University of Natural Resources and Life Sciences, Vienna

Department of Sustainable Agricultural Systems

Division of Livestock Sciences

Association of Somatic Cell Count and Differential Somatic Cell Count with acute and chronic mastitis events of dairy cows

Master Thesis

Mathias Marginter, BSc

Supervisor:

Univ.Prof. DI. Dr. Johann Sölkner

Co-Supervisor:

DI. Lisa Maria Rienesl

Vienna, March 2021

Statutory declaration

I declare that I have prepared, developed and written this thesis independently and I have not used any sources, thoughts or literature of others than clearly stated in the text. The master thesis was not used to award an academic degree at any other university.

Place, date

Signature (Mathias Marginter, BSc)

Dedication

To my aunt Christina and my uncle Herbert

Your confidence and determination will always be my role model.

Acknowledgment

This work was part of the D4Dairy project, which is supported by BMK, BMDW and the provinces of Lower Austria and Vienna in the framework of COMET-Competence Centers of Excellent Technologies. The COMET program is handled by the FFG.

The data for this thesis was kindly provided by ZuchtData EDV Dienstleistungen GmbH.

First, I would like to express my special thanks to my supervisor Univ. Prof. DI. Dr. Johann Sölkner for his guidance and continuous support throughout the entire process of writing this thesis. I am also deeply grateful to my Co-Supervisor DI. Lisa Maria Rienesl for the valuable advice and constructive comments during the whole process of writing this thesis, especially for the support during statistical programming. Special thanks to both of you for the always friendly and pleasant cooperation.

I would also like to acknowledge my colleagues and friends for their great support and for the memorable moments in the years during my studies.

Finally, I must express my very profound gratitude to my family for providing me with unfailing support and continuous encouragement throughout my years of study. Special thanks to my parents who always believed in me and wanted the best for me. This accomplishment would not have been possible without you.

Thank you.

Mathias Marginter

Abstract

The somatic cell count (SCC) is a common indicator for mastitis. It shows the number of somatic cells present in milk. Additional indicators could help to distinguish between infected and uninfected cows, particularly because SCC alone is not always sufficient to detect mastitis. The differential somatic cell count (DSCC) seems to be a promising additional indicator. The DSCC represents the percentage of polymorphonuclear leukocytes (PMN) and lymphocytes in relation to the total somatic cell count using flow cytometry. Milk mid-infrared (MIR) spectroscopy is the method of choice to predict contents of fat, protein, urea and lactose in official milk recording schemes. Furthermore, MIR spectroscopy is used to predict diseases like ketone bodies or mastitis. The first aim of this study was to analyse the association of SCC and DSCC with mastitis diagnoses of the Austrian health monitoring system. In a further step we associated mastitis diagnoses with udder health groups (UHG) based on SCC and DSCC. Secondly, we aimed to evaluate the use of DSCC as additional predictor variable on the accuracy of a current mastitis prediction model based on somatic cell score (SCS) and MIR spectra data. The data for this study originated from the Austrian milk recording system and its health monitoring system (GMON). Test day data including DSCC values and MIR spectra data for one milk testing laboratory was merged with diagnosis data of Fleckvieh, Brown Swiss and Holstein Friesian cows. Mastitis diagnoses were linked with adjacent milk recording test days. Test day records in the range of -/+ 21 days before/after diagnosis were considered as mastitis cases, test day records outside this time window were considered to be healthy. The time window for diseased cases was narrowed for further analysis (-/+ 14, -/+ 7 days). For the healthy reference, only data from cows with no diagnosis in the period were used. Median values of SCC and DSCC were plotted by days in milk (DIM) and parity. First derivatives of a subset of informative MIR spectra were included, spectra were corrected for days in milk and SCC was logarithmically transformed to somatic cell score (SCS). After data preparation, a regression model was established to evaluate the association of SCC and DSCC with mastitis incidences. Furthermore, test day records with diagnoses (-/+ 21 days) were classified by udder health groups (UHG) based on SCC and DSCC. For mastitis prediction models, data was split randomly by farm into 0.6 calibration (train) and 0.4 validation (test). Prediction was done with Partial Least Squares Discriminant Analysis (PLSDA). Indicators of model fit were sensitivity, specificity, balanced accuracy and area under the ROC (Receiver Operating Characteristic) curve (AUC). SCC and the interaction between SCC and DSCC showed significant association with mastitis incidence (time window -/+ 21 days) based on diagnoses. SCC and DSCC were also significantly associated with parity and third of lactation. The UHG method classified 52 % of the mastitis diagnoses correctly as mastitis cases (acute and chronic) and 48 % correctly as healthy and suspicious. The comparison of different sets of predictor variables in the prediction model showed that the current model (MIR + SCS) reaches an AUC value of 0.76 (sensitivity 0.63, specificity 0.76, balanced accuracy 0.70) whereas MIR + SCS + DSCC reaches nearly the same values (AUC 0.76, sens. 0.62 spec. 0.76 bal.acc. 0.69). The best AUC value of 0.80 was reached with MIR + SCS in time window -/+ 7 days. Overall, prediction of mastitis was better with narrower time windows. Results show that there is an association between SCC and DSCC with mastitis diagnoses. For using UHG as practicable management tool for farmers, further research on adapting thresholds for days in milk, parity and milk yield is needed. Results also have shown that the current prediction model (SCS + MIR) could not be improved by adding DSCC as an additional predictor variable.

