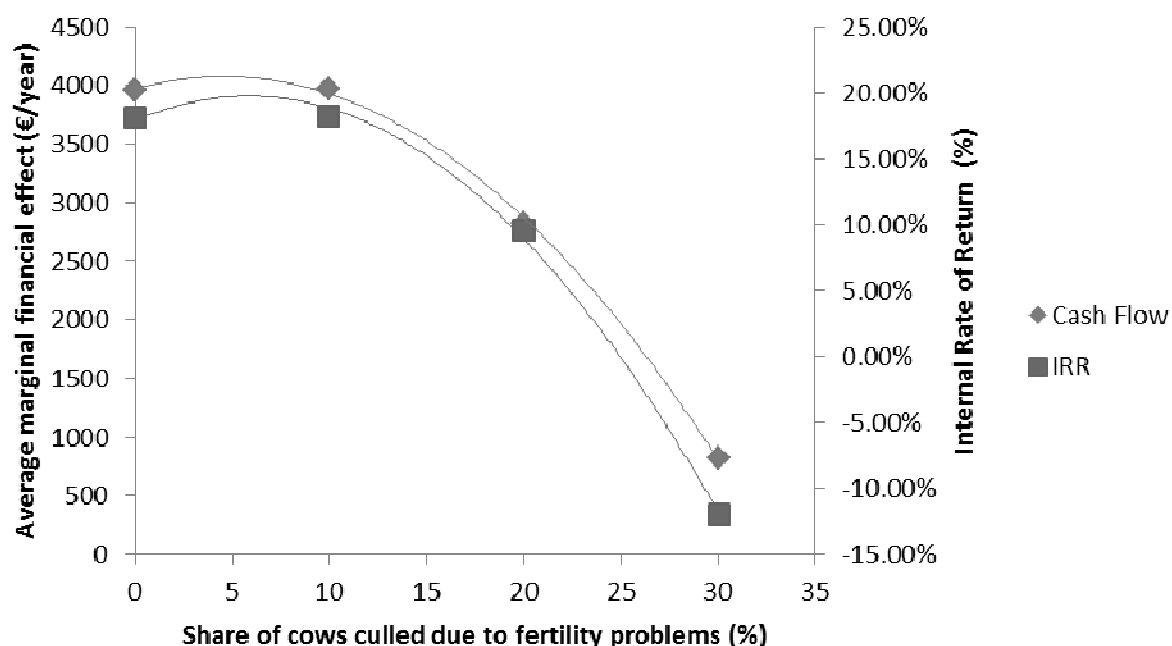
Figure 2: Effect of the share of cows culled for fertility problems on the average marginal financial effect (difference in annual net cash flow between the with and without situations) and internal rate of return (IRR) for investing in activity meters on a dairy farm with 130 cows. The x-axis represents the share (%) of cows culled for fertility problems in the total number of culled cows. Polynomial trend lines are included in the graph.

Herd size. For a herd of 65 cows the marginal financial effect was €1,779, much lower than that for a herd of 130 cows (see Table 5). However, this smaller herd had a slightly higher IRR (11%). For a herd of 195 cows, the marginal financial effect was €3,870, higher than that for a herd of 130 cows. The IRR for this larger herd (9%) was slightly lower than the baseline scenario.

Estrus detection. When sensitivity increased, the average calving interval shortened and the average annual milk production increased accordingly (data not shown). An increase in sensitivity had a stronger effect on the average calving interval at low sensitivities than at high sensitivities. For instance, increasing the sensitivity from 30% to 40% shortened the average calving interval by 7 days, from 435 to 426. Whereas increasing the sensitivity from 85% to 95% shortened the average calving interval by only 2 days, from 401 to 399.

In Figure 3 the marginal financial effect and IRR are shown for a fixed sensitivity of 80% with varying specificities, and for a fixed specificity of 95% with varying sensitivities. For a sensitivity of 70%, with specificity fixed at 95%, the IRR was below the discount rate of 5%, indicating that the investment was not profitable. Linear trend lines are included in Figure 3 to show the relationships between detection performance, and cash flow and IRR. An increase in sensitivity had a larger effect on cash flow and IRR than an increase in specificity had.
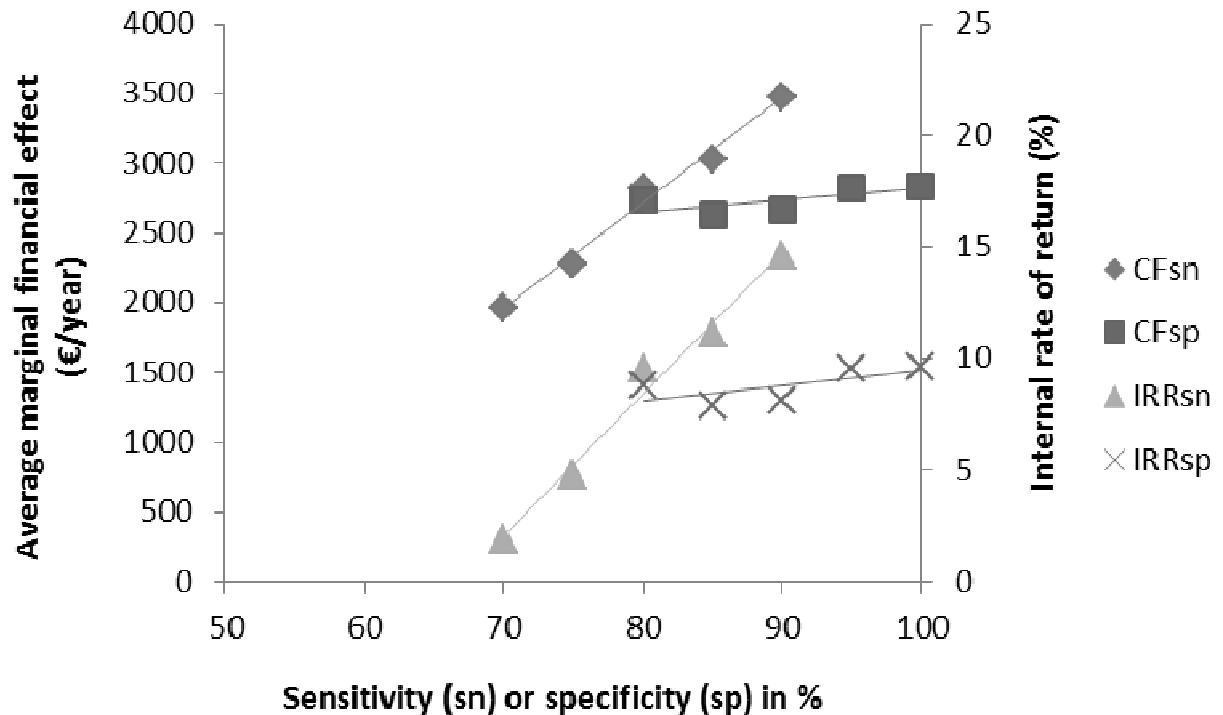
Figure 3: Effect of sensitivity and specificity of estrus detection on the average marginal financial effect (difference in annual net cash flow, CF, between the with and without situations) and internal rate of return (IRR) of investing in activity meters on a dairy farm with 130 cows. The x-axis represents different sensitivities (sn) and specificities (sp) in percent for estrus detection by activity meters. For different sensitivities, the corresponding specificity was fixed at the baseline level of 100% in the without situation. For different specificities, the corresponding sensitivity was fixed at the baseline level of 80%.

When the sensitivity of visual estrus detection changed, which was modeled by changing the labor input for visual estrus detection, the investment in activity meters was more profitable than in the baseline scenario (Table 5). Although the investment was more profitable for all scenarios, labor inputs smaller than the baseline level had a different effect than larger labor inputs. In the scenarios with a low labor input and low sensitivity for visual estrus detection, the use of activity meters meant a large improvement in estrus detection and therefore in the technical results. On the other hand, in the scenarios with a high labor input and high sensitivity for visual estrus detection, the improvement in estrus detection was relatively smaller but the use of the activity meter led to a larger saving in labor input.

When the sensitivity and specificity of activity meters was changed, the IRR was not always positive. In the scenarios with a high estrus detection rate for the activity meter (85% and 95%), the IRR was higher than in the baseline scenario (Table 5). Whereas in the scenarios with a low sensitivity (65% and 75%), the IRR was lower than in the baseline scenario. Moreover, the IRR was negative for the scenario with a sensitivity of 65%. Therefore, the corresponding NPV would not have exceeded zero. In addition to the sensitivity, the specificity of the activity meters was changed. A high specificity was combined with a low

sensitivity and vice versa. The lower specificity resulted in higher labor costs than for the scenarios with high specificities. On the other hand, in the scenarios with low sensitivity these labor costs were lower. Moreover, in all of the scenarios the investment was more profitable when a farmer confirmed estrus alerts before insemination (Table 6).

Table 6. Marginal financial effect (difference in annual net cash flow between the with and without situations) and internal rate of return (IRR) of investing in activity meters for different levels of sensitivity (SN%) and specificity (SP%) of estrus detection, and for different options for dealing with false positive estrus alerts

Scenario	Blind[1]				Confirmation[2]			
	Marginal financial effect (€/year)		IRR (%)		Cash flow(€/year)		IRR (%)	
	Average (low cash flow / high cash flow)				Average (low cash flow / high cash flow)			
SN 65, SP 99.0	844	(745 / 758)	-12	(-13 / -13)	1,620	(1,529 / 1,587)	-2	(-3 / -2)
SN 75, SP 97.5	517	(486 / 267)	-18	(-19 / -25)	2,428	(2,530 / 2,251)	6	(6 / 5)
SN 85, SP 90.0	-1,545	(-1,467 / -1,355)	NA	NA)	3,030	(3,181 / 3,127	11	(12 / 12)
SN 95, SP 85.0	-1,621	(-1,248 / -1,662)	NA	NA	3,221	(3,404 / 2,913)	13	(14 / 10)

[1] Blind refers to insemination after an alert without visual confirmation.
[2] Confirmation refers to visual confirmation of all alerts before insemination.
[3] Averages are presented with results for low and high cash flow farms in parentheses, calculated using the 5th and 95th percentile results for the marginal financial effect of activity meters.

Labor costs. Changing the opportunity costs of labor influenced both the marginal financial effect and IRR (Table 5). If the opportunity costs of labor were either zero or low (€ 9 /hour), the marginal financial effect and IRR were lower than in the baseline scenario. When high opportunity costs of labor (€ 36 /hour) were assumed, the marginal financial effect and IRR were higher than those of the baseline scenario. However, the magnitudes of the effects were relatively small for all labor scenarios; the scenarios differed from the baseline scenario by about € 200 per year for the marginal financial effect and 1% for the IRR.

DISCUSSION

A stochastic simulation model was used, which simulated the reproductive performance of a herd of dairy cows over time. This model was used to analyze the investment in a sensor system with activity meters for the automated detection of estrus. The average IRR of the investment in activity meters was 11% and the most influential inputs in the model were the assumed share of culling for fertility reasons and the increase in sensitivity of estrus detection between the without and the with situation. When a farmer followed strict culling rules and would stop inseminating cows that had not conceived by $WIM = 30$, then the average return on investment was negative. The IRR was also negative when the sensitivity in the with situation was 65% or when a farmer would blindly inseminate his cows upon an estrus alert. In all other scenarios, the average return on investment was positive. As the cow simulation model was based on the previous work of Inchaisri et al. *(2010b)*, the technical results are similar to that study.

The average calving interval was 419 days for a scenario with baseline culling rules and 427 days (data not shown) for a scenario with 'Random' culling rules (i.e. no culling rules for fertility problems). Both of these scenarios assumed a sensitivity of visual estrus detection of 50%. These intervals are in line with the average calving interval in the Netherlands of 422 days *(CRV, 2011)*. The lowest and highest average calving intervals in this study were 385 days and 435 days (data not shown). The lowest interval was observed for the scenario with a VWP of 9 weeks, and the highest interval for visual estrus detection with a sensitivity of 30%. These intervals also seem realistic when compared to the 25th and 75th percentiles of calving intervals for Dutch dairy farms, 391 days and 465 days, respectively *(CRV, 2011)*. In this study the average calving intervals did not resemble extremely long or short calving intervals. This is as expected, as no extreme values were assumed for disease incidence, ovulation rate, conception rate, VWP, and culling rules.

In the current study, farms using visual estrus detection had an average calving interval of 419 days and an average annual milk production of approximately 7,938 kg/cow. For farms with activity meters, the average calving interval was 403 days and the average annual milk production was approximately 8,030 kg/cow. These technical results can be compared to other studies. Inchaisri et al. (*2010b*) reported a calving interval of 407 days and annual milk production of 7,775 kg/cow for a visual estrus detection rate of 50%, and 362 days and 8,068 kg/cow for a detection rate of 70%. As these models differed in some areas, results were not identical. One main difference is that at an estrus detection rate of 70%, the VWP and fertility inputs were also changed in the study of Inchaisri et al. *(2010b)*. Another study, which analyzed different breeding strategies, reported calving intervals varying from 14.1-14.9 months *(Giordano, et al., 2011)*, which corresponds to approximately 420 days. Lastly, another simulation study reported an average of 120-131 days open *(Giordano, et al., 2012)*,

which would result in calving intervals in the range of 400-411 days. The results of the current study are in line with these findings.

Only a few publications on the economics of sensor systems are currently available. The economic benefits of automated BCS *(Bewley, et al., 2010a)*, the breakeven point of online progesterone measurements *(Østergaard, et al., 2005)*, and the potential economic benefits of information technology *(van Asseldonk, et al., 1999b)* have been analyzed. The results of these studies are not comparable to the current study as the focus is on different sensors. Van Asseldonk et al. *(1999b)* quantified the potential benefits of automated estrus detection. The positive financial effect of automated estrus detection was confirmed in our study. However, there are large differences between the two studies, mainly in the methodology, assumptions, and estrus detection systems.

In the simulation of culling and replacement, the framework of the current model had some limitations. An important limitation was the exclusion of herd dynamics, as only one cow place was modeled in each iteration. In reality a cow that is to be culled for fertility reasons will stay in the herd until she is ready for slaughter or is pushed out of the herd when a heifer calves. The unexpected culling of a cow (e.g. the cow dies at parturition) could cause an empty spot in the herd, if no heifer calves at the moment that the cow was culled. The effect of these dynamics on our analysis is difficult to determine. The effect would depend strongly on the WIM in which a heifer would push a cow out of the herd, and on the difference in milk production between a cow in late lactation and a heifer in early lactation. Another limitation was the estimation of slaughter values. We used average slaughter values, which did not take lactation stage into account. Although this methodology simplifies reality, no major biases are expected, as the analysis used differences between average values at the herd level.

One of the most influential assumptions in the cow simulation model was the method used to simulate culling of dairy cows. One method was random culling (used for general culling), where an individual cow was randomly selected for culling using a probability per lactation. The advantage of this random method is that the distribution of cows over various parities remained relatively constant. However, as increased estrus detection rates cause shorter calving intervals, the random culling process will lead to an increase in the annual number of culled cows for the situation with the activity meter. Culling is often not a random process, fertility problems in general, and more specifically failure to conceive, are known reasons for culling *(Brickell and Wathes, 2011, Dechow and Goodling, 2008, Demeter, et al., 2011)*. Some studies have suggested that approximately 20% of the cows were culled because of fertility problems *(Brickell and Wathes, 2011, Dechow and Goodling, 2008)*. With the improvement in estrus detection from using the activity meter, therefore a decrease in the number of cows culled for fertility problems was expected. This positive effect on culling was simulated using culling rules for fertility problems. However, these rules did not capture the complexity inherent in optimizing culling decisions for individual cows in practice *(Demeter, et al., 2011)*.

However, the culling rules in the current study attempted to approximate, given the structure of the model, the approach farmers use in practice rather than the economically optimal option. These culling rules may therefore provide more realistic results than the assumption of optimal culling decisions used in other models (e.g. *Demeter, et al., 2011*) because in practice culling decisions are not always economically optimal.

The results of the sensitivity analysis show that the lowest positive return on investment for activity meters was found at a sensitivity of 70% (IRR 3.5%) (see Figure 3). Whether the investment is profitable in this scenario, depends on the minimal return on investment that a dairy farmer requires. For Dutch circumstances, a DR of 5% might be appropriate as a minimal return on investment. Minimal returns on investment in the Netherlands may be relatively low due to low inflation rates (an average of 2.32% between 1990 and 2012, ranging from 1.1% to 4.5% *(Statistics Netherlands, 2013)* and the attitude of Dutch dairy farmers who in general tend to consider the risks involved in dairy farming to be low and not require high returns. Some Dutch dairy farmers might even consider a DR of 5% to be high. However, farmers in other regions may consider the risks involved in dairy farming to be higher and thus require higher returns. The IRR was considered to be a more flexible estimate of the return on investment than other investment measures, such as the NPV, especially when extrapolating the results of the current study to other countries and circumstances.

In scenarios with high visual estrus detection rates, higher labor inputs were assumed, and this resulted in greater labor savings for the situation with automated estrus detection. The labor inputs used in the current study were based on Roelofs et al. *(2010)*. They reported a sensitivity for visual estrus detection of 94% or 76%, when observations were done twice a day at dusk and quiet times for 60 or 30 minutes, respectively. However, detection rates were lower by 35% to 46% when observations took place twice a day for 30 minutes at milking times. This observation indicates that combining estrus detection with other activities might be less effective than observing the cows while no other work is done. From an economic perspective, spending more time on estrus detection is likely to be less profitable than investing in activity meters.

An increase in sensitivity led to higher milk production and fewer cows culled for failing to conceive. A decrease in specificity caused more false alerts, which meant higher labor and insemination costs. The changes in milk production and cows culled had a larger effect on the marginal financial effect than the higher labor input for confirmation. Therefore, an increase in sensitivity had a greater effect on the marginal financial effect and IRR, than a similar decrease in specificity. However, farmers perceive false positive alerts as a more important problem *(Mollenhorst, et al., 2012)*. Four explanations for this contradiction are possible. Firstly, the survey of Mollenhorst et al. (*2012*) considered mastitis, which is a different condition from estrus and therefore the importance of false alerts could differ. Secondly, the opportunity cost of a farmer's labor could have been underestimated in the

current study. Thirdly, the problem with a false alert may not be the actual time input, but rather the annoyance of checking a cow when nothing is going on. This annoyance could decrease the level of trust in an automated detection system, which would influence the perceived value of the system for a farmer. Fourthly farmer might want to trust the detection system blindly, which is not profitable with the estimated numbers of false alerts in the current study.

Blind insemination was not profitable in this study, because the extra costs for insemination after a false alert did exceed the benefits arising from the use of the activity meter and labor savings. Confirmation of alerts was almost always profitable, as the labor costs for confirming alerts were lower than the extra insemination costs for blind insemination. With a higher specificity this effect became stronger, however the associated lower sensitivity meant that the investment in activity meters was not profitable in these scenarios.

The influence of labor costs on the profitability of an investment in activity meters was small. This small effect was due to small labor input for visual estrus detection that was assumed in the baseline scenario. This was assumed because most farmers spend time on visual estrus detection while doing other tasks in the barn. Therefore, most of the labor input was considered to be a sunk cost. Varying results have been reported for the relationship between labor input and sensitivity of visual estrus detection *(Roelofs, et al., 2010)*. Detection methods using strict scoring systems and time blocks do not represent the current Dutch practice. Therefore, it is difficult to measure how much labor dairy farmers would save in practice by using automated estrus detection. Furthermore, studies on labor and sensitivity of visual estrus detection *(Firk, et al., 2002, Roelofs, et al., 2010)* appear unable to account for the herdmanship of the farmer. Variation between farms in the sensitivity of visual estrus detection with the same labor input is therefore likely. Another important factor is herd size. Compared to a herd of 130 cows, a farmer with a herd of 65 cows is likely to achieve a higher sensitivity of visual estrus detection with a lower labor input.

In practice, variation in detection performance will occur for both visual estrus detection and activity meters. For visual estrus detection, this variation will be related to herdsmenship of the farmer, herd size and workload of the farmer. For activity meters, this variation will depend on the detection algorithm, which may work better on one farm than on another. This effect could be caused by differences in farm structure (barn type, grazing), cow health, and cow behavior. Another potential reason is that the performance reported in studies that use a strict protocol and structured experimental set-up is not comparable to daily practice on a dairy farm.

Sensor systems for automated estrus detection based on cow behavior include activity meters, pedometers, and 3D-accelerometers. Studies on the performance of these systems reported comparable sensitivities and specificities for estrus detection, although gold standards varied *(Rutten, et al., 2013)*. Because the gold standards differed, it cannot be concluded that performances are equal for these three sensor systems. This study focused

on activity meters because their performance was validated with progesterone as a gold standard. In the current study, studies that use progesterone as a gold standard *(Hockey, et al., 2010a, Kamphuis, et al., 2012)* were considered to be the most reliable for evaluating the performance of automated estrus detection based on cow behavior. However, in practice it is possible that the performance of activity meters, pedometers and 3D-accelerometers is comparable. There is no evident reason for one type of sensor to perform better than the others in measuring cow behavior. However, the algorithm used in the sensor system will be important. The results of this study may provide some insight into the profitability of an investment in pedometers and the 3D-accelerometer.

In this study, non-reproductive diseases were excluded, although diseases such as lameness and metabolic problems do influence fertility. Inclusion of these effects could result in more realistic simulations, however it would be difficult to quantify the effects to make them useable for simulation. It would be especially relevant to include lameness because this disease influences cow walking behavior and hence activity *(Bruijnis, et al., 2010)*. Activity meters could be used to detect lameness in addition to estrus, which is likely to increase the revenues from the sensor system *(Rutten, et al., 2013)*.

The current study used a simulation model, which made it easy to explore the effects of changing a single variable, such as estrus detection rate. However, these simulation results can differ from results in practice. It would be interesting to compare the current results to profitability indicators calculated with real farm data. The advantage this comparison is that differences between farms can then be studied. Although the model was parameterized using Dutch data, the results are also applicable to dairy farming in other countries. Generally, the effect of automated estrus detection will be comparable, and investment is also likely to be profitable. However, differences in factors such as production and price levels will change the technical and financial outcomes of the calculations.

CONCLUSIONS

This study analyzed an investment in activity meters for automated estrus detection on an average dairy farm, using a simulation model parameterized with Dutch data. With an average IRR of 11%, investing in activity meters for estrus detection was profitable. Other indicators such as NPV, B/C ratio, and DPBP confirmed this conclusion. The increase in sensitivity that activity meters achieve compared to visual estrus detection (80% compared to 50%) was the most important determinant of the profitability of the investment. Increasing the sensitivity of visual estrus detection by increasing labor input was more expensive than investing in activity meters. The IRR for a small herd of 65 cows (11%) and a large herd of 195 cows (9%) deviated 1% from the IRR for a herd of 130 cows. Given the assumptions in this study, labor had a minor effect on total costs and therefore on the

profitability of the investment. With zero opportunity costs for labor the IRR only decreased from 10% to 9%. The investment in activity meters was profitable in most of the scenarios in the sensitivity analysis. The exceptions were an increase in sensitivity of estrus detection from 50% to 65%, and the use of strict culling rules. The results of this study indicate that investment in activity meters for estrus detection is likely to be profitable for most dairy farms.

ACKNOWLEDGEMENTS

APPENDIX 5.1

This appendix provides a description of the model and its software implementation. This description focuses on the model and provides more mathematical detail than the material and methods section.

Cow simulation model

The cow simulation model was developed in Microsoft Excel. For the stochastic properties of the model, the add-in software @Risk 6 for Excel *(Palisade-Corporation, 2010)* was used. New components of the cow simulation model are described in detail in this appendix. The basic specifications of the model were described by Inchaisri et al. *(2010b)*. These new components of model included: the use of calendar weeks, the culling and replacement of cows, the transition from one lactation to the next, the reproductive cycle, the estimation of feed requirements, the simulation of false positive estrus alerts, and the sensitivities and specificities of estrus detection. An overview of all model inputs is given in Table 1, Chapter 2.

The cow simulation model, which is described schematically in Figure 1, was used to simulate fertility events of an individual cow in weekly time steps. The model simulated multiple lactations, starting with a cow in an assigned parity. In the next lactation, the cow shifted to the next parity unless the cow was culled during the current lactation. In parallel, fertility was simulated with the occurrence of ovulation, which could be delayed by a reproductive disease. In addition, the model simulated the detection of estrus cases. After detection of an estrus case, the cow could be inseminated. After insemination had occurred, the model simulated whether or not the cow became pregnant and the decrease in milk production due to pregnancy.

Calendar weeks

Each iteration of the simulation model started in the first model step ($j=1$) with the first week in milk (WIM) of the first lactation block ($l=1$) of a specific cow. Parallel to each model step (j), WIM (WIM_j), calendar weeks ($WK(cal)_j$) and weeks of pregnancy ($WK(preg)_l$) were modeled. The model simulated multiple lactation blocks (l) in sequence, in the first lactation block milk production characteristics were assigned to a cow.

In order to aggregate milk production within a specific year, the concept of calendar weeks was introduced. A calendar week represented the week of the year in which a cow calved in step $j=1$, and was randomly assigned using a uniform distribution ranging from 1 to 52,

simulating a year-round calving system. From the second time step onwards, a new calendar week ($WK(cal)_j$) was simulated for each time step (j) (equation 1).

$$WK(cal)_j = \begin{cases} if\ culling_l = 0\ and\ \text{WIM}_j \leq WK(preg)_l + (APL - DOP)\ then \\ \qquad wk(cal)_{j-1} + 1 \\ if\ culling_l = 1\ then \\ \qquad WK(cal)_{j-1} + Round(Uniform(-Var_{hef}, Var_{hef}), 0) \\ if\ \text{WIM}_j > WK(preg)_l + 32\ then \\ \qquad WK(cal)_{j-1} + RCD \end{cases} \quad (1)$$

The lactation was stopped when a cow was culled or dried off. For drying off, a dry-off period (DOP) of 8 weeks was assumed. When a cow was culled, a replacement heifer was assumed to be introduced into the herd during a randomly assigned period of 2 weeks before to 2 weeks after the last lactation week of the replaced cow (VAR_{hef}). The moment of drying off was based on an average pregnancy length (APL) of 40 weeks. At the moment of calving, the new lactation started in $WIM_j = 1$. Calendar weeks were continuously counted from the last WIM_j of lactation (l) until the moment of calving in the new lactation ($l+1$). When a cow conceived, she was dried off 8 weeks before the expected calving date. The 'real' calving date (RCD) was simulated using a normal distribution as in Inchaisri et al. *(2010b)*. In each time step, milk production was simulated. The parity and 305-day milk production were randomly drawn from a discrete and a normal distribution, respectively *(Inchaisri, et al., 2010b)*. The 305-day milk production was adjusted for parity influence and average daily milk yield in that WIM_j (MY_{lj}), using Wood's function *(Wood, 1967)*, according to the methodology of Inchaisri et al. *(2010b)*. In parallel to milk production, ovulation, estrus detection, pregnancy, feed requirements, and false estrus alerts were simulated. The model simulated multiple lactation blocks (l) in which calving, culling and replacement were modeled. In the period between conception and drying off, the average daily milk production was adjusted for the week of pregnancy. The moment of calving, in calendar weeks, was determined as the difference between the 'real' pregnancy length (simulated pregnancy length), and the average pregnancy length.

The respective year (k) was calculated by dividing the number of calendar weeks by 52 and truncating this value to the highest integer.

Culling and replacement of a cow

In this study, a distinction was made between general culling and culling for fertility reasons (i.e. the cow failed to become pregnant in due time). General culling was modeled as a

random event, whereas culling for fertility reasons was modeled using culling rules. Whether a cow was culled (CUL_l) or not was determined for each lactation (l).

A set of culling rules was designed so that cows were culled for reproductive failure if they would exceed a certain threshold value. Cows were culled when they were either not pregnant in ($CUL(wim)$) or had ($CUL(ins)$) unsuccessful inseminations (Table 1). After the decision to cull was made, the cow was no longer inseminated and it was culled when average daily milk production dropped below 15 kg.

