



Definition of normal and abnormal milk at time of milking

Abilities of automatic milking systems to detect and separate abnormal milk

April 2004

Information

This report is produced within the EU project *Implications of the introduction of automatic milking on dairy farms* (QLK5 -2000-31006) as part of the EU-program 'Quality of Life and Management of Living resources'.

Liability

The content of this publication is the sole responsibility of its publisher, and does not necessarily represent the views of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use, which might be made of the following information.

Availability

This report is freely available for every person interested at:

[Http://www.automaticmilking.nl](http://www.automaticmilking.nl)

Definition of normal and abnormal milk at time of milking

Abilities of automatic milking systems to detect and separate abnormal milk

Morten Dam Rasmussen and Martin Bjerring
*Department of Animal Health and Welfare,
Danish Institute of Agricultural Sciences*

April 2004
Deliverable D7

Correspondence:
Morten Dam Rasmussen
Research Centre Foulum,
DK-8830 Tjele
Denmark
E-mail: MortenD.Rasmussen@agrsci.dk

Abstract

The general conditions for hygienic milk production in the EU are currently defined in Commission Directive 89/362/EEC (1989). The directive states that the appearance of the milk shall be checked before milking and abnormal milk should be withheld from delivery. A definition of abnormal milk was proposed in deliverable D6 of the EU-project *Implications of the introduction of automatic milking on dairy farms* (QLK5 -2000-31006). D6 defines abnormal milk as foremilk differing in colour and homogeneity from that of normal milk. In this deliverable this definition is used to test the ability of different AMS models to detect and hence divert abnormal milk at time of milking.

Five different models of AMS were tested in six herds and sampled for 13 to 48 hours to find at least 10 cow milkings with abnormal milk and 50 cow milkings with normal milk. Due to the short sampling periods, the CMT-score of the foremilk was used to identify and support classification of abnormal and normal milk. Cows and quarters with a CMT-score >3 and no clots on a 0.1 mm filter were omitted from the calculations.

The current AMS models have systems to produce alarm lists of cows that should be checked for abnormalities of their milk, but these systems are not intended for automatic diversion of milk at present. This should be taken into account when evaluating the current systems. The sensitivity of the detection for the six herds varied from 13 to 50% when calculated for the actual milking, from 22 to 100% for the test days, and from 43 to 100% when calculated for the previous week. Specificities for the same time periods were found to be 87-100%, 85-100%, and 35-100%, respectively. At present, the sensitivities and specificities are generally too low for automatic diversion of abnormal milk, and it seems that most of the models could benefit from application of more sophisticated algorithms.

It is proposed that calculation of sensitivity should be based on sampling of at least 20 cow milkings with abnormal milk from a total of three herds. Sampling should be carried out for about 36 hours in each herd. At least 200 cow milkings with normal milk should be used to calculate the specificity. We propose that at least 16 out of 20 cow milkings with abnormal milk should be detected, which ensures that the sensitivity is 62.5% or better for a 95% confidence interval. For calculation of specificity, we propose that at least 198 cows should test normal out of 200 truly normal cow milkings, which will give a minimum specificity of 97.6% for a 95% confidence interval. Calculation of specificity at the quarter level improves the statistical power and lowers the confidence interval.

One model of AMS was tested for its ability to detect and separate milk coloured by blood. The model separated milk with ≥ 6 μmol haemoglobin per L, which is the level where a red tinge to the milk can be noted. Milk mixed with blood injected into the milk stream during a short period of time in the beginning of milking was not separated. We lack data on how blood is naturally expelled into milk and to what amount. We propose that testing of sensors should be based on milk samples mixed with blood to a concentration of 0, 3, 6, and 120 μmol haemoglobin per L of milk. Milk with 120 μmol haemoglobin should be detected at the quarter level. We propose that cow composite milk with >6 μmol haemoglobin per L should be separated because at this level milk will have a red tinge.

Table of contents

1	INTRODUCTION.....	1
1.1	USE OF CONDUCTIVITY TO IDENTIFY COWS WITH INDUCED CLINICAL MASTITIS.....	1
1.2	USE OF CONDUCTIVITY TO IDENTIFY NATURALLY INFECTED COWS	2
1.3	OBJECTIVE.....	3
2	DEFINITION OF ABNORMAL MILK.....	4
2.1	DEFINITION OF COLOURED MILK	4
2.2	MILK DIFFERING IN HOMOGENEITY	4
3	TEST PROTOCOL FOR SYSTEMS FOR DETECTION OF ABNORMAL MILK.....	6
3.1	TEST OF DETECTION SYSTEMS FOR COLOURED MILK.....	6
3.2	TEST OF DETECTION SYSTEMS FOR CHANGES IN HOMOGENEITY. DETECTION AT THE QUARTER LEVEL.	6
3.3	TEST OF DETECTION SYSTEMS FOR CHANGES IN HOMOGENEITY. DETECTION AT THE COW LEVEL.	7
3.4	GUIDELINES FOR INTERPRETATION	7
4	SELECTION OF HERDS.....	8
4.1	DATA COLLECTION, CLINICAL MASTITIS	8
4.2	DATA COLLECTION, COLOURED MILK	8
4.3	STATISTICAL METHODS.....	9
5	RESULTS.....	10
5.1	NUMBER OF NORMAL AND ABNORMAL SAMPLES IN THE SIX HERDS	10
5.2	ALARM LISTS	10
5.3	CALCULATION OF SENSITIVITY AND SPECIFICITY	11
5.4	CALCULATION OF SENSITIVITY AND SPECIFICITY AT THE QUARTER LEVEL	12
5.5	DETECTION OF MILK COLOURED WITH BLOOD	12
6	DISCUSSION.....	14
6.1	AUTOMATIC DIVERSION OR ALARM LIST	14
6.2	CLASSIFICATION OF QUARTERS AND COWS	14
6.3	NUMBER OF HERDS PER MODEL AND SAMPLING PERIOD.....	15
6.4	CONFIDENCE INTERVALS FOR SENSITIVITY AND SPECIFICITY	15
6.5	IMPROVEMENT OF SENSITIVITY AND SPECIFICITY	16
6.6	DIVERSION OF COLOURED MILK.....	16
7	CONCLUSION.....	18
	REFERENCES.....	19

1 Introduction

The general conditions for hygienic milk production in the EU are defined in Commission Directive 89/362/EEC (1989) and Chapter III-4 reads:

Before the milking of the individual cow, the milker must inspect the appearance of the milk. If any physical abnormality is detected, milk from the cow must be withheld from delivery.