Key words: mastitis, differential somatic cell count, somatic cell count, MIR spectroscopy, udder health groups, PLS-DA

Zusammenfassung

Die somatische Zellzahl (Somatic Cell Count – SCC) ist ein gängiger Indikator für Mastitis. Damit wird die Anzahl der somatischen Zellen in der Milch ausgewiesen. Da SCC allein nicht immer ausreicht Mastitis zu erkennen, könnten zusätzliche Indikatoren helfen infizierte von nicht infizierten Kühen zu unterscheiden. Die differenzierte somatische Zellzahl (Differential Somatic Cell Count – DSCC) scheint ein vielversprechender zusätzlicher Indikator zu sein. DSCC repräsentiert den Anteil an polymorphkernigen neutrophilen Granulozyten (PMN) und Lymphozyten in Relation zum SCC an. Um diesen Wert zu analysieren wird Fluoreszenzmikroskopie genutzt. Die Mittlere-Infrarot (MIR) Spektroskopie ist die Methode der Wahl in der routinemäßigen Milchleistungsprüfung zur Bestimmung von Milchbestandteilen wie Fett, Protein, Laktose und Harnstoff. Weiters wird die MIR Spektroskopie verwendet um Krankheiten wie Ketose oder Mastitis vorherzusagen. Hauptziel dieser Studie war die Analyse der Assoziation von SCC und DSCC mit Mastitis Diagnosen des österreichischen Gesundheitsmonitoring (GMON). In einem weiteren Schritt wurden Mastitis Diagnosen mit Eutergesundheitsklassen, basierend auf SCC und DSCC assoziiert. Weiters wurde evaluiert, ob ein bestehendes Mastitis Vorhersagemodell, basierend auf dem Linear Somatic Cell Count (SCS, logarithmische Transformation des SCC) und MIR Spektral Daten, durch DSCC als zusätzlichen Indikator verbessert werden kann. Die Daten für diese Studie stammen aus der österreichischen Milchleistungsprüfung und dem Gesundheitsmonitoring. Die Testtagesdaten, welche DSCC und MIR Spektral Daten enthalten, stammen aus einem Labor. Diese Daten wurden mit den Diagnosedaten für die Rassen Fleckvieh, Brown Swiss und Holstein Friesian verknüpft. Die Diagnosedaten wurden mit benachbarten Testtagesdaten verlinkt. Tasttagesdaten im Zeitfenster -/+21 Tage bevor/nach Diagnose wurden als Mastitis deklariert. Testtagesdaten außerhalb dieses Zeitfensters wurden als gesund deklariert. Für nachfolgende Analysen wurde das Zeitfenster für die Definition von Mastitis eingeengt (-/+ 14, -/+ 7 Tage). Für die gesunde Referenz wurden nur Daten von Kühen herangezogen, welche im gesamten Beobachtungszeitraum keine Mastitis Diagnose aufwiesen. Der Median von SCC und DSCC im Verlauf der Laktation bzw. Laktationszahl wurden grafisch dargestellt. Die erste Ableitung der informativen MIR Spektral Daten wurden inkludiert. Weiters wurden diese Daten auf Tage in Milch korrigiert. SCC wurde logarithmisch zum Somatic Cell Score (SCS) transformiert. Nach der Datenaufbereitung wurde ein Regressionsmodell aufgestellt, um die Assoziation von SCC und DSCC mit Mastitis Diagnosen festzustellen. Weiters wurden die Testtagesdaten mit benachbarten Mastitis Diagnosen (-/+ 21 Tage) in Eutergesundheitsklassen geteilt, welche auf SCC und DSCC basieren. Für das Mastitis Vorhersagemodell wurden die Daten zufällig in einen Kalibrierungs-(0,6) und einen Validierungsdatensatz (0,4) geteilt. Partial Least Squares Discriminant Analysis (PLSDA) wurde für die Vorhersage verwendet. Die Indikatoren für die Genauigkeit des Modells waren Sensitivität, Spezifität, Balanced Accuracy und die Fläche unter der ROC (Receiver Operating Characteristic) Kurve (AUC). SCC und die Interaktion zwischen SCC und DSCC zeigten eine signifikante Assoziation mit Mastitis Diagnosen (Zeitfenster -/+ 21 Tage). Weiters bestand ein signifikanter Zusammenhang zwischen SCC bzw. DSCC mit Laktationszahl und Laktationsdrittel. Die Methode der Eutergesundheitsklassen klassifizierte 52 % der Mastitis Diagnosen korrekt als Mastitis (akut und chronisch) sowie 48 % als gesund und verdächtig. Der Vergleich verschiedener Variationen von Variablen für das Vorhersagemodell ergab, dass das aktuelle Modell (MIR + SCS) einen AUC Wert von 0,76 (Sensitivität 0,63, Spezifität 0,76, Balanced Accuracy 0,70) erreicht. Das um DSCC erweiterte Modell (MIR + SCS

+DSCC) erreichte ähnliche Werte (AUC 0,76, Sens. 0,62, Spez. 0,76, Bal.Acc. 0,69). Der beste AUC Wert mit 0,80 konnte mit dem Modell MIR + SCS im Zeitfenster -/+ 7 Tage erreicht werden. Gesamt gesehen war die Vorhersage von Mastitis besser in engeren Zeitfenstern. Die Ergebnisse zeigten, dass eine Assoziation zwischen SCC bzw. DSCC und Mastitis Diagnosen besteht. Um die Eutergesundheitsklassen als Management Tool in der Praxis anwenden zu können, bedarf es noch weiterer Forschung hinsichtlich Adaptierung der Grenzwerte für Laktationsstadium, Laktationszahl und Milchleistung. Weiters zeigten die Ergebnisse, dass das bestehende Vorhersagemodell (SCS + MIR) durch DSCC als zusätzliche Variable nicht verbessert werden konnte.