A cow was culled for general culling when the random drawing from a Bernoulli distribution was 1. The probability of culling was derived from a frequency distribution of Dutch cows over parities *(Inchaisri, et al., 2010b)*. In a specific parity, the number of cows culled was proportional to the decrease in the number of animals in the next parity in comparison to the number in the current parity. Therefore, the difference between the frequency of cows in parity n (P_n) and of cows in parity n+1 (P_{n+1}) was divided by the frequency of cows in parity n (P_n) (equation 2).

$$CUL_l = \begin{cases} if\ parity_l = 1\ then \\ \qquad bernoulli(((P_1 - P_2)/P_1)*ADJ) \\ if\ parity_l = 2\ then \\ \qquad bernoulli(((P_2 - P_3)/P_2)*ADJ) \\ if\ parity_l = 3\ then \\ \qquad bernoulli(((P_3 - P_4)/P_3)*ADJ) \\ if\ parity_l = 4\ then \\ \qquad bernoulli(((P_4 - P_5)/P_4)*ADJ) \\ if\ parity_l = 5\ then \\ \qquad bernoulli(((P_5 - (1-(P_1 + P_2 + P_3 + P_4 + P_5)))/P_5)*ADJ) \\ if\ parity_l > 5\ and\ \leq 8\ then \\ \qquad bernoulli(0.5*ADJ) \\ if\ parity_l > 8\ then \\ \qquad bernoulli(0.6*ADJ) \end{cases} \qquad (2)$$

Full random culling would cull approximately 30% of the cows. This 30% was assumed to represent the total proportion of culled cows. Therefore, the random culling probabilities were multiplied by an adjustment factor (ADJ) to account for the split in general culling and culling for fertility reasons. It was assumed that 20% of the total number of culled cows were culled for fertility reasons *(Brickell and Wathes, 2011, Dechow and Goodling, 2008, Personal Communication, 2013b)*.

The WIM_j in which a cow was culled ($WIM(cull)_j$) was simulated as a random event. A beta distribution with a higher probability of culling early in lactation was used (equation 3). This

distribution was designed by the authors to account for a higher culling probability early in lactation, reflecting the problems resulting from calving and diseases in early lactation *(Rajala-Schultz and Grohn, 1999, Seegers, et al., 1998)*.

$$WIM(cull)_j = round(Betageneral(1.3,5,1,82)$$ (3)

Transition from one lactation to the next

When a cow was culled and replaced, the parity of the cow in lactation (l) was set to 1 (for the replacement heifer). A replacement heifer had a randomly assigned 305-day milk production. When the cow entered a new lactation, the cow's parity was increased with 1. For a cow that shifted from one parity to the next, the 305-day milk production was adjusted for the influence of the increased parity *(Inchaisri, et al., 2010b)*.

Reproductive cycle. Between $WIM_j = 3$ and $WIM_j = 60$ of lactation *l*, the reproductive cycle of the cow was simulated parallel to milk production. The WIM_j of first ovulation and subsequent ovulations, and the occurrence of reproductive disorders that delay ovulation were simulated. Different sensitivities (i.e. estrus detection rates) were assumed for visual estrus detection and for estrus detection with activity meters (Table 1). For a specific cow that was ovulating, the detection sensitivity was adjusted for 305-day milk production and average daily milk yield in the respective WIM *(Inchaisri, et al., 2010b)*. During the voluntary waiting period (VWP) a cow was not inseminated, even if the cow was detected to be in estrus. The average herd conception rate was 50%, for a specific cow the conception rate was adjusted for WIM_j, parity, 305-day milk production, and occurrence of reproductive disorders *(Inchaisri, et al., 2010b)*.

Estimation of feed requirements

The feed requirements in model step *j* and in lactation *l* were calculated based on the milk production in that week. For the calculation of these requirements the first step was to calculate the daily milk yield (MY_j) as fat and protein corrected milk ($FPCM_j$) (equation 4). Dutch average fat and protein percentages of 4.25% and 3.44% were assumed *(CRV, 2011)*.

$$FPCM_j = (0.337 + 0.116 \cdot F\% + 0.06 \cdot P\%) \cdot MY_j$$ (4)

Feed requirements were calculated per week in feed units lactation (VEM) as defined by van Es *(1978)*. The feed requirements (VEM) were calculated in time step *j* and in lactation *l* with equation 5 *(Remmelink, et al., 2012)*.

$$VEM_{lj} = (5{,}323 + 400 \cdot FPCM_j + 0.73 \cdot FPCM_j^2 + ADJ(par)_l + ADJ(preg)_{lj}) * 7 \tag{5}$$

Basic feed requirements were adjusted for parity ($ADJ(par)_l$) and pregnancy ($ADJ(preg)_{lj}$) using the values in Table 1.

Simulation of false positive estrus alerts

A specificity of 100% was assumed for visual estrus detection by the farmer and a specificity of 95% was assumed for the activity meter. A daily probability of a false estrus alert existed during the period after the VWP and before the first week of pregnancy, and in a WIM in which the cow was not in estrus. The occurrence of a false alert in WIM_j (FAL_j) was simulated by a random draw from a Binomial distribution with seven trials (n=7), which accounted for the daily test protocol of the sensor in that WIM (equation 6).

$$FAL_j = \begin{cases} if\ ovulation = 0\ and\ if\ pregnancy = 0\ and\ if\ \mathrm{WIM}_j > VWP\ then \\ Binomial(1 - sensor\ specificity, n = 7) \end{cases} \tag{6}$$

Sensitivities and specificities of estrus detection

In this simulation study, activity meters that are attached to a cow with a neck collar were the analyzed sensor system. An activity meter measures the acceleration of the cows movement in three dimensions. From this data, behavioral parameters are calculated that are used to detect estrus. The performance of such sensors has been reported to have sensitivities in the range of 79% to 94% with specificities in the range of 90% to 98% *(Hockey, et al., 2010a)*. Another study indicated sensitivities in the range of 62% to 77% with a specificity of 99% *(Kamphuis, et al., 2012)*. On the basis of these reported performances, the current study used 80% sensitivity and 95% specificity for activity meters. For visual estrus detection, 50% sensitivity and 100% specificity was assumed.

Chapter 6

A prognostic model to predict the success of artificial insemination in dairy cows, based on readily available data.

C.J. Rutten

W. Steeneveld

J.C.M. Vernooij

K. Huijps

M. Nielen

H. Hogeveen

ABSTRACT

A prognosis of the likelihood of insemination success is valuable information for the decision to start inseminating a cow. This decision is important for the reproduction management of dairy farms. The aim of this study is to develop a prognostic model for the likelihood of successful first insemination. Parameters are considered for the model, which are readily available on-farm at the moment a farmer makes breeding decisions. In the first step, variables are selected for the prognostic model, which have prognostic value for the likelihood of a successful first insemination. In the second step, farm effects on the likelihood of a successful insemination are quantified and the prognostic model is cross-validated. Logistic regression with a random effect for farm was used to develop the prognostic model. Insemination and test-day milk production data from 2,000 commercial Dutch dairy farms were obtained; 190,541 first inseminations from this dataset were used for model development. The following variables were used in the selection process: parity, DIM, days to peak production, production level relative to herd mates, milk yield, breed of the cow, insemination season and calving season, log of the ratio of fat to protein content, and BCS at insemination. Variables were selected in a forward selection and backward elimination, based on the Akaike Information Criterion (AIC). The variables that contributed most to the model were the random farm effect, relative production factor, and milk yield at insemination. The parameters were estimated in a bootstrap analysis and a cross-validation was conducted within this bootstrap analysis. The parameter estimates for body condition score at insemination varied most, indicating that this effect varies most amongst Dutch dairy farms. The cross-validation shows that the prognosis of insemination success closely resembles the mean insemination success observed in the dataset. Insemination success depends on physiological conditions of the cow, which are approximated indirectly by the production and reproduction data that is routinely recorded at the farm. The model cannot be used as a detection model that distinguishes cows that conceive from cows that do not. The model validation indicates, however, that routinely collected farm data and test-day milk yield records have value for the prognosis of insemination success in dairy cows.

INTRODUCTION

Reproduction management has a large influence on dairy farm profitability *(Giordano, et al., 2011, Inchaisri, et al., 2010b)*. Economically important factors are the decision to start with insemination, the voluntary waiting period (**VWP**) *(Inchaisri, et al., 2011b)*, and the decision to stop with insemination of a dairy cow *(Inchaisri, et al., 2012)*. The aim underlying the choice of VWP is to realize a defined calving interval and milk yield at dry off, which requires a cow to conceive before a certain number of DIM. The optimal VWP within a lactation depends on factors, such as daily milk yield, production level, estrus detection rate, and the cow's health status *(Inchaisri, et al., 2011b)*. A prognosis of the likelihood of insemination success is valuable information for the decision to start with insemination, and can therefore improve reproduction management.

One study has focused on prognosis of the likelihood of insemination success *(Shahinfar, et al., 2014)*. Most studies focus on the identification of risk factors associated with insemination success *(Caraviello, et al., 2006a, Gabor, et al., 2008, Inchaisri, et al., 2010a, Inchaisri, et al., 2011a, Lane, et al., 2013, Lof, et al., 2014)*. These studies have identified the following risk factors: milk yield, parity, cow breed, insemination number, lactation stage, calving season, insemination season, use of sexed semen, and information on the course of the previous calving (e.g. whether *dystocia* occured) on the insemination success.

Other factors suggested to influence the insemination success in dairy cows include diseases *(Canu, et al., 2010, Fourichon, et al., 2000)*, BCS, ketosis, SCC, and breed of the cow *(Caraviello, et al., 2006a, Lane, et al., 2013, Lof, et al., 2014)*. An important factor is the BCS, which can be seen as a proxy of the dairy cow's body reserves *(Lof, et al., 2014)*. In particular, a post-partum decrease in BCS, indicating mobilization of body reserves, is associated with lower reproductive performance *(Hernandez, et al., 2012, Roche, et al., 2009)*. In case of extreme mobilization of body reserves, ketosis may occur, which is also associated with lower reproductive performance *(Roche, et al., 2009, Schefers, et al., 2010)*. Ketosis could be included in an analysis of the likelihood of insemination success by using a ketosis score from milk recordings, which is based on ketone bodies in milk *(de Roos, et al., 2007)*, or the ratio of fat to protein content. Many of the variables with potential prognostic value for the likelihood of successful insemination, such as fat and protein content, SCC, lactose content, and ketosis score, have been evaluated as risk factors for insemination success.

Several studies have also shown the importance of farm factors for the reproductive performance of dairy herds *(Schefers, et al., 2010, Shahinfar, et al., 2014)*. For insemination success in individual cows, the herd conception rate in the past 3 months has been suggested as an important predictor *(Shahinfar, et al., 2014)*. Other farm factors found to be influential are VWP, estrus detection accuracy, stocking density, udder health level of the

herd, and use of a resynchronization program *(Schefers, et al., 2010)*. It is therefore important that models, which give a prognosis of insemination success, account for farm effects and quantify these effects.

It is logical to assume that risk factors with a strong association to insemination success are useful for making a prognosis. However, the strength and causality of association give no indication of the prognostic power of a parameter. To implement a prognosis of the likelihood of insemination success in a decision support system, it is important to have insight in the accuracy of the prognosis. This requires knowledge of the robustness of the estimated parameters in a prognostic model, given the variation amongst farms and individual cows. Furthermore, it is important to conduct an independent validation of the prognosis. Shahinfar et al. (2014) conducted a validation for machine learning algorithms and reported an area under the receiver operating characteristic curve (**AUC**) in the range of 0.6 to 0.75.

The aim of this study is to develop a prognostic model for the likelihood of successful first insemination. In the first step, variables are selected for the prognostic model, which have prognostic value for the likelihood of a successful first insemination. The selected variables are readily available on-farm at the moment a farmer makes breeding decisions, and are obtained from data that is currently routinely recorded on Dutch dairy farms. In the second step, the farm effects on the likelihood of a successful insemination are quantified and the prognostic model is cross-validated.

MATERIALS AND METHODS

Available data

Data on inseminations, calving dates, parity and age of the cow, pregnancy checks, lactation length, relative production factor (a factor that represents the production level of an individual cow relative to her herd mates), and test-day milk yields from 2,000 randomly selected dairy farms, covering the period 2009 to 2012, were obtained from the database of the Dutch cattle breeding and milk production recording company (CRV, Arnhem the Netherlands). In 2009 and 2012, 20,268 and 18,682 farms participated in the milk production recording program of CRV, representing 81% and 85% of the total number of farms, respectively. Selected farms were required to have a farm size of at least 50 cows and to make use of the CRV artificial insemination service. Records were removed from the dataset when the data was considered biologically implausible. For example one test day was removed as the lactose percentage was 15%.

The dataset contained 1,403,194 recorded artificial inseminations and 613,668 individual lactations. Only records containing inseminations with a known preceding calving date (so the calving date that started the lactation in which the cow was inseminated at x DIM) were selected; 653,694 recorded artificial inseminations were removed from the initial dataset because no preceding calving date was registered, leaving 749,500 records. The focus in this study was on the first insemination of a lactation. A farmer will choose to start breeding a cow after a certain DIM. If insemination is unsuccessful, a farmer decides to either inseminate the cow again or stop inseminating and cull, rather than choosing to postpone a re-insemination. A prognosis of the likelihood of insemination success is therefore most relevant for the first insemination of a lactation.

Definition of successful and unsuccessful first inseminations

Inseminations recorded in the database were considered to be unsuccessful when, within the current lactation, another insemination occurred within 100 days after the date of first insemination. An insemination was considered to be successful when a calving was registered in the period from 250 to 310 days after insemination, without a subsequent insemination. This definition considered inseminations that led to conception but did not result in calving to be unsuccessful. If neither a calving nor a subsequent insemination was registered after an insemination, the pregnancy status derived from the pregnancy check was used to determine insemination success. The pregnancy check was considered to correspond with the first insemination if it occurred between 20 and 100 days post insemination. If neither a calving, nor a subsequent insemination, nor a pregnancy check were registered after a first insemination, then this insemination was removed from the dataset. This procedure resulted in a dataset containing 365,128 first inseminations.

Fitting lactation curves

Test-day milk recordings were available for 552,235 lactations. It was assumed that the farmers using this prognostic model would either have milk meters in the milking parlor or an automated milking system to record daily milk yields for each cow. However, data on daily milk yield were not available in the dataset used for model development. In addition, the test-day milk yields did not necessarily correspond with the day of insemination. The daily milk yields were therefore interpolated based on lactation curves. Lactation curves were estimated to determine milk yield and characteristics of the lactation curve corresponding with the day of insemination. A nonlinear least squares estimation with the Levenberg-Marquardt algorithm was used to fit lactation curves for individual lactations of dairy cows. The Levenberg-Marquardt algorithm is robust in reaching convergence and produces small standard errors compared to other algorithms *(Dematawewa, et al., 2007)*.

Milk. Using test-day results from the first 250 farms in the dataset, group models were estimated for the following curves: Wood *(Wood, 1967)*, Wilmink *(Wilmink, 1987)*, Dijkstra *(Dijkstra, et al., 1997)*, and Rook *(Rook, et al., 1993)*. Starting values were obtained from the 305-day milk yield models of Dematawewa et al. *(2007)*. The Bayesian Information Criterion (**BIC**) was used to compare the group models on the basis of parsimony and fit. A more parsimonious model was preferred, as individual production curves are fitted on a limited number of test-day recordings. The BIC places more emphasis on parsimony, through the use of a higher penalty for an extra variable added to the model, than other criteria, such as the Akaike Information Criteria (**AIC**).

The Dijkstra curve (Formula 1) had the lowest BIC for milk yield. Previous comparisons of production curves indicated that the Dijkstra curve provided the most accurate estimates of peak milk yield and time of peak milk yield and that smaller models are preferable *(Dematawewa, et al., 2007, Val-Arreola, et al., 2004)*. The Dijkstra curve was chosen as the production curve to model milk yield, based on the BIC in the current study and previous findings. The estimated model coefficients from the group model were used as starting values for fitting individual production curves. These starting values are presented in Table 1.

$$\text{Milk yield (kg / day)} = a * e^{\frac{b*\left(1-e^{-c*DIM}\right)}{c} - d*\dim}$$

(1)

Table 1. Starting values for the Levenberg-Marquardt to fit individual production curves for milk yield, fat content and protein content. Curves for fat and protein used different starting values dependent on the parity of the cow.

Type	Curve	Parity	a	b	c	d
Milk	Dijkstra	All	23.41	0.03557	0.07607	0.002152
Fat	Wood	1	0.7131	0.1743	0.00183	
	Wood	2	1.0558	0.1456	0.00283	
	Wood	3[+]	1.0503	0.1728	0.00329	
Protein	Wood	1	0.5209	0.1818	0.0019	
	Wood	2	0.8257	0.1464	0.00283	
	Wood	3[+]	0.7918	0.1724	0.00329	

Selection criteria for the data were designed in order to fit individual production curves.. Firstly, the test-day milk yield was required to be below 80 kg, as higher values were considered biologically implausible. Secondly, the lactation needed to include at least 5 test-day results. Thirdly, at least one test-day result within the first 60 DIM was required. It was assumed that in long lactations problems like spontaneous abortion have occurred. Therefore, lactations longer than 600 DIM were excluded from the analysis. The selection criteria were met by 440,140 lactations; a lactation curve was fitted for each lactation that

met the selection criteria. Model coefficients for a Dijkstra curve could not be estimated for 39,466 lactations, which resulted in a dataset containing 400,674 lactation curves. The maximum number of iterations for the regression analysis was set at 500 because convergence was reached for all selected lactations with this number of iterations. Milk yield curves were classified as aberrant when the coefficient d was lower than or equal to 0 and coefficient b was smaller than coefficient d, to ensure that the curve resembled the typical convex parabola production curve, similar to the approach of Inchaisri et al. *(2010a)*. In this procedure, 44,538 curves were considered aberrant and therefore removed from the dataset. The procedure of fitting production curves resulted in a Dijkstra curve for milk yield for 356,136 lactations.

Fat and protein content. The fat and protein content of the milk of an individual cow are only known on dairy farms from the results of the last test day. Therefore, the current study uses the results of the last test day before insemination for fat and protein content. However, there is a delay between the moment of insemination and the last milk recording and fat and protein content can change in this period. To approximate the effect of this delay on the prognostic model, the fat and protein content from the milk recordings in the dataset were compared with interpolated fat and protein content. Production curves were used to interpolate fat and protein content, following the same procedure used to fit individual milk curves. However, about half of the fat curves could not be fitted using this procedure.

Previous research has shown that the persistency of fat and protein content differs between heifers and older cows *(Dematawewa, et al., 2007)*. Therefore, a new approach was designed to fit fat and protein curves. Separate curves were estimated for first, second, and third or higher parity cows. The Wood model (Formula 2) *(Wood, 1967)* was chosen, as it is a more parsimonious model and starting values for these parity groups are available *(Dematawewa, et al., 2007)*. With these starting values, group models per parity were estimated for each group of 250 farms. The coefficients of the group models were used as starting values to estimate individual lactation curves; different starting values were used for each parity. This procedure resulted in Wood production curves for 438,337 lactations for protein content, and Wood production curves for 433,029 lactations for fat content.

$$Milk \text{ yield (kg / day)} = a + b * DIM + c * e^{-d*DIM} \tag{2}$$

Data editing

Milk production and insemination data were obtained as separate files and needed to be combined in a single file for further analysis. After combining the insemination records with the estimated production curves (curves for milk, fat, and protein), 279,785 first inseminations remained in the dataset. The number of first inseminations decreased

because at least one of the production curves was not fitted appropriately for 86,343 first inseminations. The estimated production curves were used, following the methodology of Dijkstra et al. *(1997)*, to estimate peak milk yield, day of peak milk yield, and milk yield at insemination. The fitted production curves for fat and protein were used to estimate the fat and protein content at insemination. Subsequently, the log of the ratio of fat to protein content at insemination was calculated.

For SCC, urea, lactose percentage, and ketosis score, the last recorded test day prior to insemination was used as the estimate for these parameters at insemination. Both SCC and urea were not distributed normally and therefore their log transformation was used for further analysis.

For each insemination, the parity of the cow and BCS (score on a 5-point scale) at insemination were available. Furthermore, class variables were determined for hair color of the cow, calving season, and insemination season. Hair color of the cow contained 3 classes: black and white, red and white, and other. The season variables contained 4 classes: spring (March - May), summer (June - August), autumn (September - November), and winter (December - February).

Records that contained one or more missing values for the independent variables were removed, which resulted in a final dataset containing 190,541 first inseminations (first dataset) from 1,475 farms. Ketosis score and urea are optional in the production recording program and therefore not conducted on all farms. Furthermore, ketosis score and lactose percentage were only available for the most recent years of the time period covered by the dataset. The variable SCC also contained many missing values. Therefore, a second dataset was created that included SCC, urea, lactose percentage, and ketosis score. The second dataset contained 59,526 first inseminations from 1,473 farms.

Statistical analysis

All data editing and analyses were conducted in R 3.0.2 *(R Development Core Team, 2008)*, extended with the following packages: dPlyr *(Wickham and Francois, 2014)*, data.table *(Dowle, et al., 2013)*, lme4 *(Bates, et al., 2013)*, reshape2 *(Wickham, 2007)* and minpack.lm *(Elzhov, et al., 2013)*.

Regression model. Before the first statistical analysis was conducted, a correlation matrix was drawn up for all independent variables. Multicollinearity was assumed to be possible for correlations above 0.75; in situations with possible multicollinearity only the most biologically plausible variable was included. Possible multicollinearity was found between fat percentage at insemination and the ratio of fat to protein; only the ratio of fat to protein was used in further analyses.

The potential prognostic value of the available explanatory variables was assessed, based on univariable analyses (data not shown) and previous studies. The order of the explanatory variables in the forward selection was designed according to two factors: potential prognostic value and ease of use, i.e., how easily the explanatory variables can be obtained and integrated in management software on dairy farms. The order was parity, DIM, log of the ratio of fat to protein content, days to peak production, relative production value, BCS at insemination, milk yield, breed of the cow, insemination season, and calving season. The independent variables were included in this order in a forward selection, using mixed logistic regression with a random intercept for farm effect (statistical analysis 1). The dependent variable was the success of first insemination. The AIC was estimated in each step of the forward selection and was used to balance model fit and parsimony. Model fit was considered to be more important than parsimony to select the best prognostic model. The dataset was considered large enough to focus less on parsimony, and therefore the AIC was chosen rather than the BIC. If the AIC decreased, then the respective independent variable was kept in the model in the next step.

For the model that resulted from the forward selection, a backward elimination was conducted. In the backward elimination, the partial contribution of each independent variable to the complete model was estimated as the difference in AIC value between the full model and the model with one variable omitted. The difference in AIC gives an indication of the contribution of an individual independent variable to the fit of the model.

In the mixed logistic regression, the maximum likelihood is approximated by a computationally demanding algorithm. The use of variables with very different scales increases the computational demand, which can cause convergence problems and result in biased parameter estimates. Three variables were rescaled to prevent these problems. The variables DIM and days to peak milk yield were rescaled by dividing all values by 100. Milk yield at insemination was rescaled by dividing all values by 10.

Bootstrap. A bootstrap procedure was applied to the model that resulted from the forward selection procedures of statistical analysis 1 to calculate confidence intervals for the estimated coefficients and estimate the bias (statistical analysis 2). For this procedure, 100 bootstrap iterations were conducted. In each iteration, farms were randomly sampled from the original dataset to produce a training dataset (two-thirds of the number of farms) and a test dataset (one-third of the number of farms). The mixed logistic model was fitted on each training dataset and the estimated coefficients were subsequently stored. Each testing dataset was used for an independent validation of the model. Similar to the methodology of van der Drift et al. *(2012)*, the models were fitted with a random herd effect to correct for clustering of cows within a farm. During the validation, only the fitted values for the fixed effects were used to estimate a prognosis as the random effects would be unknown.

The same procedure of forward selection was repeated on the second dataset (statistical analysis 3), containing all the variables of statistical analysis 1 and the log of SCC, log of urea,

lactose percentage, and ketosis score. The contribution of each independent variable to the model fit was again assessed according to the difference in AIC in a backward elimination. After the forward selection, the bootstrap procedure was repeated for this second dataset (statistical analysis 4).

Validation. In the bootstrap procedure a prognosis for insemination success was estimated for each individual cow for each test set in each bootstrap iteration. Bootstrapping is a powerful validation tool for prognostic logit models *(Grobbee and Hoes, 2015, Steyerberg, et al., 2001)*. The true likelihood of insemination success was unknown so the prognoses could only be validated against the insemination success. Insemination success was a binary variable (successful insemination or not) whereas the prognosis is a probability on a scale of 0 – 100%. For that matter it was assumed that an accurate prognosis would on average approach the mean insemination success of similar cows. Therefore, the prognoses of insemination success were summarized for each group of cows. Groups of cows with similar characteristics were defined based on the categorical variables of the dataset: parity, BCS at insemination, hair color of the cow, insemination season, and calving season. The average likelihood of insemination success and the average insemination success were estimated for each group of cows. The mean likelihood of insemination success and mean insemination success were tested for equality using a paired Wilcoxon signed rank test. The Wilcoxon signed rank test was chosen as it is appropriate for the comparison of the means of two groups when the data is not normally distributed. The test was repeated for groups of at least 100 cows to check whether the inclusion of small groups of cows influenced the difference between the mean prognosis and the mean insemination success. In some groups, all the cows became pregnant or none of the cows became pregnant. The test was repeated without these two groups of cows to check whether their inclusion influenced the difference between the mean prognosis and mean insemination success.

RESULTS

Table 2 shows the descriptive statistics of the variables in the dataset. On average, cows were inseminated for the first time on DIM 96 of a lactation and the average insemination success was 42.8%. The average milk yield at the moment of insemination was 32.9 kg per day.

Table 2. Descriptive statistics of the available data (n=190,541) on insemination and production.

variable	mean	median	sd	p10	p90
Continuous variables					
Cow per farm	129	136	59.946	40	202
Milk yield at insemination	32.87276	32.59781	7.187997	23.81472	42.35871
DIM	96	86	42.428	56	149
Log of the ratio fat to protein	-0.207	-0.21	0.109	-0.34	-0.071
Fat content (%)	4.17%	4.13%	0.647	3.40%	4.98%
Protein content (%)	3.34%	3.31%	0.301	2.98%	3.72%
Categorical variables					
AI success	42.80%	0	0.495	0	1
Parity	2.6	2	1.711	1	5
BCS at insemination[1]	4.2	4	0.931	3	5

[1] The BCS scale contains 9 level, ranging from 1 to 5 in increments of 0.5.

All of the independent variables decreased the AIC of the prognostic model in the forward selection conducted in statistical analysis 1 (Table 3). Therefore all the variables were selected and the model contained the following variables: random intercept for farm effect, parity, DIM at insemination, log of the ratio of fat to protein, days to peak milk yield, relative production factor, BCS at insemination, milk yield at insemination, breed of the cow, insemination season, and calving season. No variables were excluded in the backward exclusion procedure. The variables that decreased the AIC most were the random farm effect (1,653), relative production factor (867), and milk yield at insemination (724). The smallest AIC decrease was observed for insemination season (9), followed by DIM (33) and calving season (53). To give a prognosis of the success of insemination, farm effects are most important and insemination season is the least important.

The parameter estimates for the parametric regression analysis for the prognostic model (statistical analysis 1) and the bootstrapped regression (statistical analysis 2) are shown in Table 4. For the bootstrapped regression, the parameter estimates for all bootstrap iterations are summarized by the mean, median, and standard error. The concept of a bootstrap procedure is that inference of a statistic can be approximated by resampling the true population. The highest bias and largest confidence intervals were observed for the intercept and BCS at insemination (Table 4). In general, the bias amongst the parameter estimates ranges between 0.000 and 0.009. The bias is the approximated difference between the expected value and its population value, and describes how strongly an estimate might be under- or over-estimated.

Table 3. Comparison of Akaike Information Criterion (AIC) values for the forward selection of variables for the mixed logistic model (statistical analysis 1), and the contribution of each variable to model fit (comparison of AIC values in a backward elimination).