The proposal for the coming Hygiene Directive part B. *Hygiene during milking, collection and transport* (Off. J. EU 2004/C 48 E/23) reads:

1. Milking must be carried out hygienically, ensuring in particular:

(a) that, before milking starts, the teats, udder and adjacent parts are clean;

(b) that milk from each animal is checked for organoleptic or physico-chemical abnormalities by the milker or a method achieving similar results and that milk presenting such abnormalities is not used for human consumption;

(c) that milk from animals showing clinical signs of udder disease is not used for human consumption otherwise than in accordance with the instructions of a veterinarian;

(d) the identification of animals undergoing medical treatment likely to transfer residues to the milk, and that milk obtained from such animals before the end of the prescribed withdrawal period is not used for human consumption; and

(e) that teat dips or sprays are used only if the competent authority has approved them and in a manner that does not produce unacceptable residue levels in the milk.

Fulfilment of the old directive was a problem with automatic milking systems because normally a human is not present and visual inspection of foremilk is not performed. The proposal for the new directive opens a possibility for detection of abnormal milk at any time during milking and by any means. It seems that the term used for defining abnormal milk is too broad to be useful. Technical solutions may replace visual inspection for detection of abnormal milk either before or during milking. No direct sorting of milk is done based on these systems and it is the decision of the farmer whether a cow's milk should be delivered or not. Presently, conductivity is mainly used to identify cows and quarters on an alarm list, which the farmer can then consult and use to supplement observations of cows in the barn.

1.1 Use of conductivity to identify cows with induced clinical mastitis

In a study with experimental infections, 90% of the clinical cases were identified by changes in the electrical conductivity of foremilk before or when clinical signs were first observed (Milner et al., 1996), Table 1. A 10% increase in the conductivity of foremilk from a moving average of the last four milkings and a relative change between quarters were used to identify affected quarters. Two cases were indicated at the following milking. One case of *Staphylococcus aureus* mastitis was not found. This method was able to find 30 out of 31 clinical cases (sensitivity 97%). In all, 20 cows were inoculated with *Streptococcus uberis* and 20 cows with *S. aureus*. Five cows showed no sign of infection. The mean somatic cell count (SCC) of infected quarters was about 2M cells/ml when conductivity values changed compared with 4M cells/ml for *S. aureus* and 12M cells/ml for *Str. uberis* when clinical signs appeared. These results showed that conductivity was a good indicator for development of clinical mastitis and that treatment of such cases could be initiated before clinical signs actually occurred.

Hillerton and Semmens (1999) induced *Str. uberis* mastitis in 24 cows. Antibiotic treatment of five of the cows was initiated when the quarter conductivity was increased by 10% compared

with the four previous comparative milkings or when the conductivity of that quarter changed from being the lowest to being the highest of the four quarters. This early treatment effectively cleared the infections from quarters and no clinical signs occurred. Antibiotic treatment after the appearance of clinical signs of mastitis took longer time and used more antibiotics to achieve a bacteriological cure. These two studies show that the use of conductivity has good potential to detect truly new clinical cases of mastitis.

Table 1. No. of milkings until bacteria were present in foremilk, SCC of foremilk increased by a factor 10, conductivity of foremilk increased by 10%, or clinical signs appeared in the milk (Milner et al., 1996).

Status	No. of cows	No. of milkings after experimental infection				Found by +10% increase in conductivity
		Bacteria present	SCC 10X	Conductivity + 10%	Clots in foremilk	
<i>Str. uberis</i> , clinical	14	2.3	3.1	3.5	5.2	93
subclinical	3	2.7	9.3	2 ^a	-	-
<i>S. aureus</i> , clinical	17	1.1	2.3	2.7	4.2	88
subclinical	1	1	3	3	-	-

a: only 1 out of 3 subclinical cases were indicated by an increase in conductivity.

1.2 Use of conductivity to identify naturally infected cows

Chronically infected quarters are characterised by a high and fluctuating SCC and a varying excretion of bacteria. The foremilk may occasionally or never show clinical signs. Conductivity will be raised in the milk of such quarters.

All clinical cases were detected as mastitic by having a 15% increase in conductivity relative to the lowest quarter at the actual milking and a 20% increase over the last two milkings (Maatje et al., 1992). Additionally, half of the quarters with a SCC >500,000 cells/ml were detected as mastitic, Table 2. If these criteria had been used to divert milk automatically, the milk from 25 clinical cases and 56 cases without clinical abnormalities would have been diverted. The sensitivity for diversion of abnormal milk was in this case 100% and the specificity 85%.

Table 2. Detection of quarters with mastitis from a 15% increase in conductivity relative to the lowest quarter at the actual milking and a 20% increase over the last two milkings (Maatje et al., 1992).

Status	No. of cases	Detected	% detected
Clinical mastitis	25	25	100
Subclinical mastitis (bacteria present and SCC >500,000/ml)	26	14	53
Bacteria present and SCC <500,000/ml	106	12	11
No bacteria and SCC >500,000/ml	42	22	52
Healthy cows	200	8	4

Nielen et al. (1995) used models based on the highest conductivity values and variation within the first, middle, and last minute of milking. Quarters where milkers observed abnormal milk were defined as having clinical mastitis. Quarters without bacteria present and SCC <200,000 cells/ml were used as negative controls. Data were collected for about one year in a 60-cow experimental herd. The best model had a sensitivity of 84% and a specificity of 97%. The authors expected higher predictive values from the models when more clinical signs occurred.