Schlüsselwörter: Mastitis, Zelldifferenzierung, Zellzahl, MIR Spektroskopie, Eutergesundheitsklassen, PLS-DA

Table of content

List of tables

List of Figures

List of abbreviations

AUC	Area under the receiver operating characteristic curve
BC	Bacterial culture
DIM	Days in milk
DSCC	Differential somatic cell count
ECM	Energy corrected milk
FN	False negative
FP	False positive
GMON	Health monitoring system
GOSE	Good separation
IDF	International dairy federation
IMI	Intramammary infection
PMN	Polymorphonuclear leukocytes
ROC	Receiver operating characteristic
SCC	Somatic cell count
SCS	Somatic cell score
TN	True negative
TP	True positive

1 Introduction

1.1 General Background

Differential somatic cell count (DSCC) is a new indicator for udder health using flow cytometry, provided by FOSS Inc. Somatic cells consist of macrophages, lymphocytes, polymorphonuclear leucocytes (PMN) and epithelial cells (Burvenich et al. 2009). The DSCC represents the combined proportion of polymorphonuclear leukocytes (PMN) and lymphocytes expressed in percentage. The proportion of macrophages may be calculated by subtracting DSCC from 100 %. During a mastitis event, amount and composition of somatic cells and the total somatic cell count (SCC) change evidently (Schwarz s.a.a). SCC in healthy mammary glands is usually low with a count of < 100.000 cells/mL and macrophages represent the predominant cell type (Bobbo et al. 2019). Yet, udder inflammation was also detected with a lower SCC (< 100.000 cells/mL) (Zecconi et al. 2019). Milk from cows with mastitis have higher proportions of PMN, which can reach up to 95 % (Damm et al. 2017). Previously, predominating proportions of macrophages were detected in chronically infected cows (Leitner et al. 2000). Thus, low DSCC at high levels of SCC could be a good indicator of chronic mastitis (Schwarz s.a.a). This evidence supports aiming to apply DSCC as a method to identify mastitis, alone or in combination with SCC.

1.2 Aim of thesis

The first aim of this study was to analyse the association of SCC and DSCC with mastitis diagnoses. In a further step we compared mastitis diagnoses with udder health groups (UHG). Secondly, we aimed to evaluate the effect of the DSCC as additional predictor variable on the accuracy of a current mastitis prediction model based on somatic cell score (SCS) and MIR spectra data.

1.3 Literature review

1.3.1 Mastitis

Mastitis is the most frequent and costly disease in dairy industry (Halasa et al. 2007). The disease is defined as “inflammation of the mammary gland” which is caused by bacteria, mycoplasma, yeasts and algae (Bradley 2002). The occurrence of mastitis is grouped in clinical and subclinical types. The clinical type is an infection which can be observed due to an inflamed quarter of the udder (increased temperature, pain and swellings) and changes in the appearance of milk (e.g. clots in the milk). The subclinical type is an infection without external changes, but the somatic cell count in the milk is increased and the chemical composition of the milk changes. Pathogens can be verified in two of three cases (Blowey and Edmondson 2010; Winter and Zehle 2009). Furthermore, the clinical type can be split in acute and chronic cases. Acute cases are shown through a sudden onset and severe signs like swellings, pain increased body temperature and reduced appetite. Chronic cases persist for a long time. Only the udder shows changes like decreasing mammary gland and nodes formation, but the cow looks healthy. The secretion will be reduced but not altered, sometimes clots can be found in the milk (Blowey and Edmondson 2010; Winter and Zehle 2009). Mastitis pathogens can be classified into two types, contagious or environmental (Cervinkova et al. 2013). Contagious pathogens are organisms adapted to survive within the hosts

mammary gland. They are spread from cow to cow usually at the time of milking. Environmental pathogens invade in the mammary gland, engender an immune response in the host and are rapidly eliminated (Bradley 2002). They occur in the cows environment such as bedding, soil or manure and thus are highly influenced by management practices (Garcia 2004). The major contagious pathogens are: *Staphylococcus aureus, Streptococcus agalactiae, Streptococcus dysgalactiae.* The major environmental pathogens are: *Streptococcus uberis* and *Escherichia coli* (Blowey and Edmondson 2010).

Mastitis causes economic losses to the farm which are often underestimated by farmers (Huijps et al. 2008). These economic losses are caused by direct and indirect costs which are defined differently in literature. The most common ones are:

- Milk production losses
- Drug and veterinary costs
- Discarded milk
- Extra labour requirements for treatment
- Higher culling and replacement rates (Blowey and Edmondson 2010; Halasa et al. 2007; Heikkilä et al. 2012; Hogeveen and Winter 2009)

Aghamohammadi et al. (2018) estimated herd-level costs of clinical and subclinical mastitis in Canadian dairy farms. Figure 1 shows the relative importance of the different cost components for a median herd in Canada (100 cows per year).