Variable added to the model	AIC	Change in AIC after adding a variable[1]	Change in AIC after removing a variable from the end model[2]
Null[3]	260,163	NA	NA
Random[4]	258,012	-2,151	1,653
Parity	256,802	-1,210	146
DIM	256,750	-52	33
Log of ratio fat to protein	256,715	-36	109
Days to peak milk yield	256,638	-77	138
Relative production factor[6]	256,404	-234	867
BCS	256,278	-126	89
Milk yield	255,576	-702	724
Cow breed	255,453	-123	116
Insemination season	255,372	-81	9
Calving season	255,319	-53	53

[1] In the forward selection, variables were added to the mixed logistic model in descending order (starting with the 'Null' model and ending with adding 'Calving season'). In each step, the AIC after adding a variable was compared to the AIC of the previous step.
[2] The end model is the model resulting from the forward selection. In the backward elimination, each variable was removed from the end model and the effect on the AIC was estimated.
[3] Logistic model without random effect and with an intercept as the only independent variable.
[4] Mixed logistic model with an intercept and a random intercept for farm effect.
[6] Factor that represents the production level of an individual cow relative to her herd mates.

Confidence intervals were calculated for the mean value of the parameters estimates from the bootstrap iterations. These bootstrapped confidence intervals were small for most parameter estimates; for instance, the bootstrapped 95% confidence interval was 0.053 - 0.059 for DIM. The low biases and small confidence intervals indicate appropriately estimated coefficients, therefore the prognostic model can generally be expected to provide appropriate prognoses in practice.

Table 4. Parameter estimates for the parametric and bootstrapped mixed logistic regressions, with the estimates of bias and variation from the bootstrapped regression (statistical analysis 2).

Variable	Parametric regression		Bootstrapped regression					
	Estimate	Bias	Mean	Odds	Median	S.E.	C.I. 5%	C.I. 95%
Intercept	-0.369	-0.001	-0.370	0.690	-0.371	0.003	-0.377	-0.364
Parity								
1	Ref[1]		Ref[1]	1.000				
2	0.141	-0.001	0.140	1.150	0.139	0.001	0.138	0.142
3	0.200	-0.001	0.199	1.221	0.199	0.001	0.196	0.202
4	0.217	-0.002	0.215	1.240	0.215	0.002	0.212	0.218
5[+]	0.071	-0.001	0.070	1.073	0.068	0.002	0.067	0.073
DIM (* 100)	0.056	0.000	0.056	1.058	0.057	0.001	0.053	0.059
RFP[2]	1.250	-0.004	1.246	3.477	1.246	0.003	1.240	1.253
DTP yield[3] (* 100)	-0.155	0.001	-0.155	0.857	-0.154	0.001	-0.157	-0.152
RP factor[4]	1.485	-0.002	1.482	4.402	1.482	0.004	1.475	1.489
BCS[5]								
1	-0.059	-0.010	-0.070	0.933	-0.067	0.010	-0.089	-0.050
1.5	-0.199	0.002	-0.197	0.821	-0.199	0.002	-0.202	-0.193
2	-0.108	0.000	-0.108	0.898	-0.109	0.001	-0.110	-0.106
2.5	-0.028	0.000	-0.028	0.972	-0.028	0.001	-0.029	-0.027
3	Ref[1]		Ref[1]	1.000				
3.5	0.034	-0.001	0.033	1.034	0.033	0.001	0.031	0.036
4	0.114	-0.001	0.113	1.120	0.116	0.003	0.108	0.119
4.5	0.184	-0.004	0.180	1.197	0.195	0.007	0.166	0.193
5	-0.083	0.009	-0.074	0.929	-0.073	0.022	-0.118	-0.030
MY[6] (*10)	-0.370	0.001	-0.370	0.691	-0.371	0.001	-0.372	-0.368

Table 4. (Continued)

Cow breed[7]								
Black and white	Ref[1]		Ref[1]	1.000				
Red and white	0.163	0.000	0.163	1.177	0.163	0.001	0.160	0.165
Other	0.281	0.000	0.281	1.325	0.280	0.004	0.273	0.289
Insemination season[8]								
Winter	Ref[1]		Ref[1]	1.000				
Spring	0.023	0.000	0.022	1.022	0.022	0.001	0.020	0.024
Summer	-0.018	0.001	-0.017	0.983	-0.016	0.001	-0.019	-0.015
Autumn	-0.041	0.000	-0.041	0.960	-0.041	0.001	-0.043	-0.039
Calving season[8]								
Winter	Ref[1]		Ref[1]	1.000				
Spring	0.029	0.001	0.030	1.030	0.030	0.001	0.028	0.032
Summer	-0.037	0.001	-0.036	0.964	-0.036	0.001	-0.039	-0.034
Autumn	-0.088	0.000	-0.088	0.916	-0.087	0.001	-0.089	-0.086

[1] Reference level for the variable.

[2] Log of the ratio of fat to protein content at the last milk recording before insemination.

[3] Days to peak milk yield.

[4] Relative production factor, a factor that represents the production level of an individual cow relative to her herd mates.

[5] BCS scored on a 5-point scale at the moment of insemination.

[6] Daily milk yield at the day of insemination.

[7] Breed is defined as black and white (holstein), red and white (holstein or mry), or other. Black and white is the reference level for this variable.

[8] Season is defined as spring (March - May), summer (June - August), autumn (September - November), or winter (December - February). Winter is the reference level for this variable.

Figure 1 shows the bootstrapped variation in the coefficients for BCS at insemination and log of the ratio of fat to protein. The boxplot for the parameter estimates for the ratio of fat to protein is rather small, indicating that the estimates change only slightly across different bootstrap samples. The boxplot for the parameter estimates for BCS at insemination is much larger. Some outliers were also observed outside the boxplots for BCS. This indicates that sampling a different group of farms from the dataset causes variation in the parameter estimates. Furthermore, the boxplots of the parameter estimates for BCS levels of 1.5 to 4.5 show an ascending line. This indicates that a higher BCS at insemination has a positive effect on the likelihood of insemination success. By contrast, the parameter estimate for a BCS of 5 at insemination is lower than the parameter estimates for a BCS of 3.5 to 4.5, indicating a negative effect on the likelihood of insemination success. The boxplot for the parameter estimate was also much larger for a BCS of 5 than for the other levels. The mean estimated coefficients for BCS might therefore not be appropriate for prognostic purposes for some cows.
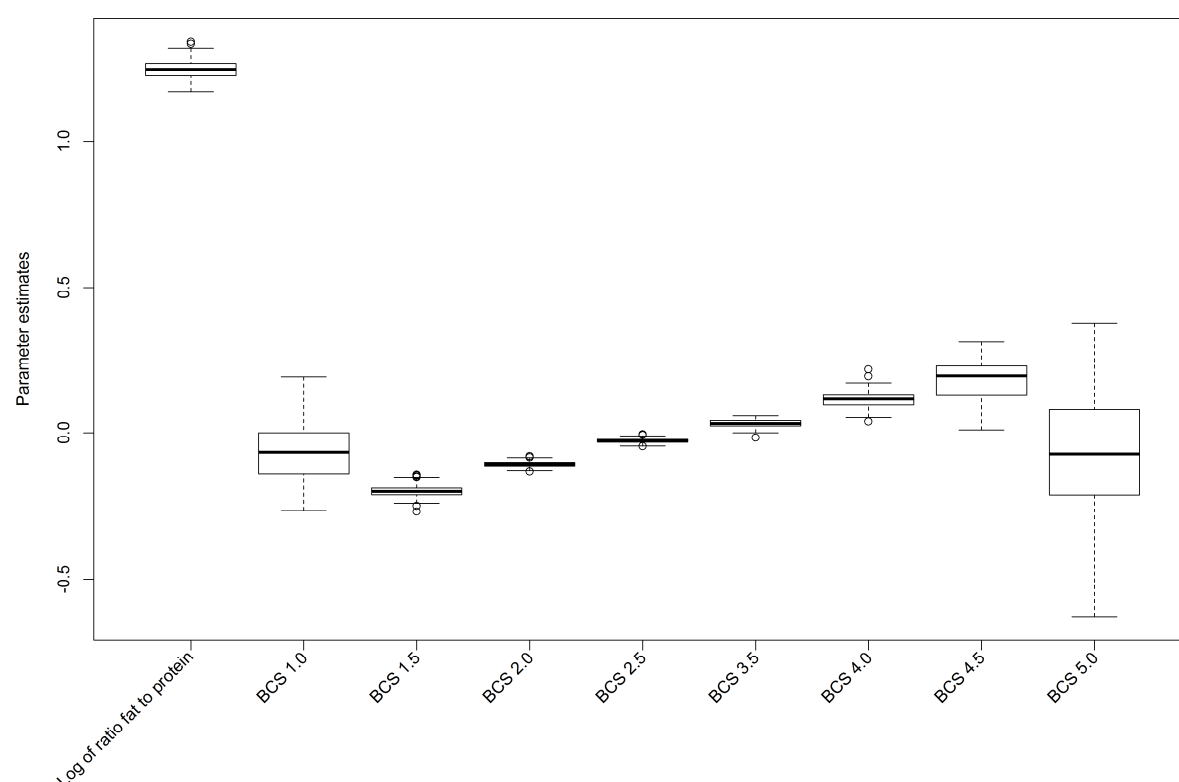
Figure 1: Boxplots of the distributions of the bootstrapped coefficients for two variables in the prognostic model for the likelihood of a successful first insemination: log of the ratio of fat to protein and BCS at insemination.

Three of the additional variables in statistical analysis 3, log of SCC, lactose percentage, and ketosis, decreased the AIC value (Table 5). None of these variables were removed in the backward elimination. In contrast, addition of log of urea content increased the model's AIC, and this variable was therefore not selected for the model. All AIC values in Table 5 are lower than the values from the forward selection in statistical analysis 1 (Table 3) because the models used for Table 5 were fitted on a smaller dataset.

The AIC for the full model with the ratio of fat to protein calculated with values of the previous milk recording was 255,319. The AIC for the full model with the ratio calculated with interpolated fat and protein content was lower, at 254,757, a difference of 562.

Table 5. Continued forward selection and backward elimination (statistical analysis 3) on a smaller dataset for the variables log of SCC, lactose content (%), log of urea, and ketosis score. Model comparison based on Akaike Information Criterion (AIC).

Variable	AIC	Change in AIC after adding a variable[1]	Change in AIC after removing a variable from the end model[2]
End model of statistical analysis 1	100,853	NA	NA
Log of SCC	100,828	-26	18
Lactose %	100,823	-5	3
Ketosis score	100,811	-12	12
Log of urea	100,813	2	NA[3]

[1] In the forward selection, variables were added to the mixed logistic model in descending order (starting with the 'End' model of statistical analysis 1 and ending with adding 'Ketosis score'). In each step, the AIC after adding a variable was compared to the AIC of the previous step.
[2] The end model is the model resulting from the forward selection. In the backward elimination, each variable was removed from the end model and the effect on the AIC was estimated.
[3] This variable was not selected for the prognostic model because the AIC value increased after addition of the variable in the forward selection.

The parameter estimates for the parametric regression analysis for the prognostic model (statistical analysis 3) and the bootstrapped regression (statistical analysis 4) for the second dataset are shown in Table 6. The biases for the variables log of SCC, lactose content, and ketosis score were -0.194, -0.012, and 0.195, respectively. The mean bootstrapped parameter estimates for BCS at insemination ranged from -0.79 to 1.203.

In Table 7, the mean insemination success and mean prognosis of insemination success is presented for each class level of parity, BCS at insemination, hair color, insemination season, and calving season. For each parity group, the mean prognosis is slightly lower than the mean insemination success and all the median prognoses are slightly lower than the mean prognoses. For instance, for the parity group 5^+, the median prognosis is 35.15%, whereas the mean prognosis is 35.59%. For each level of BCS at insemination, there is a slight difference between the mean prognosis and the mean insemination success. The mean insemination success and mean prognosis were higher for a BCS of 4.5 than for other levels of BCS. The difference between the mean prognosis (46.88%) and mean insemination success (46.18%) was -0.70% for red and white cows. This difference was slightly larger for cows with other hair color, where the difference between mean prognosis (52.06%) and mean insemination success (50.82%) was -1.24%. For the five parity groups the probability density of each prognosis on insemination success is described graphically in Figure 2. In each parity group about half of the cows has a prognosis on insemination success above the average insemination success. The probability density function for parity one is slightly smaller than those of other parities.

Table 6. Parameter estimates of the parametric and bootstrapped mixed logistic regressions for the second dataset, including SCC, lactose content, and ketosis score, with the estimates of bias and variation from the bootstrapped regression (statistical analysis 4).

Variable	Parametric regression		Bootstrapped regression					
	Estimate	Bias	Mean	Odds	Median	S.E.	C.I. 5%	C.I. 95%
Intercept	-0.969	0.003	-0.966	0.381	-0.977	0.017	-0.999	-0.933
Parity								
1	Ref[1]		Ref[1]	1				
2	0.110	-0.002	0.109	1.115	0.110	0.002	0.105	0.112
3	0.114	-0.001	0.113	1.120	0.114	0.003	0.108	0.118
4	0.100	-0.001	0.099	1.104	0.102	0.003	0.094	0.104
5[+]	-0.053	0.001	-0.051	0.950	-0.046	0.003	-0.057	-0.046
DIM (* 100)	0.254	0.000	0.253	1.288	0.255	0.002	0.250	0.257
RFP[2]	0.329	0.000	0.329	1.389	0.331	0.004	0.321	0.337
DTP yield[3] (* 100)	-0.078	-0.001	-0.079	0.924	-0.079	0.001	-0.081	-0.077
RP factor[4]	1.372	0.002	1.374	3.952	1.365	0.006	1.363	1.385
BCS ins.[5]								
1	-0.790	-0.027	-0.816	0.442	-0.801	0.024	-0.863	-0.769
1.5	-0.203	0.000	-0.203	0.817	-0.202	0.004	-0.209	-0.196
2	-0.127	-0.003	-0.130	0.878	-0.129	0.002	-0.133	-0.127
2.5	-0.041	-0.002	-0.042	0.959	-0.042	0.001	-0.045	-0.039
3	Ref[1]		Ref[1]	1				
3.5	0.045	0.002	0.047	1.048	0.050	0.003	0.042	0.052
4	0.152	-0.010	0.142	1.153	0.147	0.005	0.132	0.153
4.5	0.186	0.010	0.197	1.218	0.193	0.013	0.172	0.221
5	1.203	0.059	1.262	3.532	1.143	0.053	1.158	1.366
MY ins.[6] (*10)	-0.332	0.001	-0.331	0.718	-0.329	0.002	-0.334	-0.328

Table 6. (continued)

Cow breed[7]								
Black and white	Ref[1]		Ref[1]	1				
Red and white	0.141	-0.001	0.141	1.151	0.141	0.002	0.138	0.144
Other	0.264	-0.003	0.260	1.297	0.264	0.005	0.250	0.271
Insemination season[8]								
Winter	Ref[1]		Ref[1]	1				
Spring	0.117	0.002	0.119	1.127	0.122	0.002	0.115	0.123
Summer	-0.070	0.000	-0.070	0.932	-0.071	0.002	-0.075	-0.066
Autumn	-0.191	0.000	-0.191	0.826	-0.195	0.002	-0.195	-0.187
Calving season[8]								
Winter	Ref[1]		Ref[1]	1				
Spring	0.147	0.000	0.147	1.158	0.148	0.002	0.143	0.150
Summer	0.122	-0.001	0.122	1.130	0.121	0.002	0.118	0.126
Autumn	-0.038	-0.001	-0.039	0.961	-0.040	0.002	-0.043	-0.036
Log of SCC	-0.031	-0.194	-0.225	0.798	-0.226	0.004	-0.233	-0.218
Lactose content	1.190	-0.012	1.177	3.246	1.173	0.034	1.111	1.244
Ketosis score	-0.227	0.195	-0.032	0.969	-0.032	0.000	-0.033	-0.031

[1] Reference level for this variable.
[2] Log of the ratio of fat to protein content at the last milk recording before insemination.
[3] Days to peak milk yield.
[4] Relative production factor, a factor that represents the production level of an individual cow relative to her herd mates.
[5] BCS scored on a 5-point scale at the moment of insemination.
[6] Daily milk yield at the day of insemination.
[7] Breed is defined as black and white (holstein), red and white (holstein or mry), or other. Black and white is the reference level for this factor.
[8] Season is defined as spring (March - May), summer (June - August), autumn (September - November) or winter (December - February). Winter is the reference level for this factor.

Table 7. Observed insemination success (mean value) and the prognosis of insemination success (mean, median, 10th and 90th percentiles), for each group of cows grouped according to the class variables. The average group size is given under n / iteration as group sizes could differ for each bootstrap iteration to another.

Group		AI success	Prognosis				n/iteration
		Mean	Mean	Median	10% percentile	90% percentile	
Parity							
	1	47.41%	47.26%	47.04%	41.40%	53.46%	19,760
	2	43.58%	43.33%	43.12%	36.66%	50.25%	16,725
	3	41.03%	40.74%	40.48%	33.93%	47.82%	11,176
	4	39.61%	39.28%	38.98%	32.43%	46.51%	7,181
	5^+	35.95%	35.59%	35.15%	29.04%	42.66%	8,551
BCS							
	1	38.44%	39.02%	38.92%	30.39%	48.03%	56
	1.5	36.81%	36.80%	36.43%	29.20%	44.92%	1,459
	2	40.12%	40.05%	40.05%	31.89%	48.12%	12,110
	2.5	43.04%	42.78%	42.92%	34.53%	50.63%	29,377
	3	44.22%	43.90%	44.16%	35.40%	51.86%	16,207
	3.5	45.63%	45.21%	45.38%	36.30%	53.64%	3,365
	4	48.09%	47.57%	47.80%	38.77%	56.01%	665
	4.5	50.41%	49.67%	49.92%	40.79%	58.36%	137
	5	44.19%	43.90%	43.55%	32.99%	55.27%	15
Cow breed							
Black and white		41.44%	41.35%	41.60%	33.14%	49.05%	47,728
Red and white		46.88%	46.18%	46.48%	37.70%	54.04%	14,969
Other		52.06%	50.82%	51.30%	42.43%	58.52%	695
Insemination season							
Winter		42.06%	41.73%	41.85%	32.98%	50.11%	16,228
Spring		43.33%	43.03%	43.26%	34.33%	51.22%	16,585
Summer		43.68%	43.53%	43.72%	34.93%	51.69%	13,935
Autumn		42.42%	42.23%	42.31%	33.70%	50.45%	16,644
Calving season							
Winter		43.61%	43.27%	43.49%	34.54%	51.44%	17,333
Spring		44.01%	43.87%	44.11%	35.37%	51.87%	13,740
Summer		42.31%	42.18%	42.21%	33.70%	50.49%	14,929
Autumn		41.61%	41.28%	41.43%	32.69%	49.50%	17,390
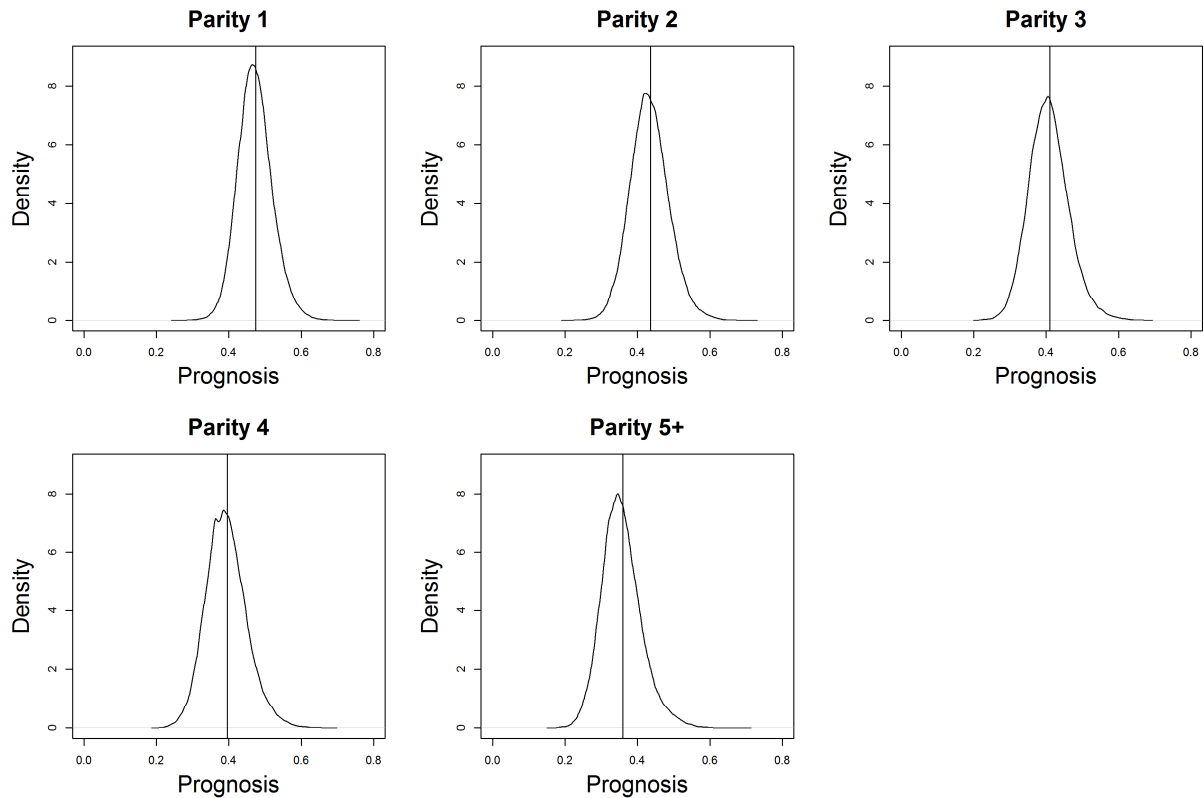
Figure 2: The probability density function of the prognosis for insemination success, with the average insemination success (solid vertical line) for each parity group.

The value of the test statistic, V, was 531,435 for the Wilcoxon signed rank test for statistical analysis 1, with a corresponding p-value of 0.258. For statistical analysis 2, the value of the test statistic V was 367,813 with a corresponding p-value of 0.251. Therefore, no significant difference was observed at the 5% level.

Table 8 shows the median value of some variables in the dataset for each level of BCS at insemination. The BCS between 2 and 3.5 were the largest groups of cows. Cows inseminated at a low BCS (1-2) were in general older than the other cows, parity 3 vs parity 2. In general cows with a lower BCS had a higher DIM at insemination with the exception of a BCS of 5.

Table 8. Median values of variables milk yield, parity and DIM at first insemination with the respective number of records (n) for each level of BCS at insemination.

BCS	n	Milk yield (kg/day)	Parity	DIM (days)
1	168	33.45	3	92
1.5	4,366	34.33	3	93
2	36,372	33.6	3	89
2.5	88,313	32.69	2	86
3	48,713	32.03	2	83
3.5	10,121	30.68	2	82
4	2,004	29.78	2	81
4.5	411	29.42	2	80
5	45	27.63	2	99

DISCUSSION

The overall percentage of successful first inseminations was 42.8%, which is in line with previous studies in the Netherlands *(Inchaisri, et al., 2010a, Windig, et al., 2005)*. The first insemination was on average observed at 96 DIM. Under Dutch circumstances a VWP of 84 DIM was assumed realistic (Inchaisri et al., 2010b), and a first insemination at 96 DIM seems to fit with this VWP. The variables contributing most to the prognosis were farm effect, relative production factor, and milk yield at the moment of insemination. The odds ratios for breed were comparable to those found by Inchaisri et al. *(2010a)*. The other effects are also in line with previous research. For instance, a higher parity, smaller number of DIM, and higher milk yield decrease the likelihood of a successful insemination *(Bello, et al., 2012, Hernandez, et al., 2012, Inchaisri, et al., 2011a, Norman, et al., 2010, Roche, et al., 2009)*. Some deviations in the results compared to other studies can be attributed to differences between the studies, such as the consideration of repeated inseminations *(Inchaisri, et al., 2011a)* or only non-Holstein cows *(Schneider, et al., 2007)*. Furthermore, the current study was focused on the development of a prognostic model and not on a detailed analysis of individual risk factors.

For the current analysis a large dataset was obtained. But, some data were not useful as information missed, like calving dates that mark the start of the lactation in which a cow was inseminated. This reason was mainly caused by the selection of the data for a specific time period. It is unlikely that this caused a selection bias data missed for all cows in the same period. The fitting of lactation curves resulted in an important data reduction as either too few milk recordings were available or a non-aberrant curve were fitted. Both reasons for exclusion were possibly related to health problems. Cows that were culled early in lactation would not have the required number of milk recordings available. Health problems would form a likely reason, as cows that for instance failed to conceive would most likely finish

their lactation before culling. Health problems (e.g. a caesarean section or a severe case of clinical mastitis) cause strong shocks in milk production, which could result in a non-aberrant production curve. Registration of diseases was not available so correction for such shocks and incorporation in the prognostic model was not possible. The reason is that recording diseases is optional for farmers, therefore disease records are not as complete and reliable as the milk recording and insemination data. These two selections of the data cause the model to be unsuitable for cows that experience a severe disease episode. Finally, an important data reduction was the focus on the first insemination within a lactation. This selection of data did fit the aim of the current study best.

Etiological models assess causality of individual factors for a condition (e.g. a disease), whereas diagnostic models determine the condition based on symptoms. Prognostic modeling differs from etiologic and diagnostic models in the sense that the sole focus is to give a prognosis on future development *(Grobbee and Hoes, 2015)*. Prognostic models are therefore mainly evaluated by the accuracy of the prognosis *(Grobbee and Hoes, 2015)*. Furthermore, effects such as confounding are ignored, because they have no relevance for the accuracy of a prognosis. For the current study the recommendations on reporting a prognostic study were considered in a so called TRIPOD statement, which was originally designed for predictive modelling in human medicine *(Moons, et al., 2015)*, see Appendix 6.1. Prognostic models are frequently used in clinical epidemiology, mainly in human medicine. For example, the prognosis of therapy success for colon cancer *(Hugen, et al., 2015)*. The purpose of prognostic modeling is to generate a prognosis (or expectation) of further or future development of a disease. For instance, a prognosis of the success of surgery on a dislocated abomasum in dairy cattle *(Buczinski, et al., 2015, Croushore, et al., 2013)*. A prognostic approach was considered appropriate in this study, because the practical decision-making will be influenced by the expected outcome and the reliability of the expectation, and not by the causal relations that may play a role.

Previous studies have demonstrated the application of machine learning algorithms on the prediction of insemination success *(Caraviello, et al., 2006a, Shahinfar, et al., 2014)*. Machine learning algorithms are powerful solutions to classification problems *(Shahinfar, et al., 2014)*. In general, machine learning algorithms use an adaptive approach of model selection and cross validation to find the model that most correctly assigns individuals to a class *(Hastie, et al., 2009)*. For the prognostic approach explicit classification was not the goal but rather the estimation of a continues probability. Furthermore, logistic regression enables the individual evaluation of variables in terms of model contribution and odds ratio. However, logistic regression requires a separate independent validation, which was conducted in the bootstrap analysis that assessed potential bias as well.

In general, prognostic studies give a prognosis of a disease outcome, such as life expectancy or pain level, on a continuous scale that is divided into classes. The accuracy of the model is then evaluated by analyzing how accurate the prognosis assigns patients to their respective

class. More diagnostic models focus on a binary outcome, such as whether surgery for displaced abomasum for an individual cow was successful. These models give a likelihood, which is then classified on a binary scale. Both methods were considered to be inappropriate for the current model. The focus was on estimation of a prognosis of insemination success, rather than predicting insemination success for individual cows. This means the model estimates a likelihood of success, which could not be validated on individual cow level as the insemination outcome (pregnant or not) was a binary outcome and not a probability. Therefore, the validation was conducted for groups of cows with similar characteristics. This means that groups were defined only based on categorical variables, but all variables of specific cows within these groups were used to estimate the prognosis.