Maatje et al. (1997) found a sensitivity of 90% for detection of clinical mastitis and 76% for subclinical mastitis and a specificity of 98% for healthy quarters (no bacteria present and SCC <500,000 cells/ml) by the use of time series analysis and Kalman-filters. The authors concluded that although the specificity seems high, it may still be too low for practical application. Even better figures were found by applying fuzzy logic (deMol and Woldt, 2001) to classify clinical mastitis. All clinical cases of mastitis (without measurement errors) were classified correctly. The number of false positives was reduced from 1266 to 64 alerts from 25 cows without signs of mastitis in 29033 milkings. Consequently, the sensitivity was 100% and the specificity 99.2% and fuzzy logic models seem to be very useful in improving conductivity measurements for the detection of clinical mastitis. These principles were applied to datasets from four farms with 212 cases of clinical mastitis (de Mol et al., 2001). The manufacturer's model had a sensitivity of 33% and a specificity of 98.6%. Operating with a 99.9% confidence interval for alerts improved the sensitivity to 62% and the specificity to 99.3% for the fuzzy logic models. Lower confidence intervals improved sensitivity further but decreased specificity.

Trilk (2002) tested one model of AMS for the ability to issue an alert on quarters with visibly changed foremilk. The sensitivity was found to be 70.3% of 219 quarters with changed foremilk. The specificity was 93.3% and 54% of the warnings concerned quarters without changes in the foremilk. The author concluded that the identification rate of changes in foremilk and clinical mastitis was insufficient for practical use. However, there seems to be a possibility of using colour scanning of milk as an aid in the differentiation between normal and abnormal appearance of milk (Ouweltjes & Hogeveen, 2001; Espada & Vijverberg, 2002). The sensitivity and specificity for detection of clinical mastitis are still not high enough. Presently, one model of AMS is using colour measurement to automatically divert milk but alerts mainly concern milk coloured by blood.

1.3 Objective

The objective of this paper is to determine state-of-the-art ability for AMS to detect and separate milk based on quality. Quality is defined as milk being normal or abnormal according to the definitions in chapter 2.

2 Definition of abnormal milk

In order for AMS companies to develop sensors to detect abnormal milk, a precise definition of abnormal or unacceptable milk was needed. Such a definition was suggested in deliverable D6 (Rasmussen, 2003) where milk was classified into four categories:

- *Normal milk*: Milk suitable for human consumption.
- *Abnormal milk*: Milk which differs from normal milk in respect of colour or homogeneity.
- *Contaminated milk*: Milk which, prior to the milking of the animal, is known to be unfit for human consumption following treatment of the animal with antibiotics or other veterinary products which require that the milk must be withheld from sale for such use.
- *Undesirable milk*: Milk which, prior to the milking of the animal, is known or expected to be unsuitable for human consumption, e.g. colostrum, high somatic cell count.

A better term for contaminated milk is “milk with a withholding period” since “contaminated milk” describes something happening unintentionally, and consequently does not cover the situation with antibiotic treatment of cows. Foremilk is suggested as the reference for testing in order to avoid double standards for conventional and automatic milking.

2.1 Definition of coloured milk

Following consensus at the workshop held in Foulum 2002, milk that has changed in colour because of the content of red blood cells is regarded as abnormal milk (Rasmussen, 2003). The frequency of visible blood in the milk is low. Against a reference of white milk, test panels of consumers and professionals can detect samples with about 0.1% of blood (about 6 μmol haemoglobin in 1 L milk), but even 1% of blood does not show clearly in a black strip cup, which is the conventional reference method when foremilk takes place in a conventional parlour. About 2% of blood (about 120 μmol haemoglobin in 1 L milk) in milk is detectable in a black strip cup and should consequently be detected by conventional as well as automatic systems. The amount of haemoglobin in blood, i.e. the red blood cell count, has to be considered when determining the reference because it is the haemoglobin that colours the milk.

2.2 Milk differing in homogeneity

If clots appear in the foremilk, the milk from that quarter and cow is abnormal. The reference method is proposed as the appearance of clearly visible clots on a filter with a pore size of 0.1 mm (Rasmussen, 2003). It is important that the reference method for classification of the milk is based on science, is applicable, repeatable, and objective. The Foulum workshop was in favour of defining abnormal milk caused by clinical mastitis on the homogeneity of the milk and not on the colour (besides red) since the colour of the milk changes with breed, stage of lactation, feedstuffs, etc. Rasmussen (2003) proposed that the reference method should be based on filtration of the milk through a filter with a pore size of 0.1 mm. Milk where clots were clearly visible in such a filter was then defined as being abnormal. Incidences of watery and yellowish milk may or may not be detected by this method. The cell count or other supportive measurements should not be included in the reference, because it has to apply to all kinds of manual and mechanical milking methods.

The current standard is to inspect the appearance of foremilk. Clinical mastitis develops in response to invasion of pathogens through the teat canal. Clinical signs are not always seen in the foremilk but may appear later in the milk flow, and the Foulum workshop proposed to include all milk from the quarter in the definition, i. e. if clots are detected at any stage of the milking the milk is abnormal. It is not practical to base the daily judgement on milk fractions other than the foremilk. However, the frequency of cows with no clots in the foremilk but clots appearing late into the milking is expected to be very low, and clots are likely to appear in the foremilk of subsequent milkings.

3 Test protocol for systems for detection of abnormal milk

This protocol describes test and reference methods for detection of abnormal milk. The automatic detection systems should be tested for sensitivity and specificity in relation to the colour and homogeneity of the milk.

3.1 Test of detection systems for coloured milk

Four samples of milk from each of ten cows are prepared to contain 0, 3, 6, and 120 μmol (about 0, 50, 100, and 2000 mg) haemoglobin per L of milk. Samples should be homogeneously mixed. Haemoglobin is measured as moles of iron atoms with a molecular weight of about 16 kDa. Sensitivity of the detection system is calculated from the samples with 120 μmol haemoglobin/L as the number of positively detected samples divided by 10 samples and multiplied by 100%. The remaining samples with 0, 3 and 6 μmol haemoglobin are used to calculate the specificity as the number of negatively detected samples divided by 30 samples and multiplied by 100 to give a percentage.

Note: If applicable, standards giving the same red colour (e.g. painted plates or sticks) as 0, 3, 6, and 120 μmol haemoglobin per L milk can be used as test method. Such methods will need validation before use.

Note: The 120 $\mu\text{mol/L}$ threshold is scored as visually red with a sensitivity of >80% by a representative group of people when scoring the coloured milk on a black surface. The 3 and 6 $\mu\text{mol/L}$ thresholds are scored as visually white with a sensitivity of >80% by a representative group of people when compared with milk without haemoglobin.