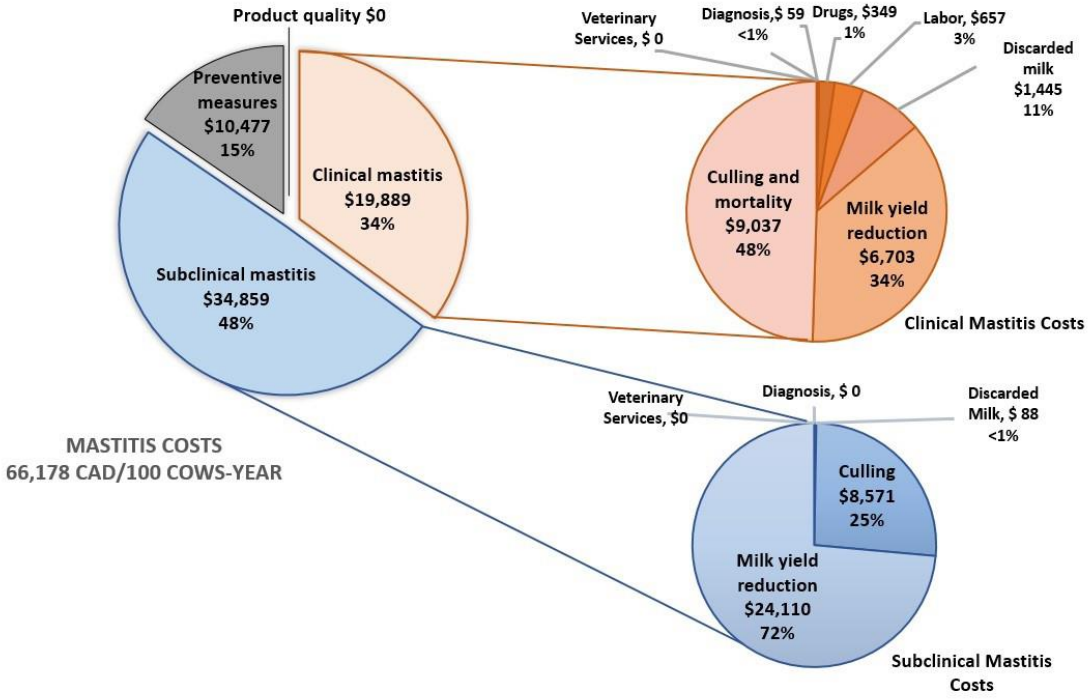


Figure 1: Absolute values and relative importance (in %) of the different cost-components for the median herd in Canada (100 cows-year) (Aghamohammadi et al. 2018).

1.3.2 Somatic cell count (SCC)

Right now, somatic cell count (SCC) is the most important indicator of mastitis. It shows the number of cells presented in milk. The cells are a combination of leucocytes and epithelial cells (Blowey and Edmondson 2010). Epithelial cells get shed from the tissue and renewed. 75 % of somatic cells are leucocytes, which enter the mammary gland as a response to injury or infection. 25 % are represented by epithelial cells (Sharma et al. 2011). The SCC of healthy cows is quite constant over lactation and increases immediately after calving and before the dry period (Burvenich et al. 2009). According to International Dairy Federation (IDF 2013) uninfected mammary glands show a SCC of under 100,000 cells per ml. The SCC increases if bacteria entry the mammary gland. Therefore, the SCC can be used as a good indicator for intramammary infection (IMI). SCC is also influenced by breed, parity and season (Sharma et al. 2011), which should be considered for using as indicator. The IDF (2013) sets the cut-off to distinguish between cows likely to be infected and normal at 200,000 cells/ml. These thresholds for uninfected mammary glands (SCC < 100,000 cells/ml) and cows infected (SCC > 200,000 cells/ml) are also used in Austrian milk recording program (Egger-Danner 2018).

1.3.3 Differential somatic cell count (DSCC)

Somatic cells consist of macrophages, lymphocytes, polymorphonuclear leucocytes (PMN) and epithelial cells (Burvenich et al. 2009). These immune cells play a major role when bacteria invade the mammary gland and cause an IMI (Nickerson 1989; Paape et al. 1979; Sordillo and Nickerson 1988). Lymphocytes regulate the induction and suppression of immune response (Nickerson 1989). Phagocytes (PMN and macrophages) digest and kill bacteria causing mastitis, but also absorb globules of milk fat and casein. That leads to a remarkable morphology in milk, which is easy to observe through microscopy or flow cytometry (Burvenich et al. 2009). Phagocytosis plays a major part in defending against invading pathogens (Sordillo and Nickerson 1988). PMNs have a lifetime of few hours and recognize, digest and kill microorganisms. In early stage of mastitis, the proportion of PMNs can be up to 90 % of SCC. The digestion of pathogens takes place trough phagocytosis (Burvenich et al. 2009). Macrophages are the prevalent cells in healthy udders and recognize invading pathogens. They are involved in the immune response and arrange a fast influx of PMNs in the mammary gland (Burvenich et al. 2009).

In Table 1 the proportion of macrophages, lymphocytes, PMNs and epithelial cells is shown with different values of SCC.

Table 1: Distribution of milk cells in milk with different SCC (Burvenich et al. 2009)

Cells	Healthy milk (SCC <100,000 cells/ml, physiological)	SCC 100,000 – 400,000 cells/ml	SCC >400,000 cells/ml
PMN	12 %	63 %	87 %
Lymphocytes	28 %	11 %	9 %
Macrophages	58 %	25 %	3 %
Epithelial cells	2 %	1 %	1 %

The SCC and also the differentiation of the cells help to draw conclusions about the health status of the mammary gland (Burvenich et al. 2009). To analyse and use the differentiation of SCC, FOSS Inc. provided a new indicator called differential somatic cell count (DSCC). DSCC represents the combined proportion of PMN and lymphocytes expressed in percentage using flow cytometry. The proportion of macrophages can be calculated by subtracting DSCC from 100 % (Schwarz s.a.a).

Foss Inc. developed a new patented method which can identify SCC and DSCC simultaneously. The measuring module in the Fossomatic 7 DC has a long-lasting LED laser as light source. The fluorescence and morphology information of each cell is received by three detectors. The fluorescence signals are measured in two channels, FL 1 and FL 2 (Schwarz s.a.b). Fluorescence emission signals from channels 1 and 2 are used for determination of the SCC (Figure 2A) (Damm et al. 2017).

A parameter called Good Separation (GOSE) defines the accuracy of the sample. If the blue dots (milk cells) in the orange rectangle are separated from the grey dots (background), as seen in Figure 2A, the GOSE value will be 1 and the analyses of this sample can be trusted. If the blue dots are not noticeable separated from the grey dots, the GOSE value will be 0 and the analyses of this sample cannot be trusted (Schwarz 2020).

In a second step only cells (i.e. SCC) identified in Figure 2A are further investigated in a dot plot seen in Figure 2B. The cells are differentiated in macrophages (grey dots) and PMN together with lymphocytes (orange dots) (Damm et al. 2017).