The mean prognosis resembled the mean insemination success, as no statistical difference between the two variables could be observed. The Wilcoxon signed rank test was repeated for groups of at least 100 cows and for a set of groups, where all groups containing either only pregnant or only non-pregnant cows were excluded. The aim of this repetition was to see if these two groups might have influenced the results. No significant difference between the mean prognosis of insemination success and the mean insemination success was found for these additional tests.

The difference in AIC between the two models with different methods for calculating the ratio of fat to protein (interpolated values versus the values of the last milk recording) was larger than the differences observed in the forward selection and backward elimination. This indicates that the time between insemination and the date at which fat and protein content are measured is important for a prognostic model of the likelihood of insemination success. Interpolated fat and protein content cannot be used in practice, as the curve for these factors over the lactation is not available. An alternative to the last milk recording, which more closely approximates the fat and protein content at the day of insemination, could improve the prognosis of the model. Options might be fat and protein content estimated by inline sensors in the milking equipment or a predicted value based on milk recordings and milk yield *(Brandt, et al., 2010)*. The reliability in practice of such alternatives to milk recording data will be important in order to achieve an improvement in the prognosis of insemination success.

In general, the mean and median prognosis of insemination success are some tenths of a percent lower than the mean insemination success. This indicates that the prognostic model tends to slightly underestimate the likelihood of insemination success. However, the differences were considered small, which indicates no problem for application of the model in practice. The underestimations are higher than 0.5% for the classes BCS 4 and 4.5, red and white cows, and cows of a different hair color. Cows with a BCS of 1, 1.5 and especially 5 were in general inseminated later than other cows. The cows with a BCS of 1 had a lower milk production than cows with a BCS of 1.5 which could explain the lower prognosis of insemination success of the latter group. The prognosis on insemination success was still

above 35% for BCS 1 and 1.5 although a low BCS is often associated with a low insemination success *(Roche, et al., 2009)*. Three things are important to note, firstly, the number of records was small for BCS 1 and 1.5. Secondly, the change in BCS, which has a biological relation with fertility, was not assessed and BCS might be less accurate for thin cows *(Roche, et al., 2009)*. Thirdly, farmers might have specific reasons to inseminate these cows regardless of their BCS, for instance because these cows received fertility treatment. The difference between the median and mean prognosis is less than 0.3% for most classes, indicating that the skewedness of the distributions of prognosis is limited and the respective confidence intervals were small. The difference between the mean and median prognosis is relatively high for red and white cows (0.3%) and cows with another hair color (0.48%). This may indicate that cows in these two classes of hair color have a strong influence on the average prognosis. All these findings suggest that the prognostic model produces an useful prognosis of the likelihood of insemination success. Some caution seems appropriate, however, for cows that are not black and white, and cows with an extreme BCS (1, 1.5, 4, 4.5, or 5) at insemination. For these cows the prognosis might be over- or underestimated in practice.

Changes in a cow's BCS are important approximations of a cow's energy balance *(Roche, et al., 2009)*. The cow's energy balance influences her fertility and BCS is, therefore, related to insemination success *(Bello, et al., 2012, Hernandez, et al., 2012, Roche, et al., 2009)*. Only the BCS at insemination was available in the current dataset, which is an estimate at a single moment in time. BCS at insemination did contribute to the prognostic model and its parameter estimates were robust, but this contribution was small compared to other variables. BCS at insemination might not be an accurate approximation of the cow's energy balance, as the change in BCS more accurately reflects the energy balance than the BCS at a single point in lactation *(Roche, et al., 2009)*. In contrast to BCS, the log of the ratio of fat to protein decreased the model's AIC more and thus gives a more direct indication of the cow's energy balance. For a prognostic model, this might mean a better prognosis could be possible if the cows energy balance could be approximated more accurately. Nevertheless, BCS and milk production provide proxies of underlying diseases that affect fertility in dairy cows. Diseases are known to influence insemination success *(Inchaisri, et al., 2010a, Shahinfar, et al., 2015)*.

For farm management the likelihood of insemination success is important for the economic outcome of breeding decisions *(De Vries, 2006, Olynk and Wolf, 2009, Tenhagen, et al., 2004)*. For instance, when the prognosis of insemination success is low (e.g. 0.25), a farmer might want to postpone the start of breeding (depending on the lactation stage) or use a (beef) bull with a higher conception rate. Alternatively, a farmer might want to use sexed semen on cows with a high prognosis of insemination success. Farmers could compare the prognosis on insemination success to the average AI success observed in a group of similar cows, for instance in the same parity as illustrated by the graphs of Figure 2. These decisions could be economically optimized in a decision support model, such as the model developed

by Steeneveld and Hogeveen *(2012)*. Previous work has indicated that the VWP *(Inchaisri, et al., 2011b)*, the decision to stop inseminating cows *(Inchaisri, et al., 2012)* and inseminating only a proportion of the herd with the highest chance on insemination success could be optimized economically *(Shahinfar, et al., 2015)*. Therefore, potential for economically optimized decision support systems for insemination decisions, which would need a prognostic model like the current model exists.

In addition, sensor technology has developed rapidly in recent years and one of the most applied technologies is automated estrus detection, although integrated decision support systems are lacking *(Rutten, et al., 2013)*. Activity meters can improve estrus detection over the detection achieved by the farmer *(Kamphuis, et al., 2012, Rutten, et al., 2014)*. A better estrus detection enables farmers to register estruses automatically and thereby determine more easily whether cows exhibit a normal estrus cycle. Furthermore, a better detection rate increases the chance that a subsequent estrus will be detected and therefore postponing an insemination will have a lower risk of missing the next estrus. The downside of using automated estrus detection is that an automated detection system will produce false positive alerts. Those may cause famers to inseminate cows that are not actually in estrus. A prognostic model could be integrated in a decision support system that uses information of an automated estrus detection system. The information of the prognostic model could be used to produce additional information that could help farmers identify false positive alerts of the automated estrus detection system. One type of automated estrus detection is of special interest, namely sensor that measure progesterone levels in milk *(Kappel, et al., 2007, Østergaard, et al., 2005)*. For a prognostic model these progesterone measurements might not only be useful for estrus detection but also for the prognosis of insemination success. A stronger drop in progesterone level might indicate a better prognosis for insemination success.

CONCLUSIONS

Results of this study indicate that routinely collected farm data and test-day milk yield records have value for the prognosis of insemination success in dairy cows. The prognosis of insemination success closely resembled the mean insemination success observed in the dataset. The most important variables in this prognostic model were the farm effect, relative production factor, and milk yield at insemination. The variables SCC, urea, and lactose percentage also contributed to the prognostic power of the model. Robust estimates can be obtained for most parameters in the prognostic model.

ACKNOWLEDGEMENTS

TRIPOD

APPENDIX 6.1

Section/Topic	n		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Line 10, page 2
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Line 29-55, page 2-3
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Line 60-76, page 4
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	Line 107-113, page 6
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation datasets, if applicable.	Lin 118-128, page 6-7
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	N.A.
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	Line 120-126, Page 6
	5b	D;V	Describe eligibility criteria for participants.	Line 124-128, page 6-7
	5c	D;V	Give details of treatments received, if relevant.	N.A.
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Line 141-152, page 7-8
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N.A.
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Line 219-238, page 11
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N.A.
Sample size	8	D;V	Explain how the study size was arrived at.	Line 239-246, page 12
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Line 239-246, page 12
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	Line 229-238, page 11
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	Line 253-290, page 12-14
	10c	V	For validation, describe how the predictions were calculated.	Line 284-300, page 14-15
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	Line 301-320, page 14-15
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N.A.
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N.A.
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Line 286-290, page 14
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	N.A.
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical	Table 1,

			features, available predictors), including the number of participants with missing data for predictors and outcome.	Line 324-326, page 15
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	N.A.
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	N.A.
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	Table 4 and 6
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Table 4 and 6
	15b	D	Explain how to the use the prediction model.	Line 291-294, page 14
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	Table 7, line 382-401, page 18
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N.A.
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	Line 514-543, page 23-24
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	N.A.
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Line 606-612, page 27
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	Line 563-602, page 25-26
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and datasets.	N.A.
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	Line 616-622, page 27

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Chapter 7

General discussion

This thesis aimed to evaluate the utility of sensor technology to support reproductive management on dairy farms. The potential of sensor technology to enable more precise management of dairy farms could improve farm profitability, animal health and welfare, environmental impact, and risk management, and reduce (manual) labor requirements *(Berckmans, 2008, Ipema, et al., 2011, Lehr, 2011, Wathes, et al., 2005)*. Although the potential is widely acknowledged, only a few applications have been commercialized successfully *(Banhazi, et al., 2012)*. One reason for farmers not to adopt a new technology is that it produces too much information, which farmers do not know what to do with *(Russell and Bewley, 2013)*. To ensure that sensors achieve their full potential, future research needs to shift from technical development towards practical applications and integration with operational farm management.

This general discussion first provides an overview of the main findings of this thesis. Next, the discussion focuses on four main points following from the research in this thesis. The first point addresses the relevance and novelty, to the field of sensor research, of the three new approaches used in the different chapters. The second point discusses the potential value of sensors and the reasons for slow adoption. The third point addresses the importance of connecting sensor systems to the health management of dairy farms. The final point discusses the relevance of societal acceptance of technological developments for sensor research and dairy farming. This chapter finishes with the identification of areas for future research and the general conclusions of this thesis.

Main findings

Chapter 2 proposed and demonstrated a framework for sensor development, in the context of a critical literature overview (Figure 1 represents a simplified version of the framework). The framework describes four levels of development. Level I, technique, deals with whether a parameter can be measured automatically by a sensor. Level II, data interpretation, deals with whether the sensor can provide information on a dairy cow. Level III, integration of information, considers what other information is needed to make a management decision. Finally, level IV, decision making, deals with the ultimate management decision that is supported by the sensor system. Sensors that measure a parameter and hence produce data were described as the first level in sensor development. Further development would include combining data from various sources and integrating available information to produce information or advice that can be used by famers in their operational management.

The structured review showed that many sensors have been developed in recent years and most sensors produce data (level I). Sensor applications for estrus and mastitis detection are the most developed. No sensor systems with an integrated decision support model (level III) were found in the literature. No studies integrated information from a sensor system with

other information (level III), and no studies included the relation between the sensor system and the operational management of a dairy farm (level IV).
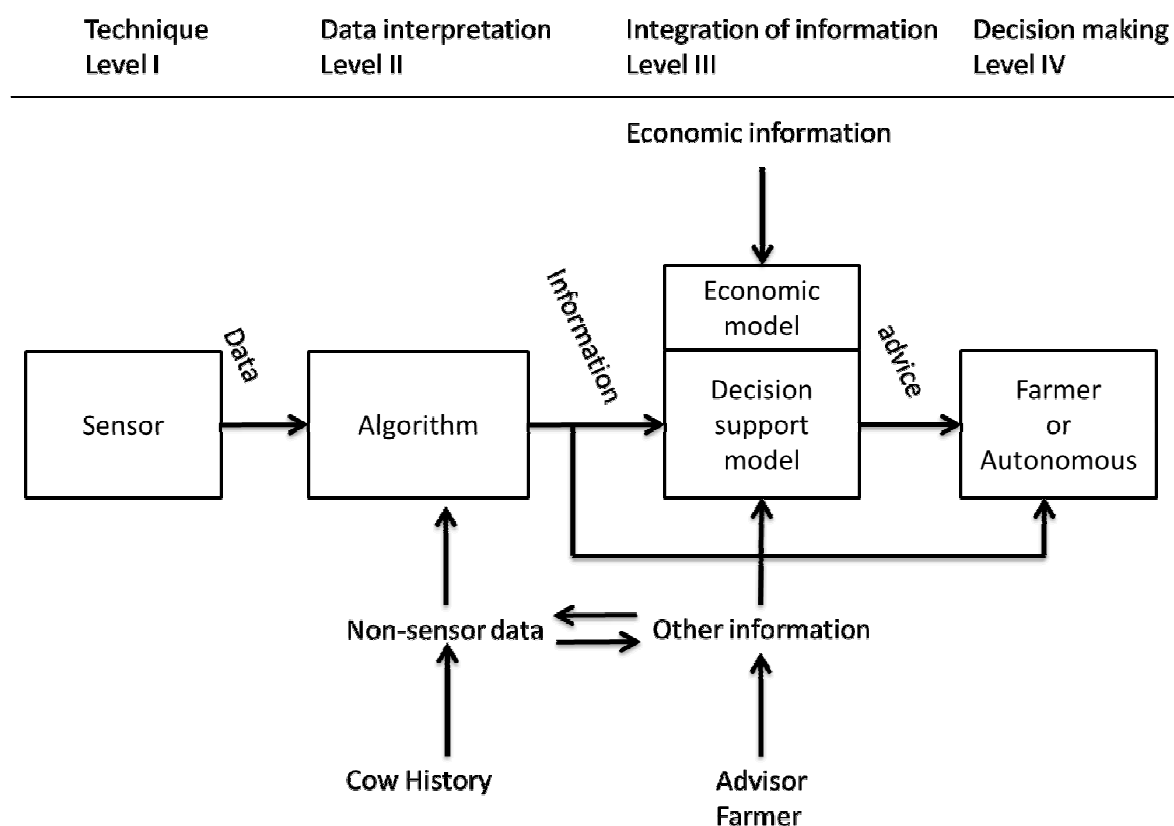
Technique	Data interpretation	Integration of information	Decision making
Level I	Level II	Level III	Level IV

Figure 1. Simplified version of the framework for the use of sensor information in dairy farm management (chapter 2).

The theoretical framework developed in chapter 2 provided a structured overview of the scientific literature on sensors. The overview also identified opportunities for further development, namely decision support and economic benefits of sensors in reproductive management. For the automated detection of mastitis, many applications of electrical conductivity sensors, color sensors, and automated somatic cell count sensors are available. The detection performance (levels I and II) of the electrical conductivity and color sensors studied did not meet the level specified by ISO recommendations. Lameness is an important problem in dairy farming in terms of health, welfare, and economics. Although sensors for lameness detection have been proposed, their detection performance is unclear and these systems have not yet been applied on commercial dairy farms. Some sensors that measure parameters relevant for metabolic disorders, such as rumen pH or rumination, have been tested. At the time the overview of chapter 2 was made, these technologies were not technically ready to be developed into detection systems on commercial dairy farms.

The structured overview in chapter 2 concluded that further improvement of mastitis detection would require novel sensors. For lameness, better detection performance and

better understanding of intervention strategies are required, whereas better sensors and improved understanding of what information these sensors can provide are needed for metabolic disorders. For estrus detection, many sensors have been studied and applied in practice on commercial dairy farms worldwide. The structured overview clearly showed that the area of reproduction management has the most potential to explore the utility of sensors. This thesis therefore focuses on the utility of sensors in reproductive management.

In this thesis, a sensor system to support reproductive management on dairy farms was studied according to the theoretical framework of chapter 2. Level I (technical level) was omitted. Interpretation of data (level II of the theoretical framework) was the focus of chapter 3. This chapter developed a model that predicts the start of calving. Level IV was also considered in chapter 3, as we discussed the relation between the detection model and decision making. Chapter 6 focused on levels II and III: interpretation of sensor data and integration of data from different sources to produce information that can support management decisions. This chapter developed a prognostic model for insemination success that uses routinely collected farm data. Finally, chapters 4 and 5 focused on levels II and IV of the theoretical framework. These chapters estimated the effects on reproductive management of using sensors for estrus and calving detection and the associated economic benefits.

Chapter 2 concluded that sensors for estrus detection based on activity (activity and rumination sensors) are available and that a good performance (levels I and II) is achievable. Chapter 3 developed an additional application for such a sensor, which was already in use for estrus detection. The additional application concerned calving detection. The start of calving was determined based on camera snapshots, and was used as the gold standard for this analysis. The start of calving was determined for 114 calvings using this approach. Compared to the expected calving date, a better prediction was possible when sensor data and the expected calving date were used jointly in a predictive model. Prediction of the start of calving in a one-hour time window at 99% specificity resulted in a sensitivity of 21.2%, a more relaxed time window of six hours increased sensitivity to 48.5%. Many false alerts were generated in the 12 hours prior to the start of calving. Within the group of 35 cows used for validation, 5 cows did not receive a calving alert at all. Chapter 3 concluded that the calving prediction is not a perfect alert and that an exact prediction of the start of calving does not seem possible with this type of sensor.

Chapter 4 developed a stochastic Monte Carlo simulation model to simulate the economics of calving management. The economic benefits of a sensor for calving detection were evaluated and a comparison made between a specialized calving sensor and an estrus detection sensor with calving detection as extra application (as developed in chapter 3). Both sensor systems were profitable, but the specialized sensor decreased the costs of *dystocia* most. Although the estrus sensor had a much lower reduction in the costs of *dystocia*, this was compensated by the lower cost of using the sensor. The main financial

advantage of the estrus sensor was that no extra investment was required. Next to this the estrus sensor was not invasive and did not require additional labor for attachment of the sensor to the cow. Due to a lack of data, many inputs were based on the expertise of the authors or experts, e.g., expected reduction in *dystocia*, stillbirths, and number of days open. Furthermore, the simulations showed large variation in the financial results, which indicates uncertainty about the profitability of sensors for calving detection.

Chapter 5 appraised the investment in activity meters for estrus detection. A Monte Carlo simulation model was used to account for variation in the important parameters influencing fertility in dairy cows. Assuming that detection sensitivity increased from 50% (visual) to 80% (sensor), the investment in activity meters was profitable. A 15%-20% increase in sensitivity would result in a profitable investment. The return on investment was on average 11%. The financial consequences of false alerts were small in all simulations. The most important effects were increased annual milk production, fewer culled cows, and reduced labor costs.

Chapter 6 used routinely collected farm data to build a prognostic model for the likelihood of insemination success. The aim was to develop a prognostic tool to support the insemination decisions of dairy farmers. A reasonable prognosis for an individual cow could be estimated using the tool. The prognosis can help determine a breeding strategy for an individual cow, based on the prognosis and whether this prognosis is above or below the expected average likelihood. Such a prognosis could be integrated with an estrus detection system. Other sensors could also provide useful data for a prognostic model for the likelihood of insemination success. In the model used in chapter 6, data for many of the variables were obtained from milk recordings, which were generally conducted each month. Sensors can monitor such parameters in real time, which means that current data could be used instead of days- or weeks-old data. Promising developments could be fat and protein content, body condition score, or progesterone measurements.

New approaches to study the utility of sensors

This thesis introduced three new approaches to study the utility of sensors. A novel theoretical framework was introduced and demonstrated to evaluate sensor research (chapter 2). Culling rules were introduced in a stochastic dynamic Monte Carlo simulation model to represent the heuristic culling procedures used by dairy farmers (chapter 5). Lastly, a prognostic approach was used to estimate the likelihood of insemination success (chapter 6).

Framework to evaluate sensor development. Chapter 2 developed a framework to structure sensor research. The framework provides a schematic representation of the development of sensor systems in four levels. These levels indicate how advanced a sensor system is and what practical information it provides to a farmer or advisor. Such a framework is new to the field of sensor research. The framework especially highlights the

decision making on dairy farms and the relation between sensor systems and operational farm management. Attention is given to the importance of decision making, the various data sources and advisors that can be involved, and the need to combine data and information in smart systems.

To date, the potential of sensors has not yet been reached. It is therefore important to focus research on sensor systems that support farm management. The framework describes the steps needed to develop sensor systems that support farm management and can therefore be used to design and focus future studies on sensor systems. Furthermore, the structured overview in chapter 2 clearly demonstrated the level to which current sensor systems have been developed. This demonstration is the first structured overview of sensors and also the first overview that evaluated the level of development of sensor systems. The structured overview clearly indicated what is lacking in current sensor research and what is needed to further the development of sensor systems.

Simulation model with rules of thumb for culling. Culling costs are an important factor in studies that estimate the costs of production diseases on dairy farms. Fixed culling costs per cow are generally used in most studies on the costs of mastitis *(Down, et al., 2013, Geary, et al., 2012, Huijps, et al., 2008, van Soest, et al., 2016)*, lameness *(Bruijnis, et al., 2010, Charfeddine and Pérez-Cabal, 2017)*, and reproductive failure *(Hutchinson, et al., 2013, Inchaisri, et al., 2010b)*. These fixed costs are often obtained from studies on the economics of culling decisions, which focus on estimating culling costs or optimizing culling approaches. Dynamic programming models are typically used to estimate an optimal approach *(de Vries, 2004, Demeter, et al., 2011, Gröhn, et al., 2003, Heikkila, et al., 2012, Houben, et al., 1994, Kennedy and Stott, 1993)*. In the economics of culling decisions, many factors are involved, such as milk production, feed consumption, and animal health *(Demeter, et al., 2011)*. Farmers consider these factors in their decision making but tend to make suboptimal decisions in practice *(Rajala-Schultz and Gröhn, 2001)*. This contrasts with the dynamic programming approach, in which optimal decision making is assumed in every step of the decision-making process. Furthermore, the availability of replacement heifers is assumed in these approaches. Optimal decision making is unrealistic and replacement heifers are not available on all farms, which limits the value of the dynamic programming approach.

Reproductive performance is an important factor in culling decisions. Rajala-Schultz and Ghrön *(2001)* showed that the sooner in lactation a cow became pregnant the lower the culling risk. Chapter 5 used a simulation model based on culling rules. For example, the rule that a cow was culled if she failed to conceive within 35 weeks in milk. The model thus captured actual practice, as farmers are more likely to use a rule of thumb to determine whether a cow has not conceived in time and should be culled. A similar approach was used in the Danish Simherd® model. This model allows for a longer breeding period, namely 301 or 259 days in milk (**DIM**) for high- and low-yielding cows *(Ettema, et al., 2011)*. This is more

liberal than the 245 DIM we assumed in chapter 5 and would result in different culling outcomes.

Culling rules are important in the evaluation of automated estrus detection because an improved estrus detection will lead to a higher likelihood of conception and thereby a lower likelihood of culling. This is supported by the results in chapter 5. For instance, less strict culling rules increased the average return on investment by 7 percentage points, whereas more strict culling rules decreased this return by 23 percentage points (chapter 5). The decisions were not optimized and the costs were estimated specifically for each cow. This approach aimed to provide a realistic estimate of culling costs.

Prognostic model for insemination success. Chapter 6 proposed a prognostic model for the likelihood of insemination success. In etiological studies regarding fertility, much emphasis has been placed on identifying risk factors *(Giordano, et al., 2012, Inchaisri, et al., 2010a, Shahinfar, et al., 2014)*. These risk factors and their respective odds ratios provide important information. Factors that influence fertility or disease occurrence can potentially be influenced by farm management. Risk-factor analysis is therefore important input for the design of control and improvement strategies for dairy farms. However, a different focus is relevant in the context of sensor research and decision support. The goal here is to generate information to be used in decision making. A prognosis provides information on the expected outcome or progression and can therefore be used to optimize the chosen intervention. A prognosis is therefore important information for decisions on animal health and reproduction.

The prognostic approach deviates from etiological studies because it focuses on predicting how likely an expected outcome is given the animal's current status. Prognostic research therefore focuses on prediction, the accuracy of prediction, and how much predictive value individual factors have. Most studies that analyzed farm data have been etiological. Some examples of prognostic modelling exist in veterinary medicine *(Buczinski, et al., 2015, Croushore, et al., 2013)*. However, using farm data for prognostic modelling with the aim of generating management information is novel.

The approach and line of reasoning used for prognostic modelling have relevance for sensor research in general. It is important to consider whether information is available ex ante or ex post. The goal of the system should be clearly defined: what should the system detect and which management actions are connected to this. These considerations have implications for the type of model: prognostic or more diagnostic. They also affect the time window, i.e., shorter or longer, that can be used in the validation of the sensor system. We demonstrated this reasoning with the prediction of the start of calving in chapter 3. The prevention of *dystocia* is important for the health and welfare of the cow and her calf. Therefore, a diagnostic model that detects whether the cow is calving or has calved is of little use. A prognostic model that predicts whether calving will start is more suitable. The expected calving date was information that was available ex ante and therefore included. Other

relevant factors, such as calf size, gender, and whether it was a singleton calf, were only available ex post and therefore ignored.

Utility of sensors

Precise management of individual cows. The potential of sensors to increase profitability on dairy farms, reduce labor input, improve animal health, and reduce environmental impact has been widely recognized in the field of sensor research *(Berckmans, 2008, Lehr, 2011, Wathes, et al., 2005)*. One of the main ideas behind the development of sensors is that the use of sensors in livestock farming enables precisely optimized management of individual animals *(Berckmans, 2008)*. The Dutch dairy sector has, for instance, shown great interest in these developments and started the research program Smart Dairy Farming in 2010 *(Anonymous, 2017)*. Individual and precise management can only be achieved if interventions can be tailored to the individual animal. In estrus detection, for instance, the decision to inseminate or not is a binary choice without possibilities for individual tailoring. However, the voluntary waiting period can be individually optimized, based on the prognosis of insemination success, production level, and expected persistency.

When standard protocols are used for every problem, the use of sensors is not likely to result in individually optimized management. With increasing herd sizes, farmers generally start to use more standardized protocols for groups of cows. A notable example is the timed AI protocol, which is used frequently in the USA as a breeding strategy. In larger herds, visual estrus detection becomes too labor intensive and therefore a structured approach has advantages. In contrast, Israeli dairy farmers predominantly use sensors to detect estrus *(Galon, 2010)*. Their reproduction results are comparable with those of farmers using timed AI protocols *(Dolecheck, et al., 2016b, Galon, 2010, Neves, et al., 2012)*. Results of a deterministic model suggest that automated estrus detection is more profitable than both visual estrus detection and timed AI substitution *(Dolecheck, et al., 2016a)*. This finding is consistent with the findings in chapter 5 of this thesis.

Economic benefits of sensors. Despite the many sensors developed for dairy farms, chapter 2 showed that only two types of sensors are applied in practice, namely sensors for estrus detection and sensors already included in an automated milking system (**AMS**) *(Steeneveld and Hogeveen, 2015)*. Adoption rates for other sensors are low. A possible reason for this low adoption rate is uncertainty about the economic benefits of investing in sensors. Surveys of farmers in the Netherlands and the USA indicated that the benefit-to-cost ratio, simplicity and ease of use, and availability of local support are important aspects for farmers in their decision to adopt sensors *(Borchers and Bewley, 2015, Steeneveld and Hogeveen, 2015)*. These surveys clearly indicate that economic considerations are important to farmers. Although the economic benefits of sensors have been suggested for years *(Hogeveen, et al., 1994, van Asseldonk, et al., 1999a, van Asseldonk, et al., 1999b)*, only a

few studies have actually estimated the potential benefits (e.g., the technical benefits for estrus detection *(Dolecheck, et al., 2016b, Galon, 2010)*).

Chapter 5 used normative modelling to quantify the economic benefits of an estrus detection system and showed investment to be profitable. The potential economic benefits were also quantified in a normative way for calving detection by sensors (chapter 4). Only Steeneveld et al. *(2015)* have investigated the economics of sensor investments in an empirical way. Based on accountancy data, they found that sensors did not significantly improve technical efficiency on Dutch dairy farms *(Steeneveld, et al., 2015)*. Furthermore, they found no economic effect, including no difference in labor costs. The potential of automated estrus detection that was shown in a normative way in chapter 5 is not supported by the analysis of accountancy data.

The reasons for the differences found between normative and empirical analyses are discussed in the following section. This discussion focuses on the AMS, which is the main application of precision livestock farming (**PLF**) in dairy farming and more extensively studied than individual sensors. The AMS serves as an example of sensor development. Furthermore, the discussion focuses on labor as this is an important factor in the decision making of dairy farmers. In this context, the utility of sensors is broader than only labor costs. Utility also captures other factors, such as leisure, family time, and peace of mind because a system monitors the cows.