3.2 Test of detection systems for changes in homogeneity. Detection at the quarter level.

More than 100 foremilk samples (each about 10 ml) of all milking quarters of at least 10 cows are taken immediately before the individual automatic milking. The foremilk is sampled on to a filter with the pore size of 0.1×0.1 mm. Milk from quarters with visible clots (more than 2 clots >2 mm) on the filter and CMT-score >3 is abnormal. At least 10 samples shall test abnormal. Milk from quarters where foremilk samples from two consecutive milkings do not show clots on the filter and have a CMT-score <4 are denoted normal samples. Milk from quarters with clots on the filter is used to calculate the sensitivity of the detection system as test positives divided by number of quarters with abnormal milk and multiplied by 100. Milk from quarters denoted normal are used to calculate the specificity as number of test negative quarters divided by number of normal quarters and multiplied by 100. Quarters with CMT scores 4 and 5 but without visible changes of the milk, quarters with watery milk, and quarters with small flakes on the filter were discarded from the analysis.

The expected cell counts of the CMT-scores were: 1) <150,000, 2) 150-300,000, 3) 300-800,000, 4) >800,000, and 5) > 3×10^6 cells/ml. CMT-score was not intended to be included in the classification of normal and abnormal milk but was used here as supportive information due to the short sampling periods. Short sampling periods make it more difficult to safely categorise quarters and cows correctly. For further discussion see paragraphs 6.2 and 6.3.

3.3 Test of detection systems for changes in homogeneity. Detection at the cow level.

For detection systems based on measurements in composite milk, foremilk samples are taken from at least 100 cow milkings of >40 cows. Milkings where foremilk from one or more quarters have visible clots on the filter and CMT-score >3 are denoted abnormal. Milk where foremilk from all quarters for at least two consecutive milkings does not have visible clots on the filter and a CMT-score <4 is denoted normal. Cows with abnormal milk are used to calculate the sensitivity as test positives divided by number of cows with abnormal milk and multiplied by 100. Cows with normal milk are used to calculate the specificity as test negatives divided by number of cows with normal milk and multiplied by 100. Cows with a CMT score of 4 or 5 of one or more quarters but without visible changes of the milk, cows with watery milk, and cows where one or more quarters had small flakes but no clots on the filter were discarded from the analysis.

Note: Finding of clots in the foremilk on the filter means that the quarter and cow suffer from clinical mastitis.

3.4 Guidelines for interpretation

The sensitivity of devices for detecting abnormal milk should be >80%, which means that 8 out of 10 cow milkings with abnormal milk should be diverted automatically. Note that the confidence interval for an expected sensitivity of 80% on 10 cow milkings is 55-105%, and for 20 cow milkings it is 63-98%. In a preliminary study, Rasmussen et al. (2002b) found that the sensitivity for visual detection of abnormal milk was 60% when the filter method was reference for the true status.

The specificity of devices for detecting abnormal milk should be >99%, which means that 1 out of 100 cow milkings with normal milk should be automatically diverted. Note that the confidence interval for an expected specificity of 99% on 100 cow milkings is 97-101% and 98-100% for 200 cow milkings with normal milk.

4 Selection of herds

The Danish distributors of AMS were contacted to nominate suitable herds with at least 100 cows and a technically well functioning AMS for inclusion in a trial. Technicians from the distributors visited the herds before the days of testing and were present during the test hours. Herds with bulk milk SCCs of 300,000 cells/ml or more were selected since such herds can be expected to have cows with subclinical as well as clinical mastitis. For herds with more than 100 cows, we expected to be able to find at least 10 cow milkings with abnormal milk from at least five different cows. Data were collected from the five AMS models present in Denmark, i.e. 1) DeLaval VMS, 2) Fullwood Merlin, 3) Gascoigne Melotte, 4) Insentec Galaxy, and 5) Lely Astronaut. One herd was selected for each of the AMS models 1-4 and two herds for model 5. The models remain anonymous in the data. Only one model of AMS was equipped with a colour sensor to divert milk automatically when blood was detected. The test for diversion based on colour was carried out in one herd.

4.1 Data collection, clinical mastitis

The six selected herds were sampled over a range of hours (13-48 hours). At least 50 cows with normal milk were sampled twice in the herds sampled for the shortest times. Cows were foremilked in the milking box just before the automatic milking. Normally, the interval from cow identification to start of movement of the robot arm is very short, but for some AMS models, it was possible to add a time lag to allow enough sampling time. For other models, one person was allocated to handling the cows of one box each so that sampling could start as soon as the cow entered the box. Foremilking was done into a four-chambered strip cup with 0.1 mm filters mounted at the outlet. A CMT-scoring plate was collecting the foremilk from each quarter. Visual scoring was done during foremilked. A small amount of water was run through the filters to remove foam before the visual inspection of the filters. CMT-scoring was also done immediately after foremilked. Cows and quarters were classified as normal, abnormal, dropped, or discarded:

- Normal: Two consecutive milkings without clots on the filter, no visual abnormality, and CMT-score ≤ 3 .
- Abnormal: Milkings with clots on the filter and a CMT-score > 3 .
- Dropped: First milking of cows with normal milk.
- Discarded: CMT-score > 3 or visually changed in colour but no clots on the filter.

Only milkings denoted normal and abnormal are used to calculate sensitivity and specificity.

4.2 Data collection, coloured milk

Fresh blood was taken from one cow into a bottle with heparin to avoid clotting. Haemoglobin (measured as Fe-atoms) content of the blood was 8.0 mmol/L (Hb 201+, HemoCue AB, Ängelholm, Sweden). Composite milk from one cow was used to mix 4x1 L of milk homogeneously with 0, 3, 6, and 120 μmol (about 0, 50, 100, and 2000 mg) haemoglobin, respectively. This milk was sucked into the AMS through an artificial udder. The flow was adjusted to about 0.5 L/min. The AMS had difficulties with the attachment due to the “shape” of the artificial udder and manual operation was not possible. Consequently, the test protocol was modified to injection of blood into an additional 1-m looped hose of one of the “short” milk hoses before milk entered the sensor. A milk separator was inserted immediately after the sensor to be able to measure the milk yield and avoid the added blood to be mixed with the milk for delivery. Blood was injected at a flow of about 0.5, 1.0 and 10 ml/min in order to reach the three planned concentrations of haemoglobin in the milk. This was done on 11 cows. Additionally, 1, 2, and 20 ml of blood was injected from 15 to 30 sec

after start of milking of three cows, again to reach the planned levels of haemoglobin for the quarter milk but during a short time period to simulate milking of a injured teat.