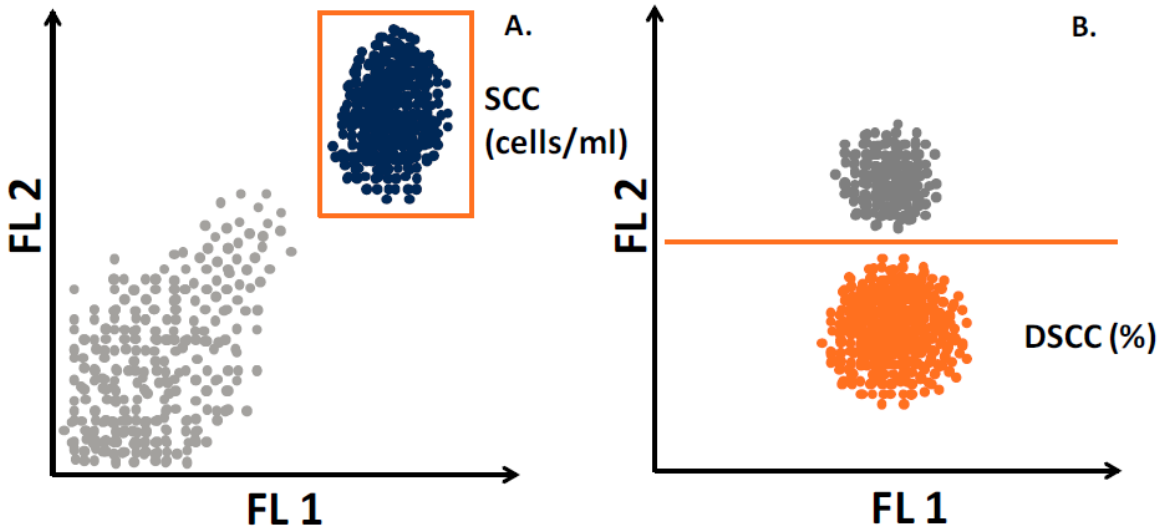


Figure 2: Schematic diagram of the measuring principle of the new Fossomatic 7 DC. A: Separation between background (grey dots) and milk cell (dark blue dots) fluorescence information and determination of SCC (orange box). B: Only somatic cells from Figure 2A are used for determination of DSCC by differentiating macrophages (grey dots) from lymphocytes and PMN (orange dots). FL = fluorescence emission (Schwarz s.a.b)

Kirkeby et al. (2020) found that DSCC in addition to SCC shows a significantly improved indication of IMI compared with SCC alone. They also found that SCC is significantly affected by IMI status, DIM and parity and concluded, that this also should be considered when interpreting DSCC in relation to IMI. The results also showed that DIM, parity and pathogen group were associated with DSCC (Kirkeby et al. 2020). Another study of Schwarz et al. (2020b) compared SCC alone, DSCC alone and the combination of both based on test characteristics using exemplary cut-offs. The IMI

status was detected by bacterial culture (BC). For example, a cut-off of SCC 200,000 cells/ml alone compared with the combination cut-off DSCC 65 % and SCC 200,000 cells/ml to classify cows as infected by major pathogens shows, that sensitivity increased from 78 % to 92 % and specificity decreased from 87 % to 66 % (Schwarz et al. 2020b).

Zecconi et al. (2019) analysed DSCC data under field conditions by Receiver Operating Characteristic (ROC) procedure. They especially investigated subclinical mastitis cases, which are defined as test records with SCC > 200,000 cells/ml. Three DSCC thresholds (66.3 % 69.2 % and 69.3 %) were for each third of lactation (≤ 100 DIM, 101 – 200 DIM, > 200 DIM) which performed best with area under the ROC curve (AUC) values of 0.91, 0.92 and 0.82. Bobbo et al. (2020) used these DSCC thresholds in combination with SCC threshold of 200,000 cells/ml to create udder health groups (UHG). Their findings highlighted that significant differences exist among UHGs in milk production and quality. Also Schwarz et al. (2020a) compared different UHGs but with another DSCC threshold. They classified four UHGs with thresholds based on former studies (Leitner et al. 2000; Schwarz et al. 2011; Schwarz et al. 2019; Schwarz et al. 2020b), as shown in Table 2.

Table 2: Definition of Udder health groups (UHG) (Schwarz et al. 2020a)

Group	Health status	SCC (cells/ml)	DSCC (%)
A	Healthy/normal	≤ 200,000	≤ 65
B	Suspicious/ onset of mastitis	≤ 200,000	> 65
C	Mastitis	> 200,000	> 65
D	Chronic/persistent mastitis	> 200,000	≤ 65

The authors also reached the conclusion that there are significant differences between UHGs for milk weight, energy corrected milk (ECM) and milk components (fat, protein, lactose) (Schwarz et al. 2020a).

As mentioned in chapter 1.3.2, healthy mammary glands show a SCC of under 100,000 cells/ml and the cut-off to distinguish between healthy and infected cows is 200,000 cells/ml (IDF 2013). Schwarz et al. (2010) confirmed the quarter foremilk threshold of 100,000 cells/ml for differentiating between infected and noninfected mammary glands. But they also found 8.5 % prevalence of mastitis pathogens in mammary glands with SCC range from 1,000 to 100,000 cells/ml. These inflammations could be better indicated by using DSCC as an additional parameter (Schwarz et al. 2010).

1.3.4 Milk mid-infrared (MIR) spectroscopy

Spectroscopy is a technique based on study of interaction between matter and electromagnetic waves (Marchi et al. 2014). Milk mid-infrared (MIR) spectroscopy is the method of choice to predict contents of fat, protein, urea and lactose in official milk recording schemes around the world (Grelet et al. 2015; Grelet et al. 2016).