Labor benefits of the AMS. Normative studies have suggested that an AMS could potentially improve milk production and reduce labor, and therefore be a profitable investment *(de Koning, 2010, Hyde and Engel, 2002, Jacobs and Siegford, 2012)*. In contrast, empirical studies have concluded that investment in an AMS provides no financial gain compared to a conventional miling system (**CMS**) *(Bijl, et al., 2007, Steeneveld, et al., 2012)*. Furthermore, labor costs were observed to be similar for CMS and AMS farms (Steeneveld et al., 2012). Despite the inconclusive evidence on the economic benefits, the AMS continues to be adopted on an increasing number of commercial dairy farms throughout the world *(Barkema, et al., 2015, de Koning, 2010)*. This suggests that other benefits are also important in the decision to invest in AMS. Farmers have stated that important reasons to invest in an AMS are more flexibility, which enables a more modern lifestyle, and reduction in production costs *(Floridi, et al., 2013, Hansen, 2015, Mathijs, 2004)*. This suggests that farmers who adopt an AMS experience the practice of milking two to three times a day in a CMS as a strain on their (working) life. Famers indicate that the AMS gives them more time for family and leisure *(Mathijs, 2004)*, and enables them to get more sleep and be better rested *(Hansen, 2015)*. These benefits form the more intangible parts of utility, which are excluded from current studies that estimate the economic benefits using tangible labor costs.

Accountancy data is unable to record a lower workload or increased flexibility. Empirical studies based on accountancy data cannot quantify such non-monetary benefits, which

results in a negative economic outcome. In normative studies (such as in chapters 4 and 5), a value needs to be assigned to labor savings, usually by using the opportunity cost of labor. The opportunity cost represents the income a farmer could earn by an alternative allocation of labor. With regard to the reasons stated by farmers in adoption studies, the opportunity costs in the literature do not capture the value of increased family time or leisure. It seems fair to state that economic evaluations of the AMS use concepts of labor costs that do not quantify the benefits that are important for farmers. To appraise the utility of sensor technology, it is therefore important to quantify all benefits, especially labor aspects, realistically.

Labor benefits of sensors. For the application of sensors, it seems logical to expect parallels with the adoption of the AMS. Dutch dairy farmers indicated in a recent survey that sensors have reduced the amount (hours) of labor used per cow *(Steeneveld and Hogeveen, 2015)*. Similar observations have been made in the USA *(Borchers and Bewley, 2015)*. Accordingly, reduced costs because of labor savings were assumed in chapters 3 and 4. Effects on labor seem to be an important factor for farmers in the decision to adopt sensors. In the field of sensor and robotics research, agreement exists that the focus of sensor technology should lie with the processes that influence farm profit most, and that applications should monitor and control these processes in the financially most optimal way *(Banhazi, et al., 2012, Wathes, et al., 2005)*. The labor reduction and changes in workload that farmers seem to prefer are not by definition linked to the farm processes that influence profitability most. A reason for the low adoption of sensor technology may therefore be that the development of sensors aims at the wrong applications, benefits, and goals. These benefits could be reducing the amount of labor or changing the type of labor e.g., to enable a more flexible way of working, to replace physically heavy work or annoying tasks by different types of work, or providing peace of mind by ensuring that a process is monitored adequately.

Supporting reproduction and health management with sensors

The potential benefits of sensors will only be gained by farmers if they use sensors in their operational management. As argued previously, a sensor system should provide information that is useful and suitable for operational management (chapter 2). On the other hand, the farmer also needs to use or act upon the information originating from sensor systems. If health problems or potential problems are signaled, the farmer should be willing and able to react to these signals.

Sensors can monitor cows constantly and thereby produce more detailed information than visual observation. However, sensors are unable to solve management problems on a farm, such as suboptimal feed, housing, or treatment strategies. The only value of sensors is in signaling problems more often, earlier, or both. Sensors can help to improve management in

terms of timely intervention and enable more detailed and individual management. An example is lameness detection, where many systems can detect cows with high lameness scores well (chapter 2). However, cows with a lameness score of four or five are severely ill animals *(Thomsen, et al., 2008)*. These cows can hardly walk or only walk with visible pain and discomfort. A farmer who does not notice these cows might have poor animal skills or is not paying attention. If a herd is so large that the farmer misses severe cases of lameness, this might be an indication that the farmer is managing too many cows, without enough labor. A useful sensor system for lameness should detect mildly lame cows or cows that are likely to develop severe lameness, so the farmer can intervene to prevent (severe) lameness. Current sensor systems use one physiological or behavioral parameter and register changes (chapter 2), without consideration of the broader context, obtained knowledge, and previous experience. Good herdmanship is therefore still required to gain the full potential from current sensor systems. In the future, sensor systems may be able to automate some of these capabilities and provide more support to farmers.

Societal acceptance of technological development

Perceptions of animal welfare and good husbandry differ between farmers, veterinarians, and scientists on the one hand, and the general public and animal rights activists on the other *(Barkema, et al., 2015, Bergstra, 2014)*. Increased herd sizes and technological developments have strengthened the perception gap between farmers and the general public. For instance, the Dutch animal rights group "Wakker dier" claims that increased herd sizes and AMS reduce the farmers' time and attention for the welfare of their cows and that technology is used in an attempt to substitute for this lack of care *(Anonymous, 210)*. Many issues are under debate. Examples include the naturalness of housing systems, the use of pain management in dehorning, and access to pasture *(Anonymous, 210, Barkema, et al., 2015)*.

Dutch animal rights activists claim that the five freedoms for farm animals *(Anonymous, 210)* are severely violated on a daily basis in dairy farming and blame this on increased herd size, in which technology (especially the AMS) plays a role *(Anonymous, 210)*. A license to produce is of vital importance for the future of dairy farmers in western society. Therefore, the general public needs to trust and understand the production methods and farm practices used. Three issues are regarded as important by Western-European citizens and are therefore important for dairy farmers to take into account. Firstly, the freedom for dairy cows to express natural behavior, specifically access to pasture *(Barkema, et al., 2015, Millar, 2000)*. Secondly, proper care of the animals to prevent diseases and discomfort and to resolve problems as quickly as possible *(Barkema, et al., 2015)*. Thirdly, it seems important to manage dairy cows as individual animals.

In the light of societal concerns, sensor technology is a development in dairy farming that needs to be utilized carefully. As sensors monitor behavioral or physiological parameters of an individual dairy cow 24 hours per day, it is possible to monitor individual animals. The case of estrus detection is a clear example of how sensors can replace standardized protocols that manage cows in groups. For the future of dairy farming in western countries, characterized by increasing herd sizes, sensors offer the opportunity for individually optimized management, rather than group management based on the average animal.

However, it could also be argued that the farmers themselves are not observing the animals, thereby reducing the animals to mere objects represented by graphs and numbers on the computer. This is one of the critiques animal rights activists use in their campaigns, that cows are treated as mere objects in an industrial setting *(Anonymous, 210)*. The romantic notion of a farmer who knows all the cows by name is an important sentiment in the discussion on individual cow management. It seems unlikely to make a difference for the cow whether the farmer observes her personally or uses technology to do so *(Bos and Munnichs, 2016)*. Ethically speaking, it is therefore not the sensor itself that is an issue, but the way it is used *(Bos and Gremmen, 2013, Millar, 2000)*. For future developments, it seems important to consider the concerns of society by giving attention to welfare and ethical aspects *(Millar, 2000)*.

If not used sensibly and communicated smartly, sensor technology can be framed as industrialized livestock farming. For sow farming, Bergstra et al. *(2014)* showed that knowledge of and commitment to farming was associated with a less negative attitude. Therefore, it would seem wise for the dairy sector to communicate openly about PLF applications with citizens and regard their values. For successful implementation of sensors, both the needs of farmers and the perceptions of society are important *(Wathes, et al., 2005)*. Ethical considerations and consideration of the concerns of citizens and the way sensors are used, are all vital for societal acceptance.

Only using technical arguments and considerations to explain the benefits of sensor technology could harm the image of sensor technology and subsequently its acceptability to the general public. Emphasis, especially in public debate, should therefore be on the opportunities PLF provides to improve the welfare and health of dairy cows and, in this regard, its added value over human observation, rather than on improved production and increasing herd sizes. The Dutch Rathenau institute has, for instance, emphasized the opportunities of PLF to manage animals on a more individual basis and to intervene more directly, which would also benefit the animal itself, rather than the possible concerns *(Bos and Munnichs, 2016)*.

Future research

If sensor technology is to be successful in dairy farming, farmers need to be able to see the benefits of the technology, adopt it, and be able to improve their management by using it. Labor is an important issue for farmers in their decision to adopt new technology. Future research on sensors should consider which tasks consume much (physical) labor or could be made more flexible. Sensor systems that help farmers with such labor issues are more likely to be successful. This first requires a better understanding of how farmers regard their farm tasks and work load. Furthermore, for an accurate economic evaluation of sensor technology on dairy farms, a better method should be developed to quantify the labor benefits of sensors. Farmers clearly state that labor is important and especially mention flexibility. Labor cannot therefore be ignored in economic evaluations. However, current analyses and methods that use the opportunity cost of labor or accountancy data seem unable to capture the value of more flexibility, leisure, and family time.

A further aspect for future research concerns the type of information provided to farmers. Farmers want information that they can interpret and use. Decision support is an approach that can integrate information and provide advice to farmers. Some recommendations can be made to develop sensor systems that produce useful and understandable information for farmers or their advisors. Data mining and big data approaches can help to obtain information from sensor data, because with such approaches new patterns in sensor data may be found and exploited. For instance, an important goal in the research program Smart Dairy Farming was the development of a data portal that enables exchanges of sensor data between different stakeholders *(Anonymous, 2017)*. However, it remains important to choose an appropriate gold standard in such studies. The gold standard should be a reflection of the information one wishes to obtain from a sensor system. Furthermore, the ideas and concepts of prognostic modelling, as used in epidemiology, are valuable for sensor research. The information (prognosis) farmers obtain from a system is what they use in their operational management. The prognosis given by a sensor system should therefore be carefully developed and validated. For successful application in practice, sensor systems need to be developed that provide integrated decision support rather than only measure physiological or behavioral parameters.

A final avenue for future research on sensor technology is societal acceptance. It is important for the whole dairy sector that society finds production methods acceptable. No principle ethical or welfare issues have yet been raised regarding sensor technology. However, animal rights activists may frame developments such as sensor technology as a more industrial form of dairy farming. Future research that focuses only on technology and improving production may hamper societal acceptance. A better understanding of how citizens regard sensor technology in dairy farming is therefore needed. Future research and commercialization should, in this context, also focus on clearly presenting tangible advantages of sensor technology for animal health and welfare.

Conclusions

We demonstrated that utility of sensors for reproductive management can be found in economic benefits (estrus and calving detection), reduction of labor (calving and estrus detection), and more detailed management information (prognosis of insemination success). We therefore conclude that automated estrus detection aids reproductive management. From this thesis the following conclusions can be drawn:

The developed theoretical framework describes four levels of sensor development, which should all be included for proper development of sensor systems. The literature review showed that no studies developed sensor systems with regard to management and decision support.

It was possible to improve the prediction of the start of calving compared to a model that only uses the expected calving date. However, predicting the start of calving within an hour was not possible with a high sensitivity and specificity.

There was financial merit in the use of calving detection, because the sensor system enables more timely intervention by the farmer. The uncertainty about the positive effects was large, which caused a wide range in the simulated financial benefits.

Investment in a sensor for estrus detection was on average profitable with a return on investment of 11%. Profitability was influenced most by the culling strategy and the expected effect of the sensor on the estrus detection rate, compared to detection by visual observation.

Routinely collected farm data can be used to estimate a prognosis of insemination success and be used to determine whether an individual cow has a higher or lower than average likelihood of insemination success. Integration of this prognostic model with an estrus detection sensor has potential.

Farmers currently only adopt sensors for estrus detection, or sensors that are standard with an AMS. A reason for this is that sensor systems do not produce clear information for farmers. Sensor technology should focus on management support of applications. Labor benefits of sensors are important for the adoption of sensors by farmers; farmers value flexibility, increased family time, and less physical workload as benefits. However, economic evaluations of technical solutions are currently unable to quantify these benefits. Sensor research should consider the preferences of farmers regarding labor. New methods to value the labor benefits of sensors are needed for the appraisal of sensor technology. Societal acceptance should be an important consideration in sensor development. Animal rights activists may frame the use of sensors as a form of industrialized farming. Societal acceptance of modern dairy farming will be hampered by only using technical arguments and considerations to explain the benefits of sensors. Application of sensors on dairy farms should be communicated smartly to society in terms that relate to the values of citizens.

References

CBS. (Statistics Netherlands, CBS 2013. Dutch statistics on: Consumer prices; inflation, Accessed: June 20, 2013.

M. Akerstedt, L. Bjorck, K. P. Waller and A. Sternesjo, 2006. Biosensor assay for determination of haptoglobin in bovine milk. Journal of Dairy Research 73, 299-305.

J. I. Alawneh, N. B. Williamson and D. Bailey, 2006. Comparison of a camera-software system and typical farm management for detecting oestrus in dairy cattle at pasture. New Zealand Veterinary Journal 54, 73-77.

O. AlZahal, H. AlZahal, M. A. Steele, M. Van Schaik, I. Kyriazakis, T. F. Duffield and B. W. McBride, 2011. The use of a radiotelemetric ruminal bolus to detect body temperature changes in lactating dairy cattle. Journal of Dairy Science 94, 3568-3574.

O. AlZahal, E. Kebreab, J. France and B. W. McBride, 2007a. A mathematical approach to predicting biological values from ruminal ph measurements. Journal of Dairy Science 90, 3777-3785.

O. AlZahal, B. Rustomo, N. E. Odongo, T. F. Duffield and B. W. McBride, 2007b. Technical note: A system for continuous recording of ruminal ph in cattle. Journal of Animal Science 85, 213-217.

O. AlZahal, M. A. Steele, E. V. Valdes and B. W. McBride, 2009. The use of a telemetric system to continuously monitor ruminal temperature and to predict ruminal ph in cattle. Journal of Dairy Science 92, 5697-5701.

G. Andre, P. B. M. Berentsen, B. Engel, C. J. A. M. de Koning and A. G. J. M. Oude Lansink, 2010a. Increasing the revenues from automatic milking by using individual variation in milking characteristics. Journal of Dairy Science 93, 942-953.

G. Andre, P. B. M. Berentsen, G. Van Duinkerken, B. Engel and A. G. J. M. Oude Lansink, 2010b. Economic potential of individual variation in milk yield response to concentrate intake of dairy cows. Journal of Agricultural Science 148, 263-276.

G. Andre, B. Engel, P. B. M. Berentsen, G. Van Duinkerken and A. G. J. M. Oude Lansink, 2011. Adaptive models for online estimation of individual milk yield response to concentrate intake and milking interval length of dairy cows. Journal of Agricultural Science 149, 769-781.

Anonymous, 210. Vijf voor twaalf voor de nederlandse melkveehouderij - wake up call voor melkveesector en maatschappij, Stichting Wakker Dier, Online, Amsterdam.

Anonymous, 2017. Smart dairy farming, www.smartdairyfarming.nl

L. H. Aramyan, A. G. J. M. Oude Lansink and J. Verstegen, 2007. Factors underlying the investment decision in energy-saving systems in dutch horticulture. Agricultural Systems 94, 520-527.

H. U. Balzer, K. Kultus and S. Kohler, 2009. A new generation of fertility monitoring in cattle herds, In: C. Lokhorst, J. F. M. Huijsmans and R. P. M. de Louw, The Fourth European Conference on Precision Livestock Farming, Wageningen Academic Publishers, Wageningen, the Netherlands, Wageningen, the Netherlands, p. 225-234.

T. M. Banhazi, H. Lehr, J. L. Black, H. Crabtree, P. Schofield, M. Tscharke and D. Berckmans, 2012. Precision livestock farming: An international review of scientificand commercial aspects. Internation Journal of Agricultural and Biological Engineering 5, 1-9.

D. Bar and R. Solomon, 2010. Rumination collars: What can they tell us, In: The First North American Conference on Precision Dairy Management, Toronto, Canada, p. 214-215.

H. W. Barkema, M. A. G. von Keyserlingk, J. P. Kastelic, T. Lam, C. Luby, J. P. Roy, S. J. LeBlanc, G. P. Keefe and D. F. Kelton, 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. Journal of Dairy Science 98, 7426-7445.

A. C. Barrier, M. J. Haskell, S. Birch, A. Bagnall, D. J. Bell, J. Dickinson, A. I. Macrae and C. M. Dwyer, 2013. The impact of dystocia on dairy calf health, welfare, performance and survival. Veterinary Journal 195, 86-90.

J. A. Bartolome and L. F. Archbald, 2011. Reproductive management in dairy cows, In: Dairy production medicine, Blackwell Publishing Ltd., p. 73-79.

D. Bates, M. Maechler, B. Bolker and S. Walker, 2013. Linear mixed-effects models using eigen and s4, lme4 version 1.0-5

N. M. Bello, J. S. Stevenson and R. J. Tempelman, 2012. Invited review: Milk production and reproductive performance: Modern interdisciplinary insights into an enduring axiom. Journal of Dairy Science 95, 5461-5475.

D. Berckmans, 2008. Precision livestock farming (plf). Computers and Electronics in Agriculture 62, 1.

T. J. Bergstra, 2014. Pig husbandry in a changing social and economic environment : Societal attitudes, farm economics and animal welfare, Wageningen University, Wageningen.

R. J. Berry, A. D. Kennedy, S. L. Scott, B. L. Kyle and A. L. Schaefer, 2003. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. Canadian Journal of Animal Science 83, 687-693.

J. M. Bewley, M. D. Boehlje, A. W. Gray, H. Hogeveen, S. J. Kenyon, S. D. Eicher and M. M. Schutz, 2010a. Assessing the potential value for an automated dairy cattle body condition scoring system through stochastic simulation. Agricultural Finance Review 70, 24.

J. M. Bewley, M. D. Boehlje, A. W. Gray, H. Hogeveen, S. J. Kenyon, S. D. Eicher and M. M. Schutz, 2010b. Stochastic simulation using @risk for dairy business investment decisions. Agricultural Finance Review 70, 28.

J. M. Bewley and M. M. Schutz, 2010. Recent studies using a reticular bolus system for monitoring dairy cattle core body temperature, In: The First North American Conference on Precision Dairy Management, Toronto, Canada, p. 218-219.

R. C. Bicalho, 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.

R. Bijl, S. R. Kooistra and H. Hogeveen, 2007. The profitability of automatic milking on dutch dairy farms. Journal of Dairy Science 90, 239-248.

J. R. Bikker, H. van Lear, R. Rump, J. Doorenbos, K. van Meurs, G. M. Griffioen and J. Dijkstra, 2014. Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. Journal of Dairy Science 97, 2974-2979.

G. Bobe, J. W. Young and D. C. Beitz, 2004. Invited review: Pathology, etiology, prevention, and treatment of fatty liver in dairy cows*. Journal of Dairy Science 87, 3105-3124.

M. R. Borchers and J. M. Bewley, 2015. An assessment of producer precision dairy farming technology use, prepurchase considerations, and usefulness. Journal of Dairy Science 98, 4198-4205.

M. R. Borchers, Y.-M. chang, A. E. Stone, B. A. Wadsworth and J. M. Bewley, 2015. Predicting impending calving using automatic activity and rumination measures in dairy cattle, In: M. Guarino and D. Berckmans, 7th European Conference on Precision Livestock Farming, Milan, Italy, p. 406 - 414.

J. M. Bos and B. Gremmen, 2013. Does precision livestock farming turn animals into objects, In: D. Berckmans and J. Vandermeulen, 6th European Conference on Precision Livestock Farming, Leuven, Belgium, p.. 106-113.

J. M. Bos and G. Munnichs, 2016. Digitalisering van dieren - verkenning precision livestock farming, Rathenau Instituut, Den Haag.

M. Brandt, A. Haeussermann and E. Hartung, 2010. Invited review: Technical solutions for analysis of milk constituents and abnormal milk. Journal of Dairy Science 93, 427-436.

U. Brehme, U. Stollberg, R. Holz and T. Schleusener, 2008. Alt pedometer - new sensor-aided measurement system for improvement in oestrus detection. Computers and Electronics in Agriculture 62, 73-80.

J. S. Brickell and D. C. Wathes, 2011. A descriptive study of the survival of holstein-friesian heifers through to third calving on english dairy farms. Journal of Dairy Science 94, 1831-1838.

M. R. N. Bruijnis, H. Hogeveen and E. N. Stassen, 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. Journal of Dairy Science 93, 2419-2432.

L. D. Brunassi, D. J. de Moura, I. D. Naas, M. M. do Vale, S. R. L. de Souza, K. A. O. de Lima, T. M. R. de Carvalho and L. G. D. Bueno, 2010. Improving detection of dairy cow estrus using fuzzy logic. Scientia Agricola 67, 503-509.

S. Bucher and A. Sundrum, 2014. Short communication: Decrease in rumination time as an indicator of the onset of calving. Journal of Dairy Science 97, 3120-3127.

S. Buczinski, G. Boulay and D. Francoz, 2015. Preoperative and postoperative l-lactatemia assessment for the prognosis of right abomasal disorders in dairy cattle. Journal of Veterinary Internal Medicine 29, 375-380.

O. Burfeind, V. S. Suthar, R. Voigtsberger, S. Bonk and W. Heuwieser, 2011. Validity of prepartum changes in vaginal and rectal temperature to predict calving in dairy cows. Journal of Dairy Science 94, 5053-5061.

V. Calcagno, 2013. Glmulti: Model selection and multimodel inference made easy, version 1.0.7. Automated model selection and model-averaging. Provides a wrapper for glm and other functions, automatically generating.

O. Cangar, T. Leroy, M. Guarino, E. Vranken, R. Fallon, J. Lenehan, J. Mee and D. Berckmans, 2008. Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis. Computers and Electronics in Agriculture 64, 53-60.

S. Canu, M. Boland, G. M. Lloyd, M. Newman, M. F. Christie, P. J. May, R. M. Christley, R. F. Smith and H. Dobson, 2010. Predisposition to repeat breeding in uk cattle and success of artificial insemination alone or in combination with embryo transfer. Veterinary Record 167, 44-51.

D. Z. Caraviello, K. A. Weigel, M. Craven, D. Gianola, N. B. Cook, K. V. Nordlund, P. M. Fricke and M. C. Wiltbank, 2006a. Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. Journal of Dairy Science 89, 4703-4722.

D. Z. Caraviello, K. A. Weigel, P. M. Fricke, M. C. Wiltbank, M. J. Florent, N. B. Cook, K. V. Nordlund, N. R. Zwald and C. L. Rawson, 2006b. Survey of management practices on reproductive performance of dairy cattle on large us commercial farms. Journal of Dairy Science 89, 4723-4735.

I. A. Casalinuovo, D. Di Pierro, M. Coletta and P. Di Francesco, 2006. Application of electronic noses for disease diagnosis and food spoilage detection. Sensors 6, 1428-1439.

J. Cavalieri, V. E. Eagles, M. Ryan and K. L. MacMillan, 2003a. Comparison of four methods for detection of oestrus in dairy cows with resynchronised oestrous cycles. Australian Veterinary Journal 81, 422-425.

J. Cavalieri, V. E. Eagles, M. Ryan and K. L. MacMillan, 2003b. Role of the sensitivity of detection of oestrus in the submission rate of cows treated to resynchronise oestrus. Australian Veterinary Journal 81, 416-421.

J. Cavalieri, L. R. Flinker, G. A. Anderson and K. L. Macmillan, 2003c. Characteristics of oestrus measured using visual observation and radiotelemetry. Animal Reproduction Science 76, 1-12.

D. Cavero, K. H. Tolle, C. Buxade and J. Krieter, 2006. Mastitis detection in dairy cows by application of fuzzy logic. Livestock Science 105, 207-213.

D. Cavero, K. H. Tolle, C. Henze, C. Buxade and J. Krieter, 2008. Mastitis detection in dairy cows by application of neural networks. Livestock Science 114, 280-286.

D. Cavero, K. H. Tolle, G. Rave, C. Buxade and J. Krieter, 2007. Analysing serial data for mastitis detection by means of local regression. Livestock Science 110, 101-110.

CBS (Statistics Netherlands), 2016. Farms and animals: Holdings, animals and farm size (dairy farming), Accessed 04-10-2016 ed. Wageningen Economic Research, Agrimati.

M. G. Chagunda, T. Larsen, M. Bjerring and K. L. Ingvartsen, 2006a. L-lactate dehydrogenase and n-acetyl-beta-d-glucosaminidase activities in bovine milk as indicators of non-specific mastitis. Journal of Dairy Research 73, 431-440.

M. G. G. Chagunda, N. C. Friggens, M. D. Rasmussen and T. Larsen, 2006b. A model for detection of individual cow mastitis based on an indicator measured in milk. Journal of Dairy Science 89, 2980-2998.

N. Chapinal, A. M. de Passille, M. Pastell, L. Hanninen, 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.

N. Chapinal, A. M. de Passille and J. Rushen, 2009. Weight distribution and gait in dairy cattle are affected by milking and late pregnancy. Journal of Dairy Science 92, 581-588.

N. Chapinal, A. M. de Passille, J. Rushen and S. Wagner, 2010. Automated methods for detecting lameness and measuring analgesia in dairy cattle. Journal of Dairy Science 93, 2007-2013.

N. Charfeddine and M. A. Pérez-Cabal, 2017. Effect of claw disorders on milk production, fertility, and longevity, and their economic impact in spanish holstein cows. Journal of Dairy Science 100, 653-665.

C. E. F. Clark, N. A. Lyons, L. Millapan, S. Talukder, G. M. Cronin, K. L. Kerrisk and S. C. Garcia, 2015. Rumination and activity levels as predictors of calving for dairy cows. Animal 9, 691-695.

R. W. Claycomb, P. T. Johnstone, G. A. Mein and R. A. Sherlock, 2009. An automated in-line clinical mastitis detection system using measurement of conductivity from foremilk of individual udder quarters. New Zealand Veterinary Journal 57, 208-214.

A. Colak, B. Polat, Z. Okumus, M. Kaya, L. E. Yanmaz and A. Hayirli, 2008. Early detection of mastitis using infrared thermography in dairy cows. Journal of Dairy Science 91, 4244-4248.

M. Conneely, D. R. Berry, J. P. Murphy, I. Lorenz, M. L. Doherty and E. Kennedy, 2014. Effect of feeding colostrum at different volumes and subsequent number of transition milk feeds on the serum immunoglobulin g concentration and health status of dairy calves. Journal of Dairy Science 97, 6991-7000.

W. S. Croushore, P. A. Ospina, D. C. Welch, D. J. Zawisza and D. V. Nydam, 2013. Association between beta-hydroxybutyrate concentration at surgery for correction of left-displaced abomasum in dairy cows and removal from the herd after surgery. Javma-Journal of the American Veterinary Medical Association 243, 1329-1333.

CRV, 2011. Annual report 2011, CRV (cattle breeding company), Arnhem, the Netherlands.

CRV, 2012. Crv pricelist 2012-2013, CRV (cattle breeding company), Arnhem, the Netherlands.

M. Darr and W. Epperson, 2009. Embedded sensor technology for real time determination of animal lying time. Computers and Electronics in Agriculture 66, 106-111.

I. de Blas, 2010. Winepi, Faculty of veterinary medicine, University of Zaragoza. http://www.winepi.net

C. J. A. M. de Koning, 2010. Automatic milking - common practice on dairy farms. In: The first North American Conference on Precision Dairy Management, Toronto, Canada52-67.

R. M. de Mol, E. J. B. Bleumer, P. H. Hogewerf and A. H. Ipema, 2009a. Recording of dairy cow behaviour with wireless accelerometers, In: C. Lokhorst, J. F. M. Huijsmans and R. P. M. de Louw, The Fourth European Conference on Precision Livestock Farming, Wageningen Academic Publishers, Wageningen, the Netherlands, Wageningen, the Netherlands, p. 349-356.

R. M. de Mol, R. J. H. Lammers, J. C. A. M. Pompe, A. H. Ipema and P. H. Hogewerf, 2009b. Recording and analysis of locomotion in dairy cows with 3d accelerometers, In: C. Lokhorst, J. F. M. Huijsmans and R. P. M. de Louw, The Fourth European Conference on Precision Livestock Farming, Wageningen Academic Publishers, Wageningen, the Netherlands, Wageningen, the Netherlands, p.