4.3 Statistical methods

No statistical analysis was applied to the results. Sensitivity was calculated as true test positives divided by true positives (= abnormal milk) and specificity as true test negatives divided by true negatives. The 95% confidence intervals were calculated as

$$p \pm 1.96 \times \sqrt{\frac{p \times (1 - p)}{N}}$$

where p is the probability (either sensitivity or specificity) and N is the number of test samples (either abnormal or normal).

5 Results

5.1 Number of normal and abnormal samples in the six herds

Sampling in the six herds resulted in the collection of foremilk scorings of 169 to 623 cow milkings, Table 3. A large percentage of the samples were first samples (dropped), especially in the herds with a short sampling time. About 5 to 15% of the cow milkings were discarded because CMT-scores were 4 or 5 (main reason to discard samples) or they were visually changed in colour but with no clots on the filter. The number of cow milkings with normal milk was 47 or more per herd, and the number of cow milkings with abnormal milk was from 4 to 18 per herd. In the test, 65 quarters showed clots on the filter and a CMT-score >3. Totally, we found clots on the filters from the foremilk of 83 quarters of which 58 cases were found visually in the strip cup. This gives a sensitivity of 70% for visual detection of abnormal milk when using the filter as true status.

Table 3. Number of cows and milkings in the tested herds.

Model	Herd	AMU	Cows	Hours of sampling	No. of cow milkings					Quarters	
					Total	Drop	Discard	Normal	Abn.	Normal	Abn.
1	1	2	79	48	350	89	46	206	9	936	9
2	2	3	145	13	222	113	35	56	18	256	22
3	3	4	116	13	178	90	21	61	6	243	6
4	4	2	100	14	192	104	26	47	15	227	15
5	5	2	105	16	169	102	7	54	6	223	6
5	6	3	184	36	623	184	69	366	4	1526	7

5.2 Alarm lists

Clinical mastitis may cause the milk yield to drop significantly, but cows with other health problems may also drop in milk yield or milk flow. The first three AMS models tested had produced alarm lists based on both milk conductivity and measurements of milk yield and/or milking speed (Table 4).

Table 4. Number of cow milkings during the test day(s) of each herd, the number of discarded and abnormal milkings, and divided into being on the alarm list or not at the actual milking. Alarm systems were based on milk conductivity, milk yield, and/or milking speed.

Herd	No alarm				Alarm list			
	Total	Discard	Normal	Abnorm	Total	Discard	Normal	Abnorm
1	241	36	199	6	20	10	7	3
2	101	33	56	12	8	2	0	6
3	73	14	54	5	11	7	3	1

All herds had an alarm list based on conductivity (Table 5). A varying number of samples classified as discarded came from cows appearing on the alarm lists. Based on conductivity, one to five cow milkings with abnormal milk matched the alarm directly and 2 to 13 did not.

Table 5. Number of cow milkings during the test day(s) of each herd and the number of discarded and abnormal milkings divided into being on the alarm list or not at the actual milking. Alarm systems were based on milk conductivity.

Herd	No alarm				Alarm list			
	Total	Discard	Normal	Abnorm	Total	Discard	Normal	Abnorm
1	250	41	203	6	11	5	3	3
2	102	33	56	13	7	2	0	5
3	74	14	55	5	10	7	2	1
4	73	19	41	13	15	7	6	2
5	54	4	47	3	13	3	7	3
6	406	43	361	2	33	26	5	2

5.3 Calculation of sensitivity and specificity

Table 6 presents the sensitivities and specificities calculated for the actual milkings, for the test days, and for the previous week (including the test day). Sensitivities were generally low for the actual milking and increased when looking at a longer time span. Specificities were generally high at the actual milking and dropped when looking at a full week. Herd 5 had the lowest specificity for the actual milking, but it turned out that sensors were not calibrated sufficiently. Consequently, the specificity was very low when looking at the alarm list for a week. Considerably more cows appeared on the alarm lists when milk yield or milk flow was included as deviation criterion (Table 6) than when only using conductivity deviations (Table 7). It could be that the relatively high numbers of abnormal cow milkings found in herd 2 were a result of the relatively low sensitivity in this herd. However, five of the 14 cows with abnormal milk were separated manually. For herd 4, the sum of sensitivity and specificity was 100% indicating that correct classification of abnormal quarters was purely random. From plots of herd 5, it seemed as if sensors used in this AMS were drifting, which may explain the poor specificity compared with herd 6.

Table 6. Sensitivity (SE) and specificity (SP) for appointing abnormal and normal milk from cows during an actual milking, the day(s) of testing, or the previous week. Alarm systems were based on milk conductivity, milk yield, and/or milking speed.

Herd	Abnormal		Actual milking		Day(s) of test		Previous week	
	Cows	Milkings	SE	SP	SE	SP	SE	SP
1	5	9	33	97	100	77	100	65
2	14	18	33	100	36	96	50*	87*
3	6	6	17	95	33	71	100	12

* Data only available for three days.

Table 7. Sensitivity (SE) and specificity (SP) for appointing abnormal and normal milk from cows during an actual milking, the day(s) of testing, or the previous week. Alarm systems were based on milk conductivity.

Herd	Abnormal		Actual milking		Day(s) of test		Previous week	
	Cows	Milkings	SE	SP	SE	SP	SE	SP
1	5	9	33	99	100	89	100	85
2	14	18	28	100	36	100	43*	100*
3	6	6	17	96	33	92	100	55
4	9	15	13	87	22	85	67	62
5	5	6	50	87	60	87	100	35
6	2	4	50	99	100	99	100	83

* Data only available for three days.

5.4 Calculation of sensitivity and specificity at the quarter level

All AMS models measured quarter conductivity, and calculation of sensitivity and specificity could be done at the quarter level, Table 8. Sensitivities were generally similar to the cow level. For herd 4, one cow was classified correctly at the actual milking but the wrong quarter was indicated, resulting in 0% sensitivity. Specificities were higher than found at the cow level.