The electromagnetic radiation is split in different regions according to wavelengths: x-ray region (0.5 – 10 nm), UV region (10 – 350 nm), visible region (350 – 800 nm), near-infrared region (800 – 2,500 nm), mid-infrared region (2,500 – 25,000 nm),

microwave region (100 µm – 1 cm) and radio frequency region (1 cm – 1 m) (Marchi et al. 2014).

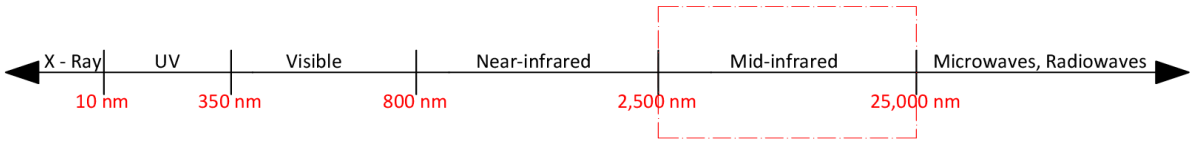


Figure 3: Electromagnetic spectrum according to Marchi et al. (2014)

According to the international norm ISO 9622:2013 for milk and liquid milk products, the milk sample is analysed after pre-treatment and homogenization in a so-called infrared spectrometer. That instrument records the quantity of radiation in transmittance at specific wavelengths in the MIR region. These obtained spectral data are transformed into estimates of constituent concentrations or other physico-chemical parameters through calibration models (ISO 2013).

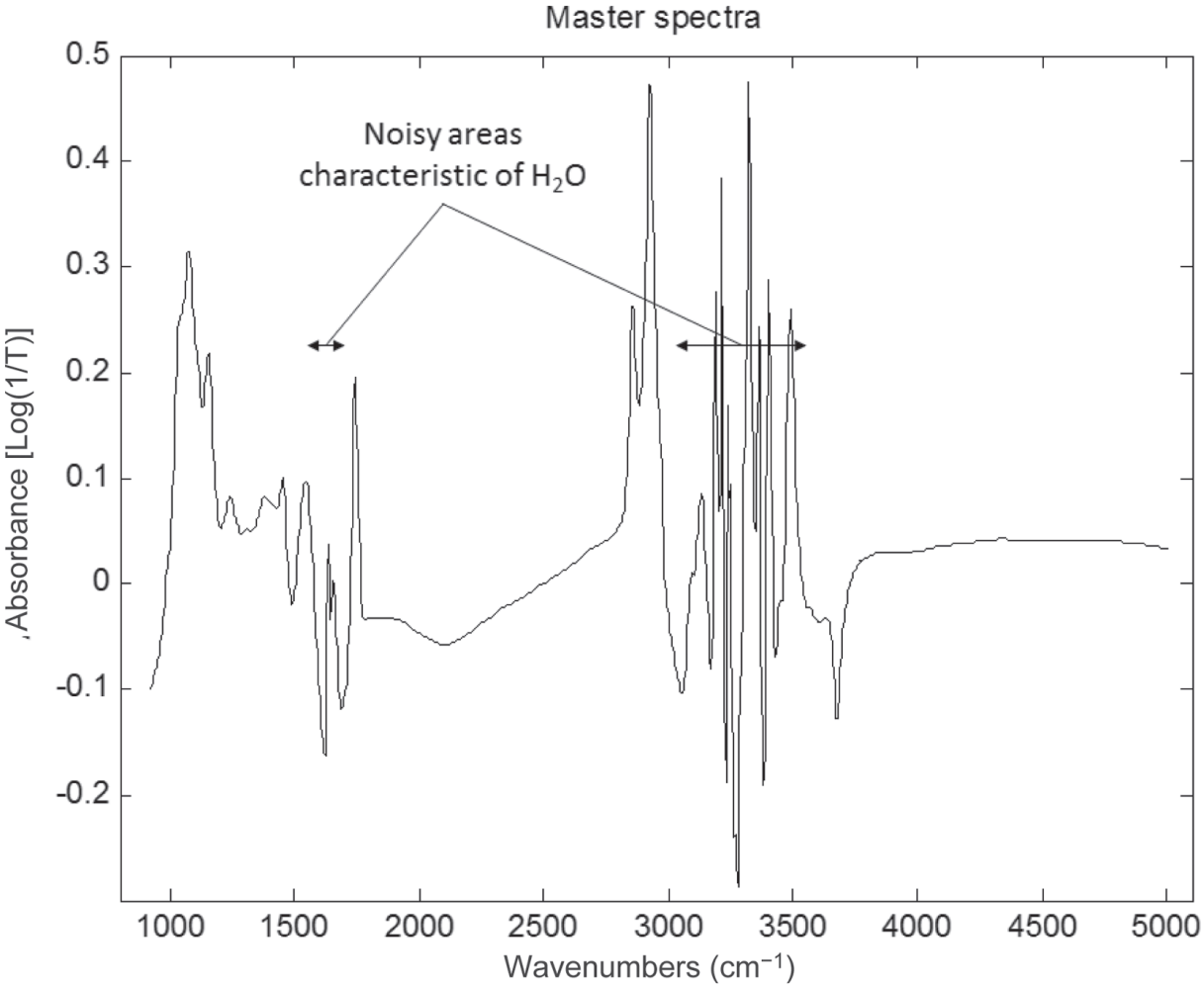


Figure 4: Typical MIR absorption curve (Grelet et al. 2015)

MIR spectroscopy is also used to predict fatty acid profile (Rutten et al. 2009), protein composition (Bonfatti et al. 2011), lactoferrin (Soyeurt et al. 2007), or minerals (Soyeurt et al. 2009). Furthermore, there are a few studies on predicting pregnancy according to MIR spectra data (Lainé et al. 2017; Rienesl et al. 2020; Toledo-Alvarado et al.