R. M. de Mol, P. H. F. M. Verhoeven, P. H. Hogewerf and A. H. Ipema, 2011. Automated behaviour monitoring in dairy cows, In: C. Lokhorst and D. Berckmans, Fifth European Conference on Precision Livestock Farming, Czech Centre for Science and Society, Prague, Czech Republic, Prague, Czech Republic, p. 70-80.

A. P. W. de Roos, H. van den Bijgaart, J. Horlyk and G. de Jong, 2007. Screening for subclinical ketosis in dairy cattle by fourier transform infrared spectrometry. Journal of Dairy Science 90, 1761-1766.

A. de Vries, 2004. Economics of delayed replacement when cow performance is seasonal*. Journal of Dairy Science 87, 2947-2958.

A. De Vries, 2006. Economic value of pregnancy in dairy cattle. Journal of Dairy Science 89, 3876-3885.

C. D. Dechow and R. C. Goodling, 2008. Mortality, culling by sixty days in milk, and production profiles in high- and low-survival pennsylvania herds. Journal of Dairy Science 91, 4630-4639.

J. M. DeJarnette, C. G. Sattler, C. E. Marshall and R. L. Nebel, 2007. Voluntary waiting period management practices in dairy herds participating in a progeny test program. Journal of Dairy Science 90, 1073-1079.

N. S. del Rio, S. Stewart, P. Rapnicki, Y. M. Chang and P. M. Fricke, 2007. An observational analysis of twin births, calf sex ratio, and calf mortality in holstein dairy cattle. Journal of Dairy Science 90, 1255-1264.

C. M. B. Dematawewa, R. E. Pearson and P. M. VanRaden, 2007. Modeling extended lactations of holsteins. Journal of Dairy Science 90, 3924-3936.

R. M. Demeter, A. R. Kristensen, J. Dijkstra, A. Lansink, M. P. M. Meuwissen and J. A. M. van Arendonk, 2011. A multi-level hierarchic markov process with bayesian updating for herd optimization and simulation in dairy cattle. Journal of Dairy Science 94, 5938-5962.

C. Derman, 1973. A guide to probability theory and application, L. J. Gleser and I. Olkin New York, Holt, Rinehart and Winston.

K. C. Dhuyvetter, T. L. Kastens, M. W. Overton and J. F. Smith, 2007. Cow culling decisions: Costs or economic opportunity?, Department of Agricultural Economics, Kansas State University.

J. Dijkstra, J. France, M. S. Dhanoa, J. A. Maas, M. D. Hanigan, A. J. Rook and D. E. Beever, 1997. A model to describe growth patterns of the mammary gland during pregnancy and lactation. Journal of Dairy Science 80, 2340-2354.

K. A. Dolecheck and J. M. Bewley, 2016. Cost of days open, University of Kentucky, http://afs.ca.uky.edu/dairy/decision-tools.

K. A. Dolecheck, G. Heersche Jr and J. M. Bewley, 2016a. Retention payoff–based cost per day open regression equations: Application in a user-friendly decision support tool for investment analysis of automated estrus detection technologies. Journal of Dairy Science 99, 10182-10193.

K. A. Dolecheck, W. J. Silvia, G. Heersche, C. L. Wood, K. J. McQuerry and J. M. Bewley, 2016b. A comparison of timed artificial insemination and automated activity monitoring with hormone intervention in 3 commercial dairy herds. Journal of Dairy Science 99, 1506-1514.

M. Dowle, T. Short and S. Lianoglou, 2013. Extension of data.Frame for fast indexing, fast ordered joins, fast assignment, fast grouping and list columns,

P. M. Down, M. J. Green and C. D. Hudson, 2013. Rate of transmission: A major determinant of the cost of clinical mastitis. Journal of Dairy Science 96, 6301-6314.

J. Dubuc, T. F. Duffield, K. E. Leslie, J. S. Walton and S. J. LeBlanc, 2010. Risk factors for postpartum uterine diseases in dairy cows. Journal of Dairy Science 93, 5764-5771.

J. H. Dufty, 1981. The influence of various degrees of confinement and supervision on the incidence of dystokia and stillbirths in hereford heifers. New Zealand Veterinary Journal 29, 44-8.

J. L. Edwards and P. R. Tozer, 2004. Using activity and milk yield as predictors of fresh cow disorders. Journal of Dairy Science 87, 524-531.

T. V. Elzhov, K. M. Mullen, A. N. Spiess and B. Bolker, 2013. R interface to the levenberg-marquardt nonlinear least-squares algorithm found in minpack, plus support for bounds, In: (Ed.)^(Eds.), 1.1-8 ed., p.^pp. The nls.lm function provides an R interface to lmder and lmdif from the MINPACK library, for solving nonlinear least-squares problems by a modification of the Levenberg-Marquardt.

E. Espada and H. Vijverberg, 2002. Milk colour analysis as a tool for the detection of abnormal milk, In: J. McLean, M. Sinclair and B. West (Ed.)^(Eds.), First North American Conference on robotic milking, ed., Toronto, p.^pp. IV-28 - IV-38.

J. F. Ettema and S. Østergaard, 2006. Economic decision making on prevention and control of clinical lameness in danish dairy herds. Livestock Science 102, 92-106.

J. F. Ettema, S. Ostergaard and M. K. Sorensen, 2011. Effect of including genetic progress in milk yield on evaluating the use of sexed semen and other reproduction strategies in a dairy herd. Animal 5, 1887-1897.

R. Firk, E. Stamer, W. Junge and J. Krieter, 2002. Automation of oestrus detection in dairy cows: A review. Livestock Production Science 75, 219-232.

R. Firk, E. Stamer, W. Junge and J. Krieter, 2003a. Improving oestrus detection by combination of activity measurements with information about previous oestrus cases. Livestock Production Science 82, 97-103.

R. Firk, E. Stamer, W. Junge and J. Krieter, 2003b. Oestrus detection in dairy cows based on serial measurements using univariate and multivariate analysis. Archiv Fur Tierzucht-Archives of Animal Breeding 46, 127-142.

M. Floridi, F. Bartolini, J. H. M. Peerlings, N. B. P. Polman and D. Viaggi, 2013. Modelling the adoption of automatic milking systems in noord-holland. Bio-based and Applied Economics 2, 73-90.

U. o. M. Food and Agricultural Policy Research Institue (FAPRE), 2015. U.S. Baseline briefing book: Projections for agricultural and biofuel markets. ed. FAPRI, University of Missouri, Columbia, MO, USA.

R. H. Foote, 1975. Estrus detection and estrus detection aids. Journal of Dairy Science 58, 248-256.

C. Fourichon, H. Seegers and X. Malher, 2000. Effect of disease on reproduction in the dairy cow: A meta-analysis. Theriogenology 53, 1729-1759.

N. C. Friggens, M. Bjerring, C. Ridder, S. Højsgaard and T. Larsen, 2008. Improved detection of reproductive status in dairy cows using milk progesterone measurements. Reproduction in Domestic Animals 43, 113-121.

N. C. Friggens and M. G. G. Chagunda, 2005. Prediction of the reproductive status of cattle on the basis of milk progesterone measures: Model description. Theriogenology 64, 155-190.

N. C. Friggens, M. G. G. Chagunda, M. Bjerring, C. Ridder, S. Højsgaard and T. Larsen, 2007. Estimating degree of mastitis from time-series measurements in milk: A test of a model based on lactate dehydrogenase measurements. Journal of Dairy Science 90, 5415-5427.

G. Gabor, F. Toth, L. Ozsvari, Z. Abonyi-Toth and R. G. Sasser, 2008. Factors influencing pregnancy rate and late embryonic loss in dairy cattle. Reproduction in Domestic Animals 43, 53-58.

N. Galon, 2010. The use of pedometry for estrus detection in dairy cows in israel. Journal of Reproduction and Development 56, S48-S52.

K. N. Galvão, P. Federico, A. De Vries and G. M. Schuenemann, 2013. Economic comparison of reproductive programs for dairy herds using estrus detection, timed artificial insemination, or a combination. Journal of Dairy Science 96, 2681-2693.

J. Gasteiner, M. Fallast, S. Rosenkranz, J. Hausler, K. Schmeider and T. Gugenberge, 2009. Measuring rumen ph and temperature by an indwelling and wireless data transmitting unit and application under different feeding conditions, In: C. Lokhorst, J. F. M. Huijsmans and R. P. M. de Louw, The Fourth European Conference on Precision Livestock Farming, Wageningen Academic Publishers, Wageningen, the Netherlands, Wageningen, the Netherlands, p. 127-133.

U. Geary, N. Lopez-Villalobos, N. Begley, F. McCoy, B. O'Brien, L. O'Grady and L. Shalloo, 2012. Estimating the effect of mastitis on the profitability of irish dairy farms. Journal of Dairy Science 95, 3662-3673.

E. H. Gillis, J. P. Gosling, J. M. Sreenan and M. Kane, 2002. Development and validation of a biosensor-based immunoassay for progesterone in bovine milk. Journal of Immunological Methods 267, 131-138.

J. O. Giordano, P. M. Fricke, M. C. Wiltbank and V. E. Cabrera, 2011. An economic decision-making support system for selection of reproductive management programs on dairy farms. Journal of Dairy Science 94, 6216-6232.

J. O. Giordano, A. S. Kalantari, P. M. Fricke, M. C. Wiltbank and V. E. Cabrera, 2012. A daily herd markov-chain model to study the reproductive and economic impact of reproductive programs combining timed artificial insemination and estrus detection. Journal of Dairy Science 95, 5442-5460.

D. Goense, H. W. J. Houwers, H. C. Muller, D. Unsenos and W. Wehren, 2009. A wireless network for measuring rumen ph in dairy cows, In: C. Lokhorst, J. F. M. Huijsmans and R. P. M. de Louw, The Fourth European Conference on Precision Livestock Farming, Wageningen Academic Publishers, Wageningen, the Netherlands, Wageningen, the Netherlands, p. 117-125.

E. Gonzalezjimenez and K. L. Blaxter, 1962. Metabolism and thermal regulation of calves in first month of life. British Journal of Nutrition 16, 199-&.

P. Gordon, 2011. Oestrus detection in dairy cattle. In Practice 33, 542-546.

G. R. Greatrix, J. C. Quayle and R. A. Coombe, 1968. Detection of abnormal milk by electrical means. Journal of Dairy Research 35, 213-&.

D. E. Grobbee and A. W. Hoes, 2015. Clinical epidemiology, Principles, methods and applications for clinical research, Second edition ed. Jones and Barlett Learning, Burlington, MA.

H. Groenendaal, D. T. Galligan and H. A. Mulder, 2004. An economic spreadsheet model to determine optimal breeding and replacement decisions for dairy cattle. Journal of Dairy Science 87, 2146-2157.

Y. T. Gröhn, P. J. Rajala-Schultz, H. G. Allore, M. A. DeLorenzo, J. A. Hertl and D. T. Galligan, 2003. Optimizing replacement of dairy cows: Modeling the effects of diseases. Preventive Veterinary Medicine 61, 27-43.

I. Halachmi, P. Polak, D. J. Roberts, M. Klopcic and J. M. Bewley, 2009. Thermally sensed, automatic cow body condition scoring, In: C. Lokhorst, J. F. M. Huijsmans and R. P. M. de Louw, The Fourth European Conference on Precision Livestock Farming, ed. Wageningen Academic Publishers, Wageningen, the Netherlands, Wageningen, the Netherlands, p. 193-200.

T. Halasa, K. Huijps, O. Osteras and H. Hogeveen, 2007. Economic effects of bovine mastitis and mastitis management: A review. Veterinary Quarterly 29, 18-31.

B. G. Hansen, 2015. Robotic milking-farmer experiences and adoption rate in jaeren, norway. Journal of Rural Studies 41, 109-117.

K. J. Hassan, S. Samarasinghe and M. G. Lopez-Benavides, 2009. Use of neural networks to detect minor and major pathogens that cause bovine mastitis. Journal of Dairy Science 92, 1493-1499.

T. Hastie, R. Tibshirani and J. Friedman, 2009. The elements of statistical learning. Data mining, inference and prediction. 2 ed. Springer-Verlag, New York.

A. M. Heikkila, J. I. Nousiainen and S. Pyorala, 2012. Costs of clinical mastitis with special reference to premature culling. Journal of Dairy Science 95, 139-150.

J. A. Hernandez, C. A. Risco, F. S. Lima and J. E. P. Santos, 2012. Observed and expected combined effects of clinical mastitis and low body condition on pregnancy loss in dairy cows. Theriogenology 77, 115-121.

T. Hertz, 2016. Usda economic research service - background.

J. H. Higginson, S. T. Millman, K. E. Leslie and D. F. Kelton, 2010. Validation of a new pedometry system for use in behavioural research and lameness detection in dairy cattle, In: The First North American Conference on Precision Dairy Management, Toronto, Canada, p. 32-133.

C. D. Hockey, J. M. Morton, S. T. Norman and M. R. McGowan, 2010a. Evaluation of a neck mounted 2-hourly activity meter system for detecting cows about to ovulate in two paddock-based australian dairy herds. Reproduction in Domestic Animals 45, E107-E117.

C. D. Hockey, J. M. Morton, S. T. Norman and M. R. McGowan, 2010b. Improved prediction of ovulation time may increase pregnancy rates to artificial insemination in lactating dairy cattle. Reproduction in Domestic Animals 45, e239-e248.

H. Hogeveen, C. Kamphuis, W. Steeneveld and H. Mollenhorst, 2010. Sensors and clinical mastitis-the quest for the perfect alert. Sensors 10, 7991-8009.

H. Hogeveen, M. A. Varner, D. S. Bree, D. E. Dill, E. N. Noordhuizenstassen and A. Brand, 1994. Knowledge representation methods for dairy decision-support systems. Journal of Dairy Science 77, 3704-3715.

S. Højsgaard and N. C. Friggens, 2010. Quantifying degree of mastitis from common trends in a panel of indicators for mastitis in dairy cows. Journal of Dairy Science 93, 582-592.

A. Holman, J. Thompson, J. E. Routly, J. Cameron, D. N. Jones, D. Grove-White, R. F. Smith and H. Dobson, 2011. Comparison of oestrus detection methods in dairy cattle. Veterinary Record 169, 47-+.

E. H. P. Houben, R. B. M. Huirne, A. A. Dijkhuizen and A. R. Kristensen, 1994. Optimal replacement of mastitic cows determined by a hierarchic markov process. Journal of Dairy Science 77, 2975-2993.

M. Hovinen, A. M. Aisla and S. Pyörälä, 2006. Accuracy and reliability of mastitis detection with electrical conductivity and milk colour measurement in automatic milking. Acta Agriculturae Scandinavica Section a-Animal Science 56, 121-127.

M. Hovinen and S. Pyörälä, 2011. Invited review: Udder health of dairy cows in automatic milking. Journal of Dairy Science 94, 547-562.

M. Hovinen, J. Siivonen, S. Taponen, L. Hanninen, M. Pastell, A. M. Aisla and S. Pyörälä, 2008. Detection of clinical mastitis with the help of a thermal camera. Journal of Dairy Science 91, 4592-4598.

N. Hugen, R. H. Verhoeven, V. E. Lemmens, C. J. van Aart, M. A. Elferink, S. A. Radema, I. D. Nagtegaal and J. H. de Wilt, 2015. Colorectal signet-ring cell carcinoma: Benefit from adjuvant chemotherapy but a poor prognostic factor. International Journal of Cancer 136, 333-339.

K. Huijps, T. Lam and H. Hogeveen, 2008. Costs of mastitis: Facts and perception. Journal of Dairy Research 75, 113-120.

I. A. Hutchinson, L. Shalloo and S. T. Butler, 2013. Expanding the dairy herd in pasture-based systems: The role for sexed semen use on virgin heifers. Journal of Dairy Science 96, 1312-1322.

J. Hyde and P. Engel, 2002. Investing in a robotic milking system: A monte carlo simulation analysis. Journal of Dairy Science 85, 2207-2214.

C. Inchaisri, A. De Vries, R. Jorritsma and H. Hogeveen, 2012. Improved knowledge about conception rates influences the decision to stop insemination in dairy cows. Reproduction in Domestic Animals 47, 820-826.

C. Inchaisri, H. Hogeveen, P. Vos, G. C. van der Weijden and R. Jorritsma, 2010a. Effect of milk yield characteristics, breed, and parity on success of the first insemination in dutch dairy cows. Journal of Dairy Science 93, 5179-5187.

C. Inchaisri, R. Jorritsma, J. C. M. Vernooij, P. Vos, G. C. van der Weijden and H. Hogeveen, 2011a. Cow effects and estimation of success of first and following inseminations in dutch dairy cows. Reproduction in Domestic Animals 46, 1043-1049.

C. Inchaisri, R. Jorritsma, P. Vos, G. C. van der Weijden and H. Hogeveen, 2010b. Economic consequences of reproductive performance in dairy cattle. Theriogenology 74, 835-846.

C. Inchaisri, R. Jorritsma, P. Vos, G. C. van der Weijden and H. Hogeveen, 2011b. Analysis of the economically optimal voluntary waiting period for first insemination. Journal of Dairy Science 94, 3811-3823.

A. H. Ipema, E. J. B. Bleumer and C. Lokhorst, 2011. Precision livestock farming claims for sustainable dairy production, In: C. Lokhorst and D. Berckmans, 5th European Conference on Precision Livestock Farming, Czech Centre for Science and Society, Prague, Czech Republic, p. 271-280.

A. H. Ipema, D. Goense, P. H. Hogewerf, H. W. J. Houwers and H. van Roest, 2008. Pilot study to monitor body temperature of dairy cows with a rumen bolus. Computers and Electronics in Agriculture 64, 49-52.

A. H. Ipema and W. Rossing, 1987. Automatic individual feeding of dairy cattle, In: Proceedings of the third symposium automation in daiyin, Instituut voor Mechanisatie, Arbeid en Gebouwen, I.M.A.G., Wageningen, the Netherlands, Wageningen, the Netherlands, p. 41-51.

K. Ito, M. A. G. von Keyserlingk, S. J. LeBlanc and D. M. Weary, 2010. Lying behavior as an indicator of lameness in dairy cows. Journal of Dairy Science 93, 3553-3560.

J. A. Jacobs and J. M. Siegford, 2012. Invited review: The impact of automatic milking systems on dairy cow management, behavior, health, and welfare. Journal of Dairy Science 95, 2227-2247.

A. W. Jalvingh, 1992. The possible role of existing models in on-farm decision support in dairy-cattle and swine production. Livestock Production Science 31, 351-365.

R. Jonsson, M. Blanke, N. K. Poulsen, F. Caponetti and S. Højsgaard, 2011. Oestrus detection in dairy cows from activity and lying data using on-line individual models. Computers and Electronics in Agriculture 76, 6-15.

S. T. Juarez, P. H. Robinson, E. J. DePeters and E. O. Price, 2003. Impact of lameness on behavior and productivity of lactating holstein cows. Applied Animal Behaviour Science 83, 1-14.

C. Kamphuis, B. DelaRue, C. R. Burke and J. Jago, 2012. Field evaluation of 2 collar-mounted activity meters for detecting cows in estrus on a large pasture-grazed dairy farm. Journal of Dairy Science 95, 3045-3056.

C. Kamphuis, H. Mollenhorst, A. Feelders, D. Pietersma and H. Hogeveen, 2010a. Decision-tree induction to detect clinical mastitis with automatic milking. Computers and Electronics in Agriculture 70, 60-68.

C. Kamphuis, H. Mollenhorst, J. A. P. Heesterbeek and H. Hogeveen, 2010b. Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. Journal of Dairy Science 93, 3616-3627.

C. Kamphuis, D. Pietersma, R. Van der Tol, M. Wiedemann and H. Hogeveen, 2008a. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. Computers and Electronics in Agriculture 62, 169-181.

C. Kamphuis, R. Sherlock, J. Jago, G. Mein and H. Hogeveen, 2008b. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. Journal of Dairy Science 91, 4560-4570.

N. D. Kappel, F. Proll and G. Gauglitz, 2007. Development of a tirf-based biosensor for sensitive detection of progesterone in bovine milk. Biosensors & Bioelectronics 22, 2295-2300.

J. O. S. Kennedy and A. W. Stott, 1993. An adaptive decision-making aid for dairy cow replacement. Agricultural Systems 42, 25-39.

D. Klein-Jobstl, M. Iwersen and M. Drillich, 2014. Farm characteristics and calf management practices on dairy farms with and without diarrhea: A case-control study to investigate risk factors for calf diarrhea. Journal of Dairy Science 97, 5110-5119.

M. Kujala, M. Pastell and T. Soveri, 2008. Use of force sensors to detect and analyse lameness in dairy cows. Veterinary Record 162, 365-368.

K. Kultus, H. U. Balzer and S. Kohler, 2011. Comparison of results using smardwatch to detect oestrus in dairy cattle parallel to progesterone test and visual detection, In: C. Lokhorst and D. Berckmans, The Fifth European Conference on Precision Livestock Farming, Czech Centre for Science and Society, Prague, Czech Republic, Prague, Czech Republic, p. 154-162.

E. A. Lane, M. A. Crowe, M. E. Beltman and S. J. More, 2013. The influence of cow and management factors on reproductive performance of irish seasonal calving dairy cows. Animal Reproduction Science 141, 34-41.

D. N. Ledgerwood, C. Winckler and C. B. Tucker, 2010. Evaluation of data loggers, sampling intervals, and editing techniques for measuring the lying behavior of dairy cattle. Journal of Dairy Science 93, 5129-5139.

H. Lehr, 2011. Practical and acceptable precision livestock farming: Results from bright animal, In: C. Lokhorst and D. Berckmans, 5th European Conference on Precision Livestock Farming, Czech Centre for Science and Society, Prague, Czech Republic, p. 281-295.

LEI, 2013. Lei's farm accountancy data network, Landbouw Economisch Instituut (Agricultural Economic Istitute), Den Haag, the Netherlands, p.^pp.

D. Liang, 2013. Estimating the economic losses from diseases and extended days open with a farm-level stochastic model, Animal and Food Sciences, University of Kentucky.

P. Liberati and P. Zappavigna, 2009. Improving the automated monitoring of dairy cows by integrating various data acquisition systems. Computers and Electronics in Agriculture 68, 62-67.

J. L. Linzell and M. Peaker, 1975. Efficacy of measurement of electrical-conductivity of milk for detection of subclinical mastitis in cows - detection of infected cows at a single visit. British Veterinary Journal 131, 447-461.

T. W. A. Little, C. N. Hebert and D. Forbes, 1968. Electrical conductivity and leucocyte count of bivine milk. Veterinary Record 82, 431-&.

E. Lof, H. Gustafsson and U. Emanuelson, 2014. Factors influencing the chance of cows being pregnant 30 days after the herd voluntary waiting period. Journal of Dairy Science 97, 2071-2080.

J. E. Lombard, F. B. Garry, S. M. Tomlinson and L. P. Garber, 2007. Impacts of dystocia on health and survival of dairy calves. Journal of Dairy Science 90, 1751-1760.

P. Løvendahl and M. G. G. Chagunda, 2010. On the use of physical activity monitoring for estrus detection in dairy cows. Journal of Dairy Science 93, 249-259.

J. M. Lukas, J. K. Reneau, R. Wallace, D. Hawkins and C. Munoz-Zanzi, 2009. A novel method of analyzing daily milk production and electrical conductivity to predict disease onset. Journal of Dairy Science 92, 5964-5976.

F. S. Machado, T. R. Tomich, A. L. Ferreira, L. F. L. Cavalcanti, M. M. Campos, C. A. V. Paiva, M. N. Ribas and L. G. R. Pereira, 2016. Technical note: A facility for respiration measurements in cattle. Journal of Dairy Science 99, 4899-4906.

W. Maertens, J. Vangeyte, J. Baert, A. Jantuan, K. C. Mertens, S. De Campeneere, A. Pluk, G. Opsomer, S. Van Weyenberg and A. Van Nuffel, 2011. Development of a real time cow gait

tracking and analysing tool to assess lameness using a pressure sensitive walkway: The gaitwise system. Biosystems Engineering 110, 29-39.

P. Martiskainen, M. Jarvinen, J. P. Skon, J. Tiirikainen, M. Kolehmainen and J. Mononen, 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Applied Animal Behaviour Science 119, 32-38.

E. Mathijs, 2004. Socio-economic aspects of automatic miling, In: A. Meijering, H. Hogeveen and C. J. A. M. de Koning, Automatic milking: A better understanding, Wageningen Academic Publishers, Wageningen, the Netherlands, p. 46-55.

F. Mazeris, 2010. Delaval herd navigator[r] proactive herd management, In: First North American Conference on Precision Dairy Management, Toronto, Canada, p. 26-27.

H. Mazrier, S. Tal, E. Aizinbud and U. Bargai, 2006. A field investigation of the use of the pedometer for the early detection of lameness in cattle. Canadian Veterinary Journal-Revue Veterinaire Canadienne 47, 883-886.

J. F. Mee, 2008. Prevalence and risk factors for dystocia in dairy cattle: A review. Veterinary journal (London, England : 1997) 176, 93-101.

J. F. Mee, C. Sanchez-Miguel and M. Doherty, 2014. Influence of modifiable risk factors on the incidence of stillbirth/perinatal mortality in dairy cattle. Veterinary Journal 199, 19-23.

C. L. Meyer, P. J. Berger and K. J. Koehler, 2000. Interactions among factors affecting stillbirths in holstein cattle in the united states. Journal of Dairy Science 83, 2657-2663.

H. M. Miedema, M. S. Cockram, C. M. Dwyer and A. I. Macrae, 2011a. Behavioural predictors of the start of normal and dystocic calving in dairy cows and heifers. Applied Animal Behaviour Science 132, 14-19.

H. M. Miedema, M. S. Cockram, C. M. Dwyer and A. I. Macrae, 2011b. Changes in the behaviour of dairy cows during the 24 h before normal calving compared with behaviour during late pregnancy. Applied Animal Behaviour Science 131, 8-14.

B. Miekley, I. Traulsen and J. Krieter, 2012. Detection of mastitis and lameness in dairy cows using wavelet analysis. Livestock Science 148, 227-236.

K. M. Millar, 2000. Respect for animal autonomy in bioethical analysis: The case of automatic milking systems (ams). Journal of Agricultural and Environmental Ethics 12, 41-50.

N. Miranda, R. Morais, M. Dias, C. Viegas, F. Silva, C. Serodio, J. Almeida, J. Azevedo and M. C. Reis, 2009. Bioimplantable impedance and temperature monitor low power micro-system suitable for estrus detection, In: J. Brugger and D. Briand, Proceedings of the eurosensors xxiii conference, Elsevier Science Bv, Amsterdam, p. 505-508.

N. Mohd Nor, W. Steeneveld and H. Hogeveen, 2014. The average culling rate of dutch dairy herds over the years 2007 to 2010 and its association with herd reproduction, performance and health. Journal of Dairy Research 81, 1-8.

H. Mollenhorst, L. J. Rijkaart and H. Hogeveen, 2012. Mastitis alert preferences of farmers milking with automatic milking systems. Journal of Dairy Science 95, 2523-2530.

H. Mollenhorst, P. P. J. van der Tol and H. Hogeveen, 2010. Somatic cell count assessment at the quarter or cow milking level. Journal of Dairy Science 93, 3358-3364.

Moocall, 2016. Moocall calving sensors.

K. G. M. Moons, D. G. Altman, J. B. Reitsma, J. P. A. Ioannidis, P. Macaskill, E. W. Steyerberg, A. J. Vickers, D. F. Ransohoff and G. S. Collins, 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration. Annals of Internal Medicine 162, W1-W73.

T. Mottram and N. J. Bell, 2010. A novel method of monitoring mobility of dairy cows, In: The First North American Conference on Precision Dairy Management, Toronto, Canada, p. 182-183.