Table 8. Sensitivity (SE) and specificity (SP) for appointing abnormal and normal milk from quarters during an actual milking, the day(s) of testing, or the previous week. Alarm systems were based on milk conductivity.

Herd	Abnormal		Actual milking		Days of test		Previous week	
	Quarters	Milkings	SE	SP	SE	SP	SE	SP
1	5	9	33	99.7	80	96.9	80	95.8
2	16	22	18	100	31.3	100	37.5	100
3	6	6	17	97.5	33.3	94.4	83.3	76.1
4	15	15	0	96.0	16.7	94.3	58.3	88.0
5	6	6	50	96.4	60.0	92.3	100	72.1
6	5	7	71	99.5	100	97.9	100	93.5

5.5 Detection of milk coloured with blood

Milk yields of the sampled quarters were from 1.0 to 5.8 L and blood concentrations ranged from 0 to 180 μmol haemoglobin per L milk. Nine samples with <6 μmol haemoglobin per L milk were not separated. Out of the 8 samples with ≥ 6.0 μmol haemoglobin per L milk, 6 were separated and 2 were not (Figure 1). These two samples had blood injected during the short period of time just after attachment of the quarter.

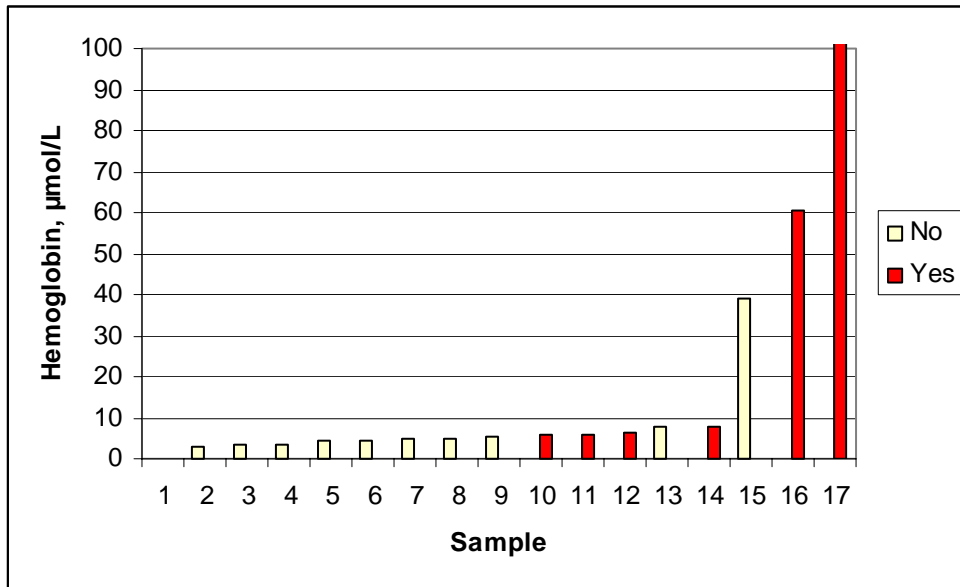


Figure 1. Detection of milk coloured with blood in the range of 0 -180 µmol haemoglobin per L milk. Samples marked with white (open) bars were not separated and samples with coloured bars were separated automatically.

6 Discussion

6.1 Automatic diversion or alarm list

Sensitivities of detection of abnormal milk were generally low when calculated for actual milkings and certainly too low for automatic diversion of abnormal milk. Herd 1 had a sensitivity at the cow level of 33% for the actual milking, but the sensitivity increased to 100% when calculated for the two sampling days, which is highly suitable for an alarm list (sometimes denoted attention or alert list). In herds 2 and 3, detections were low in sensitivity on the test day (24 hours). Herd 3 improved its sensitivity to 100% when using a one-week window. However, this high sensitivity was possibly due to a very poor specificity (55%) where about half of the cows in the herd were alerted within a week. Such a low specificity will not indicate truly abnormal cows and reduces farmers' confidence in the alert system. The model used for Herd 2 appeared restricted in its calculations of alerts and consequently achieved a very high specificity even using a three-day window (100%). The numbers calculated for Herds 1 and 6 give a fair balance between alerting to all abnormal cows and keeping a reasonable specificity for an alarm list. Herd 2 appears to miss too many cows with abnormal milk but had better performance in identifying truly normal milkings (none abnormal but two discarded). The low specificity in Herd 5 indicates a need to identify malfunctioning of sensors.

A drop in milk yield or flow may indicate that the cow is sick. However, the daily variations are large in these parameters and uncritical use of these as indicators of udder problems appears to contribute to a very low specificity. In Herd number 3, the alarm system was poor over the whole week with a specificity of 12%. Herd 1 was better in the same time span (specificity 65%) but still achieved too low a specificity for practical use if the udders of all cows on the alarm list need to be checked. The alarm lists for udder health should focus on this matter and a high sensitivity of alarm indication of udder health problems should not be achieved at the cost of poor specificity.

Automatic diversion of milk is not the primary objective for an alarm list based on conductivity where the indication of cows with potential mastitis problems is more important.

6.2 Classification of quarters and cows

There are several ways of classifying the udder health of the dairy cow, usually into clinical mastitis, subclinical mastitis, or healthy. The definition of abnormal milk in this context is milk that differs in homogeneity and colour from that of normal milk. This means that only quarters and cows with clinical mastitis are rated as abnormal and that normal milk may originate from cows with subclinical mastitis. Changes in homogeneity may not show at every milking and some quarters have teat canal infections that cause clots in the foremilk but do not cause an inflammatory response in the quarter. Consequently, classification of quarters and cows into normal and abnormal milk can be difficult if based on single milkings. We could choose to base the classification on multiple milkings or use other inflammatory parameters to support the decision. During the test, classification of quarters with abnormal milk was supported by CMT-score of the foremilk. Likewise, the prerequisites for classifying quarters with normal milk were no visual appearance of any clinical sign and a low CMT-score at two consecutive milkings, which increases the probability of being truly normal (Rasmussen et al., 2002a). This procedure, however, left a group of cows and quarters with high CMT-scores, watery, or yellowish milk. This was a relatively large group compared with the number of cows with abnormal milk. Many of the discarded cows were on the alarm list

but not included in the calculation of sensitivity and specificity. Consequently, the sensitivities and specificities are probably overestimated in our calculations.