2018). There are also several studies on predicting diseases with MIR spectroscopy, e.g. ketone bodies (Roos et al. 2007), lameness (Bonfatti et al. 2020) and mastitis (Rienesl et al. 2019; Soyeurt et al. 2012).

2 Material and Methods

2.1 Data

The data for this study originated from the Austrian milk recording system and its health monitoring system (GMON), considering milk samples processed at the milk laboratory in St. Michael, Styria, from August 2019 to July 2020 and was provided by ZuchtData GmbH. At the milk laboratory in St. Michael data of test day records from Styria and parts of Salzburg and Carinthia are analysed on two machines. One machine can analyse the DSCC value of test day milk with SCC > 52,000 cells/ml. The test day data consisted of information on breed, herd, milk weight, fat, protein, urea, lactose, SCC, GOSE (only machine with DSCC tool), DSCC and MIR spectra data for each record. The respective GMON data included mastitis diagnoses of acute and chronic mastitis events and date of diagnosis.

In a first step we deleted all data with a missing GOSE value, because the machine without the DSCC tool does not provide this value. After that step 374,375 test day records were available, 209,044 (56 %) with a DSCC value > 0. In the next step we prepared test day data and GMON data for merging. Test day records of Fleckvieh, Brown Swiss and Holstein Friesian between 3 and 305 days in milk (DIM) were included. Test day records with missing MIR spectra data, milk weight or SCC were excluded. Also records with SCC > 52.000 cells/ml who had a DSCC of 0 were deleted. Acute and chronic mastitis diagnoses (diagnostic code 51 and 52) were included in the GMON data set. In both data sets only farms with ≥ 75 % electronically transmitted diagnoses (validated farms) were included.

After merging of test day data with GMON data, we defined test day records as mastitis cases or healthy cases for prediction models. Therefore, mastitis diagnoses were linked with adjacent test day records. Test day records in the range of 21 days before and 21 days after diagnosis were considered as mastitis cases, test day records outside this time window were considered to be healthy. For further analysis, the time window was narrowed (-/+ 14, -/+7 days).

Additionally, a healthy reference was created to analyse SCC and DSCC during lactation in heathy cows. Therefore, only test day records with no diagnosis in the full period were included.

MIR data pre-treatment was done according to studies by Grelet et al. (2016), Rienesl et al. (2019) and Soyeurt et al. (2012). For final prediction model the first derivative of MIR spectra data was used. Data preparation up to here was done with SAS (SAS Institute Inc. 2018).

Further data pre-treatment of MIR spectra was done with RStudio (R Development Core Team 2008). Commonly, selected parts of the 1,060 data points of MIR spectra were used for modelling. Therefore, we selected the following spectral area: 968.1 to 1,577.5 cm^{-1}, 1,731.8 to 1,762.6 cm^{-1}, 1,781.9 to 1,808.9 cm^{-1} and 2,831.0 to 2,966.0 cm^{-1}. This selected areas with 212 datapoints contain most of the information. The removed parts are less reliable because of strong water absorbance or worse repeatability among different instruments (Grelet et al. 2016). To correct the selected datapoints for DIM we used the method according to Vanlierde et al. (2015): Each first derivative spectrum was multiplied by a constant (i.e. 1), a linear $\left(\sqrt{3} \times x\right)$ and a quadratic $\left[\sqrt{5/4} \times (3x^2 - 1)\right]$ modified Legendre polynomial (Gengler et al. 1999),

where $x = -1 + 2[(DIM - 5)/(365 - 5)]$. The SCC was transformed logarithmically to the Somatic Cell Score (SCS) with following formula:

$SCS = \log_2(SCC/100{,}000) + 3$ (Fürst et al. 2019).

The final data set for modelling included 109,084 test day records with 1,421 mastitis diagnoses of it. Details of the data are given in Table 3.

Table 3: Final data set

Test day records	**109,084**
data with DSCC > 0	57,984
	53.16 %
Healthy	107,663
	98.70 %
Mastitis	1,421
	1.3 %
Acute	1,153
	81.14 %
Chronic	268
	18.86 %
Farms	1,739
Cows	40,332
Fleckvieh	30,801
	76.37 %
Brown Swiss	4,539
	11.25 %
Holstein Friesian	4,992
	12.38 %
Healthy reference	= data with no diagnosis in the period
Test day records	98,843
	90.61 %
Cows	37,372
	92.66 %

2.2 Methodology

2.2.1 Analyses of SCC and DSCC in healthy reference

In a first step we analysed the SCC and DSCC value along the lactation period and for parities (1, 2, 3, 4, 5+) with the healthy reference data. We plotted curves with median SCC and DSCC by DIM and boxplots by parity, where for DSCC plots only data with DSCC > 0 were included. Furthermore, we also did tests for differences in means (parity and third of lactation) with a p-value threshold of 0.05. First third of lactation is defined from 5 – 100 DIM, second from 101 – 200 DIM and third from 201 – 305 DIM. Most values of the SCC are in low and middle range and only few are in high ranges. Thus the SCC does not correspond to a normal distribution (Fürst et al. 2019). Therefore, to test the significance of differences in means we transformed the SCC to SCS, which approximately apply with a normal distribution. We used the same formula as described in chapter 2.1. After calculation, SCS was back transformed to SCC by the following formula:

$$SCC = 100{,}000 \times 2^{(SCS-3)}$$

2.2.2 SCC and DSCC before/after diagnosis

As mentioned above we linked mastitis cases with adjacent test day records. We produced plots with a time window of -/+ 50 days before/after diagnoses for SCC and DSCC, respectively. Furthermore, we included a regression line for each plot.