R. Muller and L. Schrader, 2003. A new method to measure behavioural activity levels in dairy cows. Applied Animal Behaviour Science 83, 247-258.

R. C. Neves, K. E. Leslie, J. S. Walton and S. J. LeBlanc, 2012. Reproductive performance with an automated activity monitoring system versus a synchronized breeding program. Journal of Dairy Science 95, 5683-5693.

S. Neveux, D. M. Weary, J. Rushen, M. A. G. von Keyserlingk and A. M. de Passille, 2006. Hoof discomfort changes how dairy cattle distribute their body weight. Journal of Dairy Science 89, 2503-2509.

M. Nielen, H. Deluyker, Y. H. Schukken and A. Brand, 1992. Electrical-conductivity of milk - measurement, modifiers, and meta analysis of mastitis detection performance. Journal of Dairy Science 75, 606-614.

L. R. Nielsen, A. R. Pedersen, M. S. Herskin and L. Munksgaard, 2010. Quantifying walking and standing behaviour of dairy cows using a moving average based on output from an accelerometer. Applied Animal Behaviour Science 127, 12-19.

E. Norberg, 2005. Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review. Livestock Production Science 96, 129-139.

E. Norberg, H. Hogeveen, I. R. Korsgaard, N. C. Friggens, K. Sloth and P. Løvendahl, 2004. Electrical conductivity of milk: Ability to predict mastitis status. Journal of Dairy Science 87, 1099-1107.

H. D. Norman, J. L. Hutchison and R. H. Miller, 2010. Use of sexed semen and its effect on conception rate, calf sex, dystocia, and stillbirth of holsteins in the united states. Journal of Dairy Science 93, 3880-3890.

J. O'Connell, F. A. Togersen, N. C. Friggens, P. Løvendahl and S. Højsgaard, 2011. Combining cattle activity and progesterone measurements using hidden semi-markov models. Journal of Agricultural Biological and Environmental Statistics 16, 1-16.

S. Ohtani, T. Wang, K. Nishimura and M. Irie, 2005. Milk fat analysis by fiber-optic spectroscopy. Asian-Australasian Journal of Animal Sciences 18, 580-583.

N. J. Olynk and C. A. Wolf, 2009. Stochastic economic analysis of dairy cattle artificial insemination reproductive management programs. Journal of Dairy Science 92, 1290-1299.

S. Østergaard, N. C. Friggens and M. G. G. Chagunda, 2005. Technical and economic effects of an inline progesterone indicator in a dairy herd estimated by stochastic simulation. Theriogenology 64, 819-843.

A. G. J. M. Oude Lansink, J. Verstegen and J. J. Van den Hengel, 2001. Investment decision making in dutch greenhouse horticulture. Netherlands Journal of Agricultural Science 49, 357-368.

V. Ouellet, E. Vasseur, W. Heuwieser, O. Burfeind, X. Maldague and E. Charbonneau, 2016. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in holstein dairy cows. Journal of Dairy Science 99, 1539-1548.

M. W. Overton, 2005. Cost comparison of natural service sires and artificial insemination for dairy cattle reproductive management. Theriogenology 64, 589-602.

C. Pahl, E. Hartung, A. Grothmann, K. Mahlkow-Nerge and A. Haeussermann, 2014. Rumination activity of dairy cows in the 24 hours before and after calving. Journal of Dairy Science 97, 6935-6941.

Palisade-Corporation, 2010. @risk, Palisade corporation, Ithaca, USA.

M. A. Palmer, G. Olmos, L. A. Boyle and J. F. Mee, 2010. Estrus detection and estrus characteristics in housed and pastured holstein-friesian cows. Theriogenology 74, 255-264.

C. Palombi, M. Paolucci, G. Stradaioli, M. Corubolo, P. B. Pascolo and M. Monaci, 2013. Evaluation of remote monitoring of parturition in dairy cattle as a new tool for calving management. Bmc Veterinary Research 9, 9.

M. Paolucci, L. Sylla, A. Di Giambattista, C. Palombi, A. Elad, G. Stradaioli, P. Pascolo and M. Monaci, 2010. Improving calving management to further enhance reproductive performance in dairy cattle. Veterinary Research Communications 34, S37-S40.

T. J. Parkinson, G. C. W. England and G. H. Arthur, 2001a. Chapter 1 - endogenous and exogenous control of ovarian cyclicity a2 - noakes, david e, Arthur's veterinary reproduction and obstetrics (eighth edition), W.B. Saunders, Oxford, p. 3-53.

T. J. Parkinson, G. C. W. England and G. H. Arthur, 2001b. Chapter 6 - parturition and the care of parturient animals a2 - noakes, david e, Arthur's veterinary reproduction and obstetrics (eighth edition), W.B. Saunders, Oxford, p. 155-187.

T. J. Parkinson, G. C. W. England and G. H. Arthur, 2001c. Chapter 11 - fetal dystocia: Aetiology and incidence a2 - noakes, david e, Arthur's veterinary reproduction and obstetrics (eighth edition), W.B. Saunders, Oxford, p. 245-263.

M. Pastell, L. Hanninen, A. M. de Passille and J. Rushen, 2010. Measures of weight distribution of dairy cows to detect lameness and the presence of hoof lesions. Journal of Dairy Science 93, 954-960.

M. Pastell, M. Hautala, V. Poikalainen, J. Praks, I. Veermae, M. Kujala and J. Ahokas, 2008a. Automatic observation of cow leg health using load sensors. Computers and Electronics in Agriculture 62, 48-53.

M. Pastell, M. Kujala, A. M. Aisla, M. Hautala, V. Poikalainen, J. Praks, I. Veermae and J. Ahokas, 2008b. Detecting cow's lameness using force sensors. Computers and Electronics in Agriculture 64, 34-38.

M. Pastell and H. Madsen, 2008. Application of cusum charts to detect lameness in a milking robot. Expert Systems with Applications 35, 2032-2040.

M. Pastell, J. Tiusanen, M. Hakojarvi and L. Hanninen, 2009. A wireless accelerometer system with wavelet analysis for assessing lameness in cattle. Biosystems Engineering 104, 545-551.

M. E. Pastell and M. Kujala, 2007. A probabilistic neural network model for lameness detection. Journal of Dairy Science 90, 2283-2292.

O. A. Peralta, R. E. Pearson and R. L. Nebel, 2005. Comparison of three estrus detection systems during summer in a large commercial dairy herd. Animal Reproduction Science 87, 59-72.

Personal Communication CRV, 2013a. Commercially available estrus detection systems, Arnhem, the Netherlands, p.^pp.

Personal Communication CRV, 2013b. Fertility expert of cattle breeding company, Arnhem, the Netherlands.

Personal Communication Royal Agrifirm, 2013c. Average live weigths of dutch dairy cows, Apeldoorn, the Netherlands.

C. Pinzon-Sanchez, V. E. Cabrera and P. L. Ruegg, 2011. Decision tree analysis of treatment strategies for mild and moderate cases of clinical mastitis occurring in early lactation. Journal of Dairy Science 94, 1873-1892.

D. Piwczynski, Z. Nogalski and B. Sitkowska, 2013. Statistical modeling of calving ease and stillbirths in dairy cattle using the classification tree technique. Livestock Science 154, 19-27.

A. Pluk, C. Bahr, T. Leroy, A. Poursaberi, X. Song, E. Vranken, W. Maertens, A. Van Nuffel and D. Berckmans, 2010. Evaluation of step overlap as an automatic measuure in dairy cow locomotion. Transactions of the Asabe 53, 1305-1312.

B. Polat, A. Colak, M. Cengiz, L. E. Yanmaz, H. Oral, A. Bastan, S. Kaya and A. Hayirli, 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. Journal of Dairy Science 93, 3525-3532.

A. Poursaberi, C. Bahr, A. Pluk, A. Van Nuffel and D. Berckmans, 2010. Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques. Computers and Electronics in Agriculture 74, 110-119.

P. J. Rajala-Schultz and Y. T. Grohn, 1999. Culling of dairy cows. Part i. Effects of diseases on culling in finnish ayrshire cows. Preventive Veterinary Medicine 41, 195-208.

P. J. Rajala-Schultz and Y. T. Gröhn, 2001. Comparison of economically optimized culling recommendations and actual culling decisions of finnish ayrshire cows. Preventive Veterinary Medicine 49, 29-39.

P. G. Rajkondawar, M. Liu, R. M. Dyer, N. K. Neerchal, U. Tasch, A. M. Lefcourt, B. Erez and M. A. Varner, 2006. Comparison of models to identify lame cows based on gait and lesion scores, and limb movement variables. Journal of Dairy Science 89, 4267-4275.

P. G. Rajkondawar, U. Tasch, A. M. Lefcourt, B. Erez, R. M. Dyer and M. A. Varner, 2002. A system for identifying lameness in dairy cattle. Applied Engineering in Agriculture 18, 87-96.

G. J. Remmelink, K. Blanken, J. C. v. Middelkoop, W. Ouweltjes and H. Wemmenhove, 2012. Handboek melkveehouderij 2012 (handbook dairy farming). ed. Livestock Reserach, Wageningen UR, Lelystad, the Netherlands.

Rendac, 2013. Pricelist, Son, the Netherlands, p. 2.

J. R. Roche, N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford and D. P. Berry, 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. Journal of Dairy Science 92, 5769-5801.

J. Roelofs, F. López-Gatius, R. H. F. Hunter, F. J. C. M. van Eerdenburg and C. Hanzen, 2010. When is a cow in estrus? Clinical and practical aspects. Theriogenology 74, 327-344.

J. B. Roelofs, F. van Eerdenburg, N. M. Soede and B. Kemp, 2005. Pedometer readings for estrous detection and as predictor for time of ovulation in dairy cattle. Theriogenology 64, 1690-1703.

A. J. Rook, J. France and M. S. Dhanoa, 1993. On the mathematical-description of lactation curves. Journal of Agricultural Science 121, 97-102.

W. Rossing, 1999. Animal identification: Introduction and history. Computers and Electronics in Agriculture 24, 1-4.

W. Rossing, A. H. Ipema and P. F. Veltman, 1985. Perspectieven voor het melken in een voerbox. Instituut voor Mechanisatie, Arbeid en Gebouwen, I.M.A.G., Wageningen, the Netherlands.

R. A. Russell and J. M. Bewley, 2013. Characterization of kentucky dairy producer decision-making behavior. Journal of Dairy Science 96, 4751-4758.

C. J. Rutten, C. Kamphuis, H. Hogeveen, K. Huijps, M. Nielen and W. Steeneveld, 2017. Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows. Computers and Electronics in Agriculture 132, 108-118.

C. J. Rutten, W. Steeneveld, C. Inchaisri and H. Hogeveen, 2014. An ex ante analysis on the use of activity meters for automated estrus detection: To invest or not to invest? Journal of Dairy Science 97, 6869-6887.

C. J. Rutten, W. Steeneveld, C. Kamphuis, K. Huijps and H. Hogeveen, 2015. Development of a predictive model for the onset of calving, In: M. Guarino and D. Berckmans, 7th European Conference on Precision Livestock Farming, Milan, Italy, p. 397 - 405.

C. J. Rutten, A. G. J. Velthuis, W. Steeneveld and H. Hogeveen, 2013. Invited review: Sensors to support health management on dairy farms. Journal of Dairy Science 96, 1928-1952.

M. Saint-Dizier and S. Chastant-Maillard, 2011. Towards an automated detection of oestrus in dairy cattle. Reproduction in Domestic Animals 47, 1056-1061.

M. Saint-Dizier and S. Chastant-Maillard, 2015. Methods and on-farm devices to predict calving time in cattle. The Veterinary Journal 205, 349-356.

M. Sakaguchi, R. Fujiki, K. Yabuuchi, Y. Takahashi and M. Aoki, 2007. Reliability of estrous detection in holstein heifers using a radiotelemetric pedometer located on the neck or legs under different rearing conditions. Journal of Reproduction and Development 53, 819-828.

K. Sananikone, M. J. Delwiche, R. H. BonDurant and C. J. Munro, 2004. Quantitative lateral flow immunoassay for measuring progesterone in bovine milk. Transactions of the Asae 47, 1357-1365.

S. Sato, H. Mizuguchi, K. Ito, K. Ikuta, A. Kimura and K. Okada, 2012. Technical note: Development and testing of a radio transmission ph measurement system for continuous monitoring of ruminal ph in cows. Preventive Veterinary Medicine 103, 274-279.

J. Saumande, 2002. Electronic detection of oestrus in postpartum dairy cows: Efficiency and accuracy of the dec (r) (showheat) system. Livestock Production Science 77, 265-271.

J. M. Schefers, K. A. Weigel, C. L. Rawson, N. R. Zwald and N. B. Cook, 2010. Management practices associated with conception rate and service rate of lactating holstein cows in large, commercial dairy herds. Journal of Dairy Science 93, 1459-1467.

K. M. Scheibe and C. Gromann, 2006. Application testing of a new three-dimensional acceleration measuring system with wireless data transfer (was) for behavior analysis. Behavior Research Methods 38, 427-433.

K. Schirmann, N. Chapinal, D. M. Weary, L. Vickers and M. A. G. von Keyserlingk, 2013. Short communication: Rumination and feeding behavior before and after calving in dairy cows. Journal of Dairy Science 96, 7088-7092.

M. D. Schneider, E. Strandberg, U. Emanuelson, K. Grandinson and A. Roth, 2007. The effect of veterinary-treated clinical mastitis and pregnancy status on culling in swedish dairy cows. Preventive Veterinary Medicine 80, 179-192.

P. C. Schön, K. Hamel, B. Puppe, A. Tuchscherer, W. Kanitz and G. Manteuffel, 2007. Altered vocalization rate during the estrous cycle in dairy cattle. Journal of Dairy Science 90, 202-206.

G. M. Schuenemann, S. Bas, E. Gordon and J. D. Workman, 2013. Dairy calving management: Description and assessment of a training program for dairy personnel. Journal of Dairy Science 96, 2671-2680.

H. Seegers, F. Beaudeau, C. Fourichon and N. Bareille, 1998. Reasons for culling in french holstein cows. Preventive Veterinary Medicine 36, 257-271.

H. Seegers, C. Fourichon and F. Beaudeau, 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Veterinary Research 34, 475-491.

S. Shahinfar, J. N. Guenther, C. D. Page, A. S. Kalantari, V. E. Cabrera, P. M. Fricke and K. A. Weigel, 2015. Optimization of reproductive management programs using lift chart analysis and cost-sensitive evaluation of classification errors. Journal of Dairy Science 98, 3717-3728.

S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke and K. Weigel, 2014. Prediction of insemination outcomes in holstein dairy cattle using alternative machine learning algorithms. Journal of Dairy Science 97, 731-742.

I. M. Sheldon, J. Cronin, L. Goetze, G. Donofrio and H. J. Schuberth, 2009. Defining postpartum uterine disease and the mechanisms of infection and immunity in the female reproductive tract in cattle. Biology of Reproduction 81, 1025-1032.

T. Sing, O. Sander, N. Beerenwinkel and T. Lengauer, 2005. Rocr: Visualizing classifier performance in r. Bioinformatics 21, 3940-3941.

K. Sloth, N. C. Friggens, P. Løvendahl, P. H. Andersen, J. Jensen and K. L. Ingvartsen, 2003. Potential for improving description of bovine udder health status by combined analysis of milk parameters. Journal of Dairy Science 86, 1221-1232.

X. Song and R. van der Tol, 2010. Automatic detection of clinical mastitis in astronaut a3[tm] milking robot, In: The First North American Conference on Precision Dairy Management, Toronto, Canada, p. 154-155.

X. Y. Song, 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.

W. Steeneveld and H. Hogeveen, 2012. Economic consequences of immediate or delayed insemination of a cow in oestrus. Veterinary Record 171, 17-+.

W. Steeneveld and H. Hogeveen, 2015. Characterization of dutch dairy farms using sensor systems for cow management. Journal of Dairy Science 98, 709-717.

W. Steeneveld, H. Hogeveen and A. Lansink, 2015. Economic consequences of investing in sensor systems on dairy farms. Computers and Electronics in Agriculture 119, 33-39.

W. Steeneveld, J. Swinkels and H. Hogeveen, 2007. Stochastic modelling to assess economic effects of treatment of chronic subclinical mastitis caused by streptococcus uberis. Journal of Dairy Research 74, 459-467.

W. Steeneveld, L. W. Tauer, H. Hogeveen and A. G. J. M. Oude Lansink, 2012. Comparing technical efficiency of farms with an automatic milking system and a conventional milking system. Journal of Dairy Science 95, 7391-7398.

W. Steeneveld, L. C. van der Gaag, W. Ouweltjes, H. Mollenhorst and H. Hogeveen, 2010. Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems. Journal of Dairy Science 93, 2559-2568.

W. Steeneveld, T. van Werven, H. W. Barkema and H. Hogeveen, 2011. Cow-specific treatment of clinical mastitis: An economic approach. Journal of Dairy Science 94, 174-188.

E. W. Steyerberg, F. E. Harrell, G. Borsboom, M. J. C. Eijkemans, Y. Vergouwe and J. D. F. Habbema, 2001. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. Journal of Clinical Epidemiology 54, 774-781.

Z. B. Sun, S. Samarasinghe and J. Jago, 2010. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. Journal of Dairy Research 77, 168-175.

K. M. Svennersten-Sjaunja and G. Pettersson, 2008. Pros and cons of automatic milking in europe. Journal of Animal Science 86, 37-46.

J. M. Swinkels, H. Hogeveen and R. N. Zadoks, 2005a. A partial budget model to estimate economic benefits of lactational treatment of subclinical staphylococcus aureus mastitis. Journal of Dairy Science 88, 4273-4287.

J. M. Swinkels, J. G. Rooijendijk, R. N. Zadoks and H. Hogeveen, 2005b. Use of partial budgeting to determine the economic benefits of antibiotic treatment of chronic subclinical mastitis caused by streptococcus uberis or streptococcus dysgalactiae. Journal of Dairy Research 72, 75-85.

U. Tasch and P. G. Rajkondawar, 2004. The development of a softseparator (tm) for a lameness diagnostic system. Computers and Electronics in Agriculture 44, 239-245.

R Development Core Team, 2008. R: A language and environment for statistical computing, In: (Ed.)^(Eds.), ed. R Foundation for Statistical Computing, Vienna, Austria, p.^pp.

B. A. Tenhagen, M. Drillich, R. Surholt and W. Heuwieser, 2004. Comparison of timed ai after synchronized ovulation to ai at estrus: Reproductive and economic considerations. Journal of Dairy Science 87, 85-94.

P. T. Thomsen, L. Munksgaard and F. A. Togersen, 2008. Evaluation of a lameness scoring system for dairy cows. Journal of Dairy Science 91, 119-126.

M. Titler, M. G. Maquivar, S. Bas, P. J. Rajala-Schultz, E. Gordon, K. McCullough, P. Federico and G. M. Schuenemann, 2015. Prediction of parturition in holstein dairy cattle using electronic data loggers. Journal of Dairy Science 98, 5304-5312.

B. J. Tolkamp, M. J. Haskell, F. M. Langford, D. J. Roberts and C. A. Morgan, 2010. Are cows more likely to lie down the longer they stand? Applied Animal Behaviour Science 124, 1-10.

J. Tschmelak, N. Kappel and G. Gauglitz, 2005. Tirf-based biosensor for sensitive detection of progesterone in milk based on ultra-sensitive progesterone detection in water. Analytical and Bioanalytical Chemistry 382, 1895-1903.

USDA, 2007. Heifer calf health and management practices on u.S. Dairy operations, 2007, USDA: APHIS, Fort Collins, CO, USA.

D. Val-Arreola, E. Kebreab, J. Dijkstra and J. France, 2004. Study of the lactation curve in dairy cattle on farms in central mexico. Journal of Dairy Science 87, 3789-3799.

M. van Asseldonk, R. B. M. Huirne, A. A. Dijkhuizen and A. J. M. Beulens, 1999a. Dynamic programming to determine optimum investments in information technology on dairy farms. Agricultural Systems 62, 17-28.

M. van Asseldonk, A. W. Jalvingh, R. B. M. Huirne and A. A. Dijkhuizen, 1999b. Potential economic benefits from changes in management via information technology applications on dutch dairy farms: A simulation study. Livestock Production Science 60, 33-44.

B. H. P. van den Borne, T. Halasa, G. van Schaik, H. Hogeveen and M. Nielen, 2010. Bioeconomic modeling of lactational antimicrobial treatment of new bovine subclinical intramammary infections caused by contagious pathogens. Journal of Dairy Science 93, 4034-4044.

S. G. A. van der Drift, R. Jorritsma, J. T. Schonewille, H. M. Knijn and J. A. Stegeman, 2012. Routine detection of hyperketonemia in dairy cows using fourier transform infrared spectroscopy analysis of beta-hydroxybutyrate and acetone in milk in combination with test-day information. Journal of Dairy Science 95, 4886-4898.

P. P. J. van der Tol, J. H. M. Metz, E. N. Noordhuizen-Stassen, W. Back, C. R. Braam and W. A. Weijs, 2002. The pressure distribution under the bovine claw during square standing on a flat substrate. Journal of Dairy Science 85, 1476-1481.

P. P. J. van der Tol, J. H. M. Metz, E. N. Noordhuizen-Stassen, W. Back, C. R. Braam and W. A. Weijs, 2003. The vertical ground reaction force and the pressure distribution on the claws of dairy cows while walking on a flat substrate. Journal of Dairy Science 86, 2875-2883.

R. van der Tol and A. van der Kamp, 2010. Time series analysis of live weight as health indicator, In: The First North American Conference on Precision Dairy Management, Toronto, Canada, p. 230-231.

A. J. H. van Es, 1978. Feed evaluation for ruminants. I. The systems in use from may 1977-onwards in the netherlands. Livestock Production Science 5, 331-345.

A. T. M. van Knegsel, H. van den Brand, J. Dijkstra, W. M. van Straalen, R. Jorritsma, S. Tamminga and B. Kemp, 2007. Effect of glucogenic vs. Lipogenic diets on energy balance, blood metabolites, and reproduction in primiparous and multiparous dairy cows in early lactation. Journal of Dairy Science 90, 3397-3409.

F. J. S. van Soest, I. Santman-Berends, T. Lam and H. Hogeveen, 2016. Failure and preventive costs of mastitis on dutch dairy farms. Journal of Dairy Science 99, 8365-8374.

I. Vermeij, 2012. Kwin-v 2012-2013, handboek 23 (quantitative information livestock farming 2012-2013, handbook 23). Livestock Research, Wageningen UR, Lelystad, the Netherlands.

N. Vreeburg, 2010. Precision management on two dutch dairy farms by use of herd navigator [r], In: The First North American Conference on Presicion Dairy Management, Toronto, Canada, p. 104-105.

C. M. Wathes, H. H. Kristensen, J. M. Aerts and D. Berckmans, 2005. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Wageningen Academic Publishers, Wageningen.

K. Welbeck, P. Leonard, N. Gilmartin, B. Byrne, C. Viguier, S. Arora and R. O'Kennedy, 2011. Generation of an anti-nagase single chain antibody and its application in a biosensor-based assay for the detection of nagase in milk. Journal of Immunological Methods 364, 14-20.

H. Wickham, 2007. Reshaping data with the reshape package. Journal of Statistical Software 21, 1-20.

H. Wickham and R. Francois, 2014. Dplyr: A grammar of data manipulation, version 0.1, A fast, consistent tool for working with data frame like objects, both in memory and out of memory.

W. F. Williams, D. R. Yver and T. S. Gross, 1981. Comparison of estrus detection techniques in dairy heifers. Journal of Dairy Science 64, 1738-1741.

J. B. M. Wilmink, 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livestock Production Science 16, 335-348.

J. J. Windig, M. P. L. Calus and R. F. Veerkamp, 2005. Influence of herd environment on health and fertility and their relationship with milk production. Journal of Dairy Science 88, 335-347.

R. R. Wolfe, G. E. Ward and S. C. Sharma, 1972. Method for electronic detection of bovine mastitis. Transactions of the Asae 15, 500-&.

P. D. P. Wood, 1967. Algebraic model of lactation curve in cattle. Nature 216, 164-&.

A. Zecconi, R. Piccinini, G. Giovannini, G. Casirani and R. Panzeri, 2004. Clinical mastitis detection by on-line measurements of milk yield, electrical conductivity and milking duration in commercial dairy farms. Milchwissenschaft-Milk Science International 59, 240-243.

A. Zeileis and G. Grothendieck, 2005. Zoo: S3 infrastructure for regular and irregular time series. Journal of Statistical Software 14.

Summary

Since the 1980s, efforts have been made to develop sensors that measure a parameter from an individual cow. The development started with individual cow recognition and was followed by sensors that measure the electrical conductivity of milk and pedometers that measure activity. Some sensors like activity meters, electrical conductivity, weight floors and somatic cell count sensors are commercially available. Adoption has in general been low and mainly driven by the AMS, with a clear exception for estrus detection. In practice, the economic benefits of using sensor systems has not been proven. So, to make sensors live up to their full potential there is a need for research to shift from technical development towards practical applications and integration with operational farm management. Estrus detection sensors can have a good detection performance and are currently applied by farmers in practice, therefore this thesis focusses on sensors that support reproductive management. The main objective of this thesis is to study the utility of sensor technology to support reproductive management on dairy farms. This main objective was split in five sub objectives that each study a part of the main objective and were discussed in the separate chapters of this thesis.

The objective of chapter 2 was to provide a structured overview of the published sensor systems for dairy health management. The development of sensor systems can be described by the following four levels: I. techniques that measure something about the cow (e.g., activity); II. interpretations that summarize changes in the sensor data (e.g., increase in activity) to produce information about the cow's status (e.g., estrus); III. integration of information where sensor information is supplemented with other information (e.g., economic information) to produce an advice (e.g., whether to inseminate a cow or not); and IV. the farmer makes a decision or the sensor system takes the decision autonomously (e.g., the inseminator is called). This review has structured a total of 126 publications describing 139 sensor systems and compared them based on the four levels. The publications were published in the ISI database from January 2002 until June 2012 or in the proceedings of three conferences on precision (dairy) farming in 2009, 2010, and 2011. The performance of sensor systems varies based on the choice of gold standards, algorithms, and test sizes (number of farms and cows). Studies on sensor systems for mastitis and estrus have shown that sensor systems are brought to a higher level, however there is still a need to improve detection performance. Studies on sensor systems for locomotion problems have shown that the search continues for the most appropriate indicators, sensor techniques, and gold standards. Studies on metabolic problems show that it is still unclear which indicator reflects best the metabolic problems that should be detected. No systems with integrated decision support models have been found.

The objective of chapter 3 was to develop a sensor system that predicts the moment of calving. Observation of a cow prior to calving is important because *dystocia* can occur, which requires timely intervention to mitigate adverse effects on both cow and calf. In this study, 400 cows on a Dutch dairy farm were equipped with sensors. The sensor was a single device in an ear tag, which synthesized cumulative activity, rumination activity, feeding activity, and

temperature on an hourly basis. Data were collected during a one-year period. During this period, the exact moment of 417 calvings was recorded using camera images of the calving pen taken every five minutes. In total, 114 calving moments could be linked with sensor data. The moment at which calving started was defined as the first camera snapshot with visible evidence that the cow was having contractions or had started labor. Two logit models were developed: a model with the expected calving date as independent variable and a model with additional independent variables based on sensor data. The areas under the curves of the Receiver Operating Characteristic were 0.885 and 0.929 for these models, respectively. Results indicate that the inclusion of sensor data improves the prediction of the start of calving; therefore the sensor data has value for the prediction of the moment of calving. The model with the expected calving date and sensor data had a sensitivity of 21.2% at an one-hour time window and 42.4% at a three-hour time window, both with a false positive rate of 1%. This indicates that prediction of the specific hour in which calving started was not possible with a high accuracy.