6.3 Number of herds per model and sampling period

Only one herd was used to test four of the models. This is an advantage in terms of labour input to perform the test. Different herds have different levels of clinical and subclinical mastitis. We selected herds for the tests that had relatively high bulk milk SCC in order to increase the probability of finding cows with clinical mastitis (= abnormal milk) in the herds. For most of the herds, we were successful. However, a high bulk milk SCC could also reflect that the herd had many cows with a subclinical infection and a high and fluctuating SCC. Such cows would also have a higher conductivity in the milk of the infected quarters and were more likely to be included in the alarm list. We discarded cows with a high CMT-score without visual clots in the foremilk from the calculations of sensitivity and specificity. This was to be sure that 1) only cows with clinical mastitis were counted as abnormal; 2) only healthy cows were used for calculation of the specificity; and 3) a management decision of keeping chronically infected cows in the herd would not influence the calculations. The second reason could be modified if the sampling period was longer and more samples were used to classify the cows and quarters into normal and abnormal.

Different organisms cause clinical and subclinical mastitis within and between herds. Quarters infected with coliforms and environmental streptococci are more likely to develop clinical than subclinical mastitis whereas quarters infected with agents like *S. aureus* may stay subclinical for long periods of time and only occasionally show clinical signs in the milk. Such differences between agents will appear as differing patterns of alerts on the alarm lists when these are mainly based on conductivity. Changes in the settings of alarm thresholds could adjust the numbers of false positives on the alarm list within a herd.

In order to have enough cases of abnormal milk with sufficiently different causes, it is advisable to sample at least three herds. The sampling period of each herd should last as long as it takes to sample all cows twice, which means sampling for at least 24 hours. The protocol we have used for the present tests tended to use too few cow milkings with abnormal milk as well as normal milk. When only sampling for 13-14 hours, it will mainly be the high yielding cows that are sampled twice and fewer, if any, of the low yielding cows. This may cause a bias especially if different algorithms and thresholds are used for different stages of lactation. Some cows will have a milking interval of 18 to 24 hours (cows not coming by themselves) and these cows, besides being lazy, could have a health problem. Sampling for 36 hours will then include all cows in the herd at least twice. The difference in specificities for herds 5 and 6 highlights the need for sampling of more than one herd per tested model. If selecting three herds, there is a higher probability of having enough abnormal milkings as well as normal milkings, and calculations of sensitivity and specificity will be based on measurements from several sensors. In conclusion, sampling should be done in three herds and for 36 hours each.

6.4 Confidence intervals for sensitivity and specificity

Some of the herds only had a few cases of abnormal milk. The confidence interval will be relatively large for small numbers and in the herd with six cow milkings with abnormal milk, only two need to be found in order not to discard the hypothesis of a sensitivity of 80% with a 95% confidence. However, only one cow was detected at the actual milking in Herd 3 and for this herd the sensitivity was <80%. Likewise, 10 cows should have been found in Herd 2 in order not to reject the hypothesis of a sensitivity of 80% whereas only 5 cow milkings were flagged at the actual milking. Finding 12 out of 20 cow milkings with abnormal milk will give

a sensitivity of 60%, which, for a 95% confidence interval, will then include a sensitivity of 80%. If the 80% sensitivity is a minimum requirement, then milk from at least 19 out of 20 cow milkings with abnormal milk should be diverted. If this is the case, then it will be required that the sensors measure the property of abnormal milk rather directly. This does not leave much room for development of sensors measuring correlated properties of abnormal milk. We propose that at least 16 out of 20 cow milkings with abnormal milk should be detected, which ensures that the sensitivity is 62.5% or better for a 95% confidence interval. For an 80% confidence interval, the minimum guaranteed sensitivity would then be 68.5%.

Hillerton (2000) found a sensitivity for detection of clinical mastitis of 80% for conventional milking. The reference in that study was not the filter method. The sensitivity for appointing abnormal milk through visual appearance of the foremilk (as for conventional milking) was 70% in our dataset when using the filter method as the true status. This sensitivity is lower than the requirements set in paragraph 3.4 but fits with the minimum guaranteed sensitivity for an 80% confidence interval.

Another way to increase sensitivity is to decrease specificity. However, a low specificity for automatic sorting will directly influence the economy of milk production since normal saleable milk will be discarded. Farmers will not accept low specificities for automatic sorting and to a certain extent not for an alarm list either since the proportion of normal to abnormal cow milkings then increases. We propose that the specificity for automatic sorting should be >99%. To give that guarantee with a 95% probability, all 50 out of 50 normal cow milkings should test normal. Testing 99 as normal out of 100 truly normal cow milkings will only assure that the specificity is >97%, and testing 199 normal out of 200 truly normal cow milkings will give a minimum specificity of 98.5%. Calculation of specificity on the quarter level improves the statistical power. Appointing 995 quarters as normal out of 1000 truly normal quarter milkings will, with a 95% confidence interval, guarantee that the specificity is >99%. We propose that more than 200 cow milkings should test normal in order to calculate the specificity safely at the cow level, which will also improve the precision of specificity calculations on a quarter basis and this is actually a better goal. Calculations at the quarter level should not be confused with the fact that diversion of abnormal milk still is at the cow level.

6.5 Improvement of sensitivity and specificity

It is quite clear from the relatively low sensitivities that the current systems cannot be used for automatic sorting of milk. Alarm systems should focus more on the definition of abnormal milk if a higher sensitivity should be achieved without lowering the specificity. Automatic sorting based on a low specificity will discard a lot of milk, and relatively much for herds with a low prevalence of clinical mastitis (Rasmussen, 2003). Farmers will not accept low specificities (Ouweltjes, 2004). Fuzzy logics can be used to improve sensitivity and specificity of systems using conductivity as the main source of information for mastitis detection (de Mol and Woldt, 2001; de Mol et al., 2001). Some of the systems may obviously benefit from adopting and implementing such calculation models.