The objective of chapter 4 was to determine the economic benefits of a sensor system that predicts the start of calving in dairy cattle on typical mid-sized (300 cows) dairy farms in the United States. Furthermore a specialized calving sensor already used in the equine sector was compared to a estrus detection system with an additional algorithm for calving detection. Dynamic discrete event Monte Carlo simulation was used to estimate these benefits. Stochastic information for input variables was derived from scientific literature, survey results, and the authors' expertise. Effects on insemination costs, time spent observing close-up cows, assisting cows during calving, days open, treatment, culling and lost milk production due to metritis, stillbirth rate, and the costs, lifetime, time to apply the sensor, and subscription costs related to the sensor systems were considered. Marginal profit of the equine sensor was on average $0.97 per cow on mid-sized dairy farms with a range from - $21.70 to $963.67. This profit mainly consisted of a reduction in labor costs, and a reduction in metritis incidence and stillbirth rates. The alternative sensor was already used for estrus detection, therefore no investment cost were incurred. This caused profit for the estrus sensor the be higher that the profit of the equine sensor. The most influential inputs were the labor costs regarding calving management.

The objective of chapter 5 was to analyze the economic benefits of a sensor system for detection of estrus and appraises the feasibility of an investment in such a system. A stochastic dynamic simulation model was used to simulate reproductive performance of a dairy herd. The number of cow places in this herd was fixed at 130. The model started with 130 randomly drawn cows (in a Monte Carlo process) and simulated calvings and replacement of these cows in subsequent years. Model inputs were derived from real farm data and expertise. For the analysis, visual detection by the farmer ("without" situation) was compared to automated detection with activity meters ("with" situation). For visual estrus detection, an estrus detection rate of 50% and a specificity of 100% was assumed. For automated estrus detection, an estrus detection rate of 80% and a specificity of 95% was

assumed. The results of the cow simulation model were used to estimate the difference between the annual net cash flows in the with and without situations (marginal financial effect) and the internal rate of return (IRR) as profitability indicators. It was estimated that the initial investment in activity meters would cost €17,728 for a herd of 130 cows, with an additional cost of €90 per year for the replacement of malfunctioning activity meters. The average marginal financial effect in the with and without situations was € 2,827 for the baseline scenario, with an average IRR of 11%. The IRR is a measure of the return on invested capital. Investment in activity meters was generally profitable. The most influential assumptions on the profitability of this investment were the assumed culling rules and the increase in sensitivity of estrus detection between the without and the with situation.

The objective of chapter 6 was to develop a prognostic model for the likelihood of successful first insemination. Parameters are considered for the model, which are readily available on-farm at the moment a farmer makes breeding decisions. In the first step, variables are selected for the prognostic model, which have prognostic value for the likelihood of a successful first insemination. In the second step, farm effects on the likelihood of a successful insemination are quantified and the prognostic model is cross-validated. Logistic regression with a random effect for farm was used to develop the prognostic model. Insemination and test-day milk production data from 2,000 commercial Dutch dairy farms were obtained; 190,541 first inseminations from this dataset were used for model development. Variables were selected in a forward selection and backward elimination, based on the Akaike Information Criterion. The variables that contributed most to the model were the random farm effect, relative production factor, and milk yield at insemination. The cross-validation shows that the prognosis of insemination success closely resembles the mean insemination success observed in the dataset. Insemination success depends on physiological conditions of the cow, which are approximated indirectly by the production and reproduction data that is routinely recorded at the farm. The model cannot be used as a detection model that distinguishes cows that conceive from cows that do not.

We demonstrated that utility of sensors for reproductive management can be found in economic benefits (estrus and calving detection), reduction of labor (calving and estrus detection) and more detailed management information (prognosis of insemination success). So, automated estrus detection aids reproductive management.

From this thesis the following conclusions can be drawn:

The developed theoretical framework describes four levels of sensor development, which should all be included in proper development of sensor systems. The literature review showed that no studies developed sensor systems with regard to management and decision support.

It was possible to improve the prediction of the start of calving compared to a model that only uses the expected calving date. However, predicting the start of calving within an hour was not possible with a high sensitivity and specificity.

There was financial merit in the use of calving detection, because the sensor system enables more timely intervention by the farmer. The uncertainty about the positive effects was large, which caused a wide range in the simulated financial benefits.

Investment in a sensor for estrus detection was on average profitable with a return on investment of 11%. Profitability was influenced most by the heuristic culling rules and the expected increase of the estrus detection rate between detection by visual observation and the sensor.

Routinely collected farm data can be used to estimate a prognosis on insemination success and be used to determine whether an individual cow has a higher or lower than average likelihood of insemination success. Integration of this prognostic model with an estrus detection sensor has potential.

Currently farmers only adopt sensors for estrus detection or because they were standard with an AMS. A reason for this is that sensor systems do not produce clear information for farmers. Sensor technology should be focused on management support of applications. Labor benefits of sensors are important for adoption of sensors by farmers, farmers value flexibility, increased family time and less physical workload as benefits. However, economic evaluations of technical solutions are unable to quantify these benefits. Sensor research should consider the preferences of farmers regarding labor. For the appraisal of sensor technology new methods to value labor benefits of sensor are needed. Furthermore, in sensor development societal acceptance should be an important consideration. Animal rights activists may frame the use of sensors as a form of industrialized farming. Only using technical arguments and considerations to explain the benefits of sensors will hamper the societal acceptance of modern dairy farming. Application of sensors on dairy farms should be communicated smartly to society in terms that relate the values of citizens.

Samenvatting

In de jaren '80 van de vorig eeuw werden de eerste sensoren die een parameter op individueel koe niveau meten ontwikkeld. Koe herkenning was een belangrijke eerste stap in deze ontwikkeling en werd gevolgd door sensoren die geleidbaarheid van de melk meten en stappentellers. Sommige sensoren zoals activiteit meters (een doorontwikkelde variant op de stappenteller), geleidbaarheidssensoren, weegvloeren en celgetalmeters zijn commercieel verkrijgbaar. Vooral sensoren voor tochtdetectie worden toegepast, maar andere sensoren hebben veehouders vooral omdat deze onderdeel zijn van een melkrobot. Wat sensoren nu precies opleveren op een melkveebedrijf is tot nu toe niet vastgesteld. De potentie die sensoren hebben komt tot nu toe niet uit de verf. Om hier verandering in te brengen moet het onderzoek naar sensoren zich verplaatsen van technische ontwikkeling naar de ontwikkeling van praktische toepassingen die een veehouder in zijn dagelijks management kan integreren. Vooral voor tochtdetectie zijn er op dit moment sensoren die goede detectieresultaten laten zien en op commerciële melkveebedrijven worden gebruikt. Daarom is het hoofddoel van dit proefschrift het bestuderen van het nut van sensoren als ondersteuning van het reproductiemanagement op melkveebedrijven. Dit hoofddoel is opgesplitst in vijf subdoelen die elk een onderdeel van dit hoofddoel bestuderen en in de verschillende hoofdstukken van dit proefschrift zijn bestudeerd en uitgewerkt.

De doelstelling van hoofdstuk 2 was om een gestructureerd overzicht te geven van sensoren die gezondheid van koeien monitoren en beschreven zijn in de wetenschappelijke literatuur. Voor dit hoofdstuk is de ontwikkeling van een sensor met een theoretisch raamwerk in de volgende vier niveaus ingedeeld: I. technieken die iets meten van een koe (bijv. activiteit); II. interpretaties die veranderingen in de sensordata samenvatten (bijv. een piek in activiteit) en daarmee informatie produceren die iets over de status van de koe zegt (bijv. tochtigheid); III. integratie van informatie waarbij sensordata worden gecombineerd met andere informatie (bijv. economische informatie) om een advies te geven (bijv. of de koe op dit moment geïnsemineerd moet worden of niet); en IV. de veehouder neemt een beslissing of het sensorsysteem neemt autonoom een beslissing (bijv. de inseminator wordt gebeld). Het literatuuroverzicht bevatte totaal 126 wetenschappelijke publicaties die 139 sensorsystemen beschrijven en vergeleek deze systemen op basis van de vier niveaus van ontwikkeling. De publicaties zijn gepubliceerd in de ISI database tussen januari 2002 en juni 2012 of in de congresbundel van drie congressen over precisie melkveehouderij in 2009, 2010 en 2011. De sensorsystemen verschilden in wiskundig model, referentietest en studieomvang (aantal koeien en bedrijven) waardoor de prestaties sterk varieerden. Studies naar sensoren voor mastitis- en tochtdetectie ontwikkelden detectiesystemen tot een hoger niveau, maar er bleef noodzaak tot het verbeteren van de detectieprestaties. Studies naar sensoren voor de detectie van klauw- en pootproblemen laten zien dat de juiste indicatoren, sensortechniek en referentietest nog niet zijn gevonden. Studies naar sensoren voor metabole problemen en voedingsproblemen laten zien dat het onduidelijk is welk indicator dit soort problemen goed in beeld brengt. Geen enkele studie beschreef een sensorsysteem dat ook een geïntegreerde beslissingsondersteuning bevatte.

De doelstelling van hoofdstuk 3 was om een sensorsysteem te ontwikkelen dat het moment van afkalven voorspelt. Het is belangrijk om een drachtige koe te observeren voorafgaand aan het afkalven omdat *dystocia* kan voorkomen en dan is tijdig ingrijpen nodig om zo de schadelijke effecten voor koe en kalf te reduceren. In deze studie zijn 400 koeien op een Nederlands melkveebedrijf met een sensor uitgerust. De sensor was een apparaat in het oormerk van de koe dat per uur cumulatieve activiteit, herkauwactiviteit, vreettijd en gemiddelde temperatuur genereerd. Gedurende een jaar werd er data verzameld. In dit jaar werden er 417 afkalvingen met behulp van een camera vastgelegd. De camera nam iedere 5 minuten een foto en hiermee werd het tijdstip waarop het afkalven begon bepaald. Het begintijdstip was gedefinieerd als de eerste foto waarop de koe zichtbaar persweeën had. Voor 114 afkalvingen kon het begintijdstip aan sensordata gekoppeld worden. Twee logistische regressie modellen werden ontwikkeld: een model met de verwachte afkalfdatum als onafhankelijke variabele en een model met zowel verwachte afkalfdatum als onafhankelijke variabelen op basis van sensordata. Voor deze modellen was de respectievelijke oppervlakte onder de grafiek van de Receiver Operating Characteristic 0.885 en 0.929. Dit resultaat wijst er op dat het toevoegen van sensordata aan het model de voorspelling van de start van afkalven verbeterd en sensordata heeft dus waarde voor het voorspellen van de start van afkalven. Als 1% aan vals positieve alarmen getolereerd werd had het model met de verwachte afkalfdatum en sensordata een sensitiviteit van 21.2%voor het voorspellen van afkalven binnen een uur na een alarm van het model. Voor het voorspellen van afkalven binnen drie uur na een alarm had het model een sensitiviteit van 42.4%. Dit wijst erop dat met een hoge precisie voorspellen van het exacte uur waarin afkalven begint niet mogelijk was.

De doelstelling van hoofdstuk 4 was om de economische waarde van een sensorsysteem dat het begin van het afkalfproces voorspelt te bepalen. Deze berekeningen zijn gedaan voor een middelgroot (300 koeien) Amerikaans melkveebedrijf. Er zijn twee sensorsystemen vergeleken, namelijk een gespecialiseerde sensor uit de paardenhouderij en een sensorsysteem dat al voor tochtdetectie wordt gebruikt en waaraan een detectiemodel voor het voorspellen van afkalven is toegevoegd. De economische waarde van een sensor die afkalven voorspelt is geschat met een Monte Carlo simulatiemodel. Model variabelen werden beschreven met kansverdelingen, en deze werden op wetenschappelijke literatuur, enquêteresultaten en de expertise van de auteurs gebaseerd. De effecten op inseminatiekosten, tijd die besteed werd aan het observeren van hoog drachtige koeien, tijd besteed aan het verlossen van kalvende koeien, aantal dagen open, behandelingen, afvoer van koeien en misgelopen melkproductie als gevolg van metritis, percentage doodgeboren kalveren en de bijbehorende kosten, levensduur van de sensor, tijd die nodig is om de sensor bij de koeien in te brengen en de abonnementskosten van de sensor werden meegenomen. Per koe leverde de paardensensor $0.97 op voor een middelgroot Amerikaans melkveebedrijf. De spreiding voor dit resultaat was -$21.70 tot $963.67 per koe. Deze opbrengsten bestonden vooral uit arbeidsbesparing en een vermindering van de hoeveelheid metritis en doodgeboren kalveren. De alternatieve sensor werd al gebruikt voor

tochtdetectie, en daarom was er geen extra investering nodig. Hierdoor leverde de tochtsensor meer op dan de paardensensor. De arbeidsbesparing en kosten van arbeid hadden de meeste invloed op de economische waarde van een afkalfsensor.

De doelstelling van hoofdstuk 5 was om de economische waarde van een sensorsysteem voor tochtdetectie en winstgevendheid van investeren in een dergelijk systeem te bepalen. De reproductie en melkproductie van een melkkoe werd met een stochastisch, dynamisch simulatiemodel gesimuleerd. Er werd uitgegaan van een koppel met plek voor 130 melkkoeien. Het simulatiemodel startte met 130 willekeurig gekozen koeien (in een Monte Carlo proces) en simuleerde afkalven en vervanging van deze koeien in de daarop volgende jaren. Het model was gebaseerd op bedrijfsgegevens van Nederlandse melkveebedrijven en expertise van de auteurs. In de economische analyse werden tochtdetectie door de veehouder (de situatie "zonder") en tochtdetectie door activiteit meters (de situatie "met") vergeleken. Voor de situatie zonder tochtdetectiesensoren werd aangenomen dat de sensitiviteit 50% was en de specificiteit 100%. Voor de situatie met tochtdetectiesensoren werd aangenomen dat de sensitiviteit 80% was en de specificiteit 95%. De resultaten van het koe simulatiemodel werden gebruikt om het verschil per jaar in kasstroom (extra kasstroom) tussen de situaties met en zonder tochtdetectiesensoren te bepalen. Daarnaast werd ook de interne opbrengstvoet (IRR) bepaald, dit is een maat voor het rendement op geïnvesteerd kapitaal. De investering in activiteit meters was geschat op €17.728,- voor een koppel van 130 koeien met €90 extra kosten per jaar om kapotte activiteit meters te vervangen. De extra kasstroom per jaar was €2.827,- met een gemiddelde IRR van 11%. Investeren in een sensorsysteem voor tochtdetectie was over het algemeen winstgevend. De winstgevendheid werd vooral bepaald door hoe afvoer en vervanging van koeien werden gemodelleerd en het verschil in sensitiviteit tussen de situaties met en zonder tochtdetectiesensor.

De doelstelling van hoofdstuk 6 was om een prognostisch model te ontwikkelen voor de kans dat een koe drachtig wordt na de eerste inseminatie. Voor dit model werden parameters gezocht die op een melkveebedrijf routinematig worden vastgelegd en beschikbaar zijn op het moment dat de veehouder beslist of hij deze koe wil aanhouden voor een volgende lactatie. Stap een was het selecteren van variabelen die mogelijk prognostische waarde hebben voor de kans dat een koe drachtig wordt na de eerste inseminatie. Stap twee was om bedrijfseffecten te kwantificeren en het prognostisch model kruislings te valideren. Logistische regressie met een random effect voor bedrijf werd gebruik om het prognostisch model te ontwikkelen. Melkproductie- en reproductiedata van 2.000 commerciële Nederlandse melkveebedrijven waren beschikbaar en 190.541 eerste inseminaties uit deze dataset werden gebruikt in deze studie. Het Akaike Informatie Criterium (**AIC**) werd gebruikt om variabelen voor het model te selecteren en vervolgens onnodige variabelen te elimineren. De variabelen random bedrijfseffect, lactatiewaarde en melkproductie op het moment van insemineren droegen het meeste bij aan het model. De kruislingse validatie liet zien dat de prognose voor een inseminatiesucces het werkelijke gemiddelde inseminatiesucces dicht benaderde. Inseminatiesucces hangt af van de

fysiologische conditie van de koe en deze wordt indirect benadert door de melkproductie- en reproductiedata die op een melkveebedrijf worden vastgelegd. Het model kan daarom niet worden gebruikt om koeien te identificeren die zeker drachtig zullen worden.

We hebben gedemonstreerd dat het nut van sensoren voor reproductiemanagement gevonden kan worden in economische waarde (tocht- en afkalfdetectie), arbeidsbesparing (tocht en afkalfdetectie) en meer specifieke managementinformatie (prognose voor inseminatiesucces).

Uit dit proefschrift kunnen de volgende conclusies getrokken worden:

Het theoretisch raamwerk beschrijft vier niveaus van sensorontwikkeling die allemaal in ogenschouw genomen moeten worden bij ontwikkelingstrajecten van sensorsystemen. Het literatuur overzicht laat zien dat wetenschappelijke studies management en beslissingsondersteuning buiten beschouwing lieten bij de ontwikkeling van sensorsystemen.

De voorspelling van het startmoment van afkalven kon verbeterd worden door sensordata toe te voegen aan een voorspellend model naast de verwachte afkalfdatum. De voorspelling van het startmoment binnen een uur na een alarm van het model was echter niet mogelijk met een hoge sensitiviteit en specificiteit.

Een sensor die afkalven van hoog drachtige koeien voorspelt levert financieel gezien iets op want het maakt meer tijdig ingrijpen door de veehouder mogelijk. De onzekerheden over deze positieve effecten waren groot, dit zorgde voor een grote spreiding in de gesimuleerde financiële voordelen.

Het rendement op investeren in een sensor voor tochtdetectie was gemiddeld genomen 11%. Dit rendement werd het sterkst beïnvloed door hoe afvoer en vervanging van koeien werden gemodelleerd en het verschil in sensitiviteit tussen de situaties met en zonder tochtdetectiesensor

Melkproductie- en reproductiedata die routinematig op melkveebedrijven worden vastgelegd kunnen worden gebruikt om een prognose voor inseminatiesucces te schatten. Met die prognose kan bepaald worden of een koe een boven of beneden gemiddelde kans heeft om na één keer insemineren drachtig te worden. Integratie van dit prognostisch model met sensorsystemen heeft potentieel.

Op dit moment schaffen veehouders enkel tochtdetectiesensoren aan of sensoren die onderdeel zijn van de melkrobot. Een reden hiervoor is dat veel sensorsystemen geen heldere informatie aan de veehouder presenteren. Sensortechnologie zou meer moeten focussen op toepassingen die het management van melkvee ondersteunen. Voor veehouders is arbeidsverlichting een belangrijke reden om sensoren aan te schaffen. Veehouders vinden flexibiliteit, vrije tijd en tijd voor familie en vermindering van fysiek

zware arbeid belangrijk. Economische studies slagen er echter niet in om dit soort voordelen van technologische oplossingen te kwantificeren. Bij het ontwikkelen van sensorsystemen zou men de wensen van veehouders met betrekking tot arbeid voor ogen moeten houden. Om de economische waarde van sensortechniek te schatten zijn nieuwe methodes nodig om arbeidsverlichting te waarderen. Ook moet met de maatschappelijke acceptatie van sensortechniek rekening worden gehouden. Activisten kunnen het gebruik van sensoren framen als een vorm van geïndustrialiseerde veehouderij. Door enkel technische argumenten en overwegingen te gebruiken om de voordelen van sensoren uit te leggen kan men burgers vervreemden van de moderne veehouderij. Daarom is het van belang het gebruik van sensoren op intelligente wijze te communiceren naar de samenleving in termen die aansluiten bij de beleving van burgers.

Dankwoord

Wat ontzettend vreemd om nu te schrijven "ik heb mijn proefschrift af"! Toen ik ruim 10 jaar geleden in het diepe sprong en in Nijmegen ging studeren had ik nooit bedacht dat ik vandaag hier zou staan. Voor het eerst weg van Texel en op eigen benen proberen te staan, het is een boeiende reis geworden. Ook het starten van mijn promotietraject was een sprong in het diepe. Gerard, het is aan jou te danken dat ik bij bedrijfseconomie in beeld kwam als onderzoeker. Henk en Annet, zonder jullie overtuigingskracht en vertrouwen in mij vanaf de start was ik waarschijnlijk nooit begonnen aan het avontuur van promoveren. Het heeft mij ontzettend veel geleerd over melkveehouderij, de wereld en mijzelf, want wat kom je jezelf soms hard tegen in een promotietraject. Het was niet altijd een makkelijk traject en er zijn een paar diepe dalen geweest. Maar ik ben er toch weer uitgekomen, geloof heeft me daarin steeds gesterkt. Of het nu tegen zat of dat het juist voor de wind ging, ik voelde me op adelaarsvleugels gedragen en heb op de goede afloop durven vertrouwen.

Er zijn een aantal mensen zonder wie ik de eindstreep niet gehaald zou hebben. Henk, bedankt voor alle nieuwe ideeën en uitdagingen die je bood gedurende de afgelopen jaren. Wilma, bedankt voor je hulp om meer structuur aan te brengen en alles concreet en behapbaar te houden. Mirjam, je had meer dan eens ontzettend vervelend gelijk, je kritische blik hielp enorm om het proefschrift scherper te krijgen. Henk, Wilma en Mirjam, het was ontzettend plezierig om met jullie samen te werken.

Alle collega's uit Utrecht en Wageningen bedankt voor de collegialiteit, zinvolle en minder zinvolle gesprekken en vooral de gezelligheid. Claudia, bedankt voor je steun en betrokkenheid de afgelopen jaren. Felix, het was fijn een kantoor met je te delen. We hebben veel goeie inhoudelijke discussies gehad over onze behoorlijk verschillende promotieonderwerpen. Maar bovenal was het fijn om af en toen over iets anders dan onderzoek te praten. Jaap en Jamal, dank dat jullie vandaag mijn paranimfen zijn en de afgelopen jaren het lot van promovendus hebben gedeeld. Anne, Jeanette, Ilona, Annet, Fros, Carroll en Wout dank voor al jullie hulp en geduld bij het afhandelen van alle praktische, administratieve en financiële perikelen de afgelopen jaren.

Pa, Ma, Opa, familie en vrienden, dank voor jullie geduld de afgelopen tijd omdat ik toch wel wat druk was met mijn proefschrift en meer dan een tikje warrig. Fijn dat jullie soms een luisterend oor boden of juist even voor afleiding zorgden.

Tot slot: Bedankt!

Voor wie zich aangesproken voelt, zou moeten voelen of wil voelen. Je hulp, geduld of betrokkenheid waardeer ik enorm. Dit stuk tekst is te kort om alle hartelijke mensen die ik de afgelopen jaren ben tegengekomen persoonlijk te noemen en op de juiste manier te bedanken.

Nu op naar de volgende sprong in het diepe. Terug naar Texel en de koeien. Maar ik denk dat de sensoren nog wel even in mijn leven blijven.

About the author

Niels Rutten is geboren op 17 mei 1986 op Texel. Hij groeide op, op een boerderij met zowel melkvee als akkerbouw op. Al van jongs af aan had Niels een passie voor koeien. Hij behaalde zijn atheneum diploma aan OSG de Hogeberg in Den Burg in 2005. Dit markeerde zijn vertrek van het eiland, om het grote avontuur te starten van studeren en wonen in de stad op de vaste wal. De fysiologie van dier en mens en het ontstaan van ziekten intrigeerde Niels altijd al. Dus behaalde hij een BSc graad in de Medische Biologie aan de Radboud Universiteit Nijmegen. Dit deed Niels echter wel beseffen dat hij koeien te veel miste om met deze studie door te gaan. Daarom veranderde hij van richting en begon in Wageningen aan de MSc Dierwetenschappen met focus op fysiologie en veevoeding. Tijdens deze Master voltooide Niels een minor thesis in bedrijfseconomie bij Alfa Accountants. Hier kruiste zijn wegen Gerard Giesen die hem introduceerde bij de leerstoelgroep Bedrijfseconomie om aan een onderzoeksproject naar de kosten van monitoring van aviaire influenza te beginnen. Dit was het moment dat Henk Hogeveen en Annet Velthuis hem op de mogelijkheid van promotieonderzoek naar het nut van sensor technologie bij melkvee wezen. Na enige aarzeling startte hij en ontdekte hoe leuk en uitdagend dit onderzoek was. Daarnaast gaf het werken bij het Departement Gezondheidszorg Landbouwhuisdieren van de Faculteit Diergeneeskunde Niels de kans onderwijs aan diergeneeskunde studenten te geven, wat hij met veel plezier heeft gedaan. Niels vervolgt zijn carrière als melkveehouder op Texel en zoekt momenteel naar mogelijkheden om ook bij de ontwikkeling van sensoren betrokken te blijven.

Niels Rutten was born on Texel, the Netherlands on 17 May 1986. He grew up on a farm with both dairy cows and arable crops. From a young age Niels had a passion for dairy cattle. He obtained his atheneum diploma from OSG de Hogeberg in Den Burg (Texel), the Netherlands) in 2005. This did also marked his departure from the island, to start the big adventure of studying and living in a city on the mainland. The physiology of animals and humans and the development of diseases was always intriguing for Niels. So, he obtained a BSc degree in Medical Biology at the Radboud University Nijmegen. This made Niels realise, however, that he did miss cows too much to continue this study. Therefore, he changed directions to Wageningen. Here he started an MSc in Animal Sciences and focussed on physiology and animal nutrition. During this Master, Niels did a minor thesis in business economics at Alpha Accountants. Here he met Gerard Giesen who, subsequently introduced Niels to the business economics group to work on a research project about the costs of surveillance for avian influenza. This was the moment that Henk Hogeveen and Annet Velthuis brought up the opportunity to do a PhD on the utility of sensor technology. Hesitant at first he started and became to enjoy the challenges of research. Furthermore, working at the Department of Farm Animal Health of the Faculty of Veterinary Medicine gave Niels the opportunity to get involved in teaching veterinary students, which he did with great pleasure. Niels will continue his career as a dairy farmer on Texel and is currently searching for a possibility to stay involved in the development of sensors.

Education certificate

Cornelis Johannis Rutten

Wageningen School of Social Sciences (WASS)

Completed Training and Supervision Plan

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
Econometrics, AEP 21306	WUR	2012	6
Economic models, AEP30806	WUR	2013	6
Applied Bayesian statistics	PhD course Epidemiology UU	2012	1.5
Writing Research proposal	BEC, WUR	2011-2012	6
B) General research related competences			
Introduction course	WASS	2012	1
'Cost analysis of various Low Pathogenic Avian Influenza surveillance systems in the Dutch layer sector'	VEEC, Utrecht	2011	1
'Cost minimization analysis for low Pathogenic Avian Influenza surveillance systems based on blood or egg samples' (Poster)	ISVEE, Maastricht	2012	1
'Overview of Sensors in Health Management of Dairy Cattle'	IDF standing committee Animal Health, Amersfoort	2012	1
'Overzicht: Sensoren voor koe gezondheid in de Melkveehouderij'	Animal Health service, Nuenen	2012	
'Analysis of investment in an oestrus detection system for dairy farms'	ECPLF, Leuven	2013	1
'Overview of published sensor systems for detection of oestrus and lameness in dairy cows'	ECPLF, Leuven	2013	1
'Analysis of investment in an estrus detection system for dairy farms' (Poster)	North American Conference on Precision Dairy Farming, Rochester Minnesota	2013	1
'Sensor systems for dairy cow health management: A review'	North American Conference on Precision Dairy Farming, Rochester Minnesota	2013	1

'Can sensor technology benefit mastitis control?'	British Mastitis Conference, BirmingHam	2013	1
'The potential of using sensor data to predict the moment of calving for dairy cows'	EAAP, Kopenhagen	2014	1
'Review of sensor systems for Oestrus detection and potential of sensors to predict onset of calving'	ESDAR, Helsinki	2014	1
'Development of a predictive model for the onset of calving'	ECPLF, Milaan	2015	1
PhD-meetings Business Economics Group	WUR	2011-2015	4
C) Career related competences/personal development			
Scientific writing	Language service Centre, WUR	2012	1.8
Reviewing Scientific papers (5 papers)	Journal of Dairy Science, Biosystems Engineering and Revista Colombiana de Ciencias Pecuarias	2015-2016	5
Teaching BSc and MSc courses	UU, WUR, Wageningen Academy		4
Total			**46.3**

*One credit according to ECTS is on average equivalent to 28 hours of study load