6.6 Diversion of coloured milk

Milk may change in colour due to many different reasons but only colouring by blood will be discussed since the workshop in Foulum was in favour of not considering the colour affected by mastitis. Colostrum differs from normal milk in colour as well but the workshop defined the first three days after calving as the colostrum period and colour changes from that time onward are minor. The colour sensor tested here worked precisely when measuring a

continuous flow of milk mixed with blood. Quarter milk with up to 5.5 μmol haemoglobin per L milk was not separated and milk mixed with ≥ 6 μmol haemoglobin per L had a visible red tinge and was separated. However, two exceptions occurred when the amount of blood was injected in the beginning of the milking during a short time span. These samples contained 7.6 and 39.0 μmol haemoglobin per L milk when averaged for the full milking and would have been separated if the blood had been injected continuously during the entire milking of these quarters. These findings raise the question if coloured milk should be separated based on single high concentrations of blood or on the endpoint result after mixing with milk from the rest of the quarter and with milk from the other quarters. We think that composite milk with >6 μmol haemoglobin per L should be separated because at this level, milk will have a red tinge, which could be noted by an average producer or consumer. When milking in a conventional system with dark hoses and no recorder jar such redness will not be noted.

The test protocol was set up to test quarters with a continuous flow of coloured milk but does not simulate the situation where blood is coming in as squirts within short time periods. Presently, we do not have the data to show how blood is mixed into milk whenever the teat has an external wound, internal wound, or a punctured blood vessel. Foremilking in conventional systems will be able to detect external wounds of the teat and note if the first squirts contain >120 μmol haemoglobin per L milk. A small, leaking blood vessel will not be detected in a conventional system but probably be noted by an automated system as the one tested here. The occurrence of visible blood in milk is rare and we believe that systems being able to detect blood in milk according to the proposed protocol will also be able to detect most of the naturally occurring cases of blood in milk. Clever use of the software for detection of coloured milk could probably account for different natural ways of mixing blood into milk.

7 Conclusion

The current AMS models have systems to produce alarm lists of cows that should be checked for abnormalities of their milk, but at present the systems are not intended for automatic diversion of milk. At present, the sensitivities and specificities are generally too low for automatic diversion of abnormal milk and most of the models could benefit from application of more sophisticated algorithms or from the use of sensors that more directly measure abnormal milk. It is proposed that calculation of sensitivity should be based on sampling of at least 20 cow milkings with abnormal milk in three herds. Sampling should be carried out for about 36 hours. At least 200 cow milkings with normal milk should be used to calculate the specificity. We propose that cow composite milk with $>6 \mu\text{mol}$ haemoglobin per L should be separated because at this level milk will have a red tinge.

References

- Espada, E. and H. Vijverberg. 2002. Milk colour analysis as tool for the detection of abnormal milk. In: Proc. First North American Conference on Robotic Milking, March 20-22, Toronto, Canada, IV-28-IV-38.
- Hillerton, J.E. 2000. Detecting mastitis cow-side. *Natl. Mastitis Council Ann. Meet.* 39:48-53.
- Hillerton, J.E. and J.E. Semmens. 1999. Comparison of treatment of mastitis by oxytocin or antibiotics following detection according to changes in milk electrical conductivity prior to visible signs. *J. Dairy Sci.* 82:93-98.
- Maatje, K., P.J.M. Huijsmans, W. Rossing and P.H. Hogewerf. 1992. The efficacy of in-line measurement of quarter milk electrical conductivity, milk yield and milk temperature for detection of clinical and subclinical mastitis. *Livestock Prod. Sci.* 30:239-249.
- Maatje, K., R.M. de Mol and W. Rossing. 1997. Cow status monitoring (health and oestrus) using detection sensors. *Computers and Electronics in Agriculture.* 16:245-254.
- Milner, P., K.L. Page, A.W. Walton and J.E. Hillerton. 1996. Detection of clinical mastitis by changes in electrical conductivity of foremilk before visible changes in milk. *J. Dairy Sci.* 79:83-86.
- Mol, R.M. de, W. Ouweltjes, G.H. Kroeze and M.M.W.B. Hendriks. 2001. Detection of estrus and mastitis: field performance of a model. *Applied Engineering in Agriculture.* 17:399-407.
- Mol, R.M. de and W.E. Woldt. 2001. Application of fuzzy logic in automated cow status monitoring. *J. Dairy Sci.* 84:400-410.
- Nielen, M., Y.H. Schukken, A. Brand and S. Haring. 1995. Comparison of analysis techniques for on-line detection of clinical mastitis. *J. Dairy Sci.* 78:1050-1061.
- Ouweltjes, W. 2004. Demands and opportunities for operational management support. Operational management on farms with automatic milking systems. EU project "Implications of the introduction of automatic milking on dairy farms (QLK5-2000-31006) as part of the EU-program "Quality of Life and Management of Living Resources", Deliverable D28, 36 pp.
- Ouweltjes, W. and H. Hogeveen. 2001. Detecting abnormal milk through colour measuring. *Natl. Mastitis Council Ann. Meet.* 40:217-219.
- Rasmussen, M.D. M. Bjerring and F. Skjøth. 2002a. Visual inspection of foremilk. In Rasmussen, M.D. (ed.) Internal report no. 169 for Workshop of the EU-project (QLK5-2000-31006): Implications of the introduction of automatic milking on dairy farms. November 27, 2002, 47-62.
- Rasmussen, M.D., B. Madsen and M. Bjerring. 2002b. Available tests and reference methods. In Rasmussen, M.D. (ed.) Internal report no. 169 for Workshop of the EU-project (QLK5-2000-31006): Implications of the introduction of automatic milking on dairy farms. November 27, 2002, 47-62.
- Rasmussen, M.D., 2003. Consequences of definitions of acceptable milk quality for the practical use of automatic milking systems. EU project "Implications of the introduction of automatic milking on dairy farms (QLK5-2000-31006) as part of the EU-program "Quality of Life and Management of Living Resources", Deliverable D6, 25 pp.

- Smith, K.L., J. E. Hillerton, and R. J. Harmon. 2001. NMC guidelines on normal and abnormal milk based on SCC and signs of clinical mastitis. Natl. Mastitis Council. Madison, WI.
- Trilk, J. 2002. The use of electric conductivity to control milk quality and udder health. In: Proc. First North American Conference on Robotic Milking, March 20-22, Toronto, Canada, V-49-V-54.