2.2.3 Association of SCC and DSCC with mastitis incidences

To analyse the association of SCC and DSCC with mastitis incidences we set up the following regression model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

where Y_{ijk} denotes diagnosis of cow k, SCC class i and DSCC class j. μ is the intercept for all Y_{ijk}. α_i denotes the fixed effect of SCC class i (i=1,…,9) and β_j the fixed effect of DSCC class j (j=0,…,6). $(\alpha\beta)_{ij}$ denotes the effect of interaction between SCC class i and DSCC class j. ε_{ijk} denotes the residual effect. SCC class was divided by 50,000 cells/ml and for DSCC classes we used the 5 %, 25 %, 50 %, 75 % and 95 % quantile. The classification of SCC and DSCC can be seen in Table 4. Within this model we defined mastitis as diagnoses with linked test days in the range of -/+ 21 days before/after diagnosis and narrowed the time window to -/+ 14 and -/+ 7 days for further analysis. We did separate analyses for acute and chronic mastitis and also one analysis for mastitis, acute or chronic. Furthermore, we plotted the outputs of the models. Calculations were done with proc glm in SAS (SAS Institute Inc. 2018). We analysed the overall F-test and the Type III Sum of Squares output. The Type III test provides evidence of the significance of an effect over and above the other effects in the model. In our case as an example, DSCC was tested as an additional parameter to the model including SCC and the interaction between SCC and DSCC.

Table 4: Classification of SCC and DSCC

SCC class	x 1,000 cells/ml	DSCC class	%
1	< 52	0	0
2	≥ 52 - < 100	1	> 0 - < 45.4
3	≥ 100 - < 150	2	≥ 45.4 - < 62.9
4	≥ 150 - < 200	3	≥ 62.9 - < 72.6
5	≥ 200 - < 250	4	≥ 72.6 - < 80.0
6	≥ 250 - < 300	5	≥ 80.0 - < 87.5
7	≥ 300 - < 400	6	≥ 87.5
8	≥ 400 - < 600		
9	≥ 600		

2.2.4 Comparison Udder health groups (UHG) and Diagnoses

Following some studies (Bobbo et al. 2020; Schwarz et al. 2020a; Zecconi et al. 2019) we created UHG to compare them with mastitis diagnoses from GMON. We defined the thresholds of UHG according to Schwarz et al. (2020a). The threshold of SCC is 200,000 cells/ml and the one of DSCC is 65 %. The classification of the groups is listed in Table 2 in chapter 1.3.3. To evaluate the reliability of UHG to predict mastitis cases we set up a confusion matrix. This is a table categorising predictions (UHG in our case) according to whether they match the actual value (Diagnosis in our case) (Lantz 2015). There are four categories shown in Table 5. True positive (TP) are correctly classified mastitis diagnoses by UHG. True negative (TN) are correctly classified healthy cows by UHG. False positive (FP) are incorrectly classified mastitis diagnoses by UHG. False negative (FN) are incorrectly classified healthy cows by UHG.

Table 5: Confusion matrix of UHG and mastitis diagnoses

		Mastitis according to diagnosis	
		1	0
Mastitis according to UHG	1	**TP**	**FP**
	0	**FN**	**TN**

To measure the proportion of cases that are correctly classified we used sensitivity and specificity. Sensitivity gives the proportion of correctly classified mastitis cases and

specificity the proportion of correctly classified healthy cases. Sensitivity and specificity are calculated with following formulas:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

Balanced accuracy is defined as the mean of sensitivity and specificity and was also used as indicator of model fit (Lantz 2015).

We created two models with different definitions of healthy and mastitis. In the first model ("original") UHG 1 was classified as healthy and UHG 2, 3 and 4 were classified as mastitis cases. In the second model ("modified") UHG 1 and 2 were classified as healthy and UHG 3 and 4 as mastitis cases. We also narrowed the time window for diagnosis and calculated the indicators of model fit for each time window (-/+21, -/+14, -/+7) respectively.

Furthermore, we investigated the thresholds of SCC (200,000 cells/ml) and DSCC (65 %) for classification of UHG. We plotted the number of test day records with SCC > 200,000 cells/ml by third of lactation and parity. Therefore, we used the final data set and for comparison the data set of healthy reference (= no mastitis diagnosis in the full period). Additionally, we plotted the number of all mastitis diagnoses from GMON if they were in lactation. Furthermore, we plotted mastitis cases defined as diagnosis from GMON with adjacent test days in the range of 21 days before and after. We also did a t-test for significance of differences in means with a p-value of 0.05. The same plots and tests were done for the number of test day records with DSCC > 65 %.

2.2.5 Prediction model

For the mastitis prediction model, the final data set was split randomly by farm into 0.6 calibration (train) and 0.4 validation (test). The models were done with Partial Least Squares Discriminant Analysis (PLSDA) using the 'caret' package in R (R Development Core Team 2008). The number of latent variables was set automatically (with a maximum of 70) and we ran 10 replications per setting. The indicators of model fit were sensitivity, specificity, balanced accuracy (as described in chapter 2.2.4) and AUC (area under the ROC curve). An example for a ROC diagram is shown in Figure 5. The ROC curve is defined on a plot with the proportion of true positive rate (= sensitivity, vertical axis) and the proportion of false positive rate (= 1 – sensitivity, horizontal axis). The diagonal line represents the classifier with no predictive value, which detects TP and FP at the same rate (= random classification) (Lantz 2015). Consequently, a ROC curve above the diagonal line means that the prediction is better than random classification and a ROC curve under the diagonal line means that the prediction is worse. The AUC value measures the total area under the ROC curve and ranges from 0.5 (classifier with no predictive value) to 1.0 (perfect classifier) (Lantz 2015). The AUC value can be divided in classes (Table 6) and indicates the accuracy of a prediction model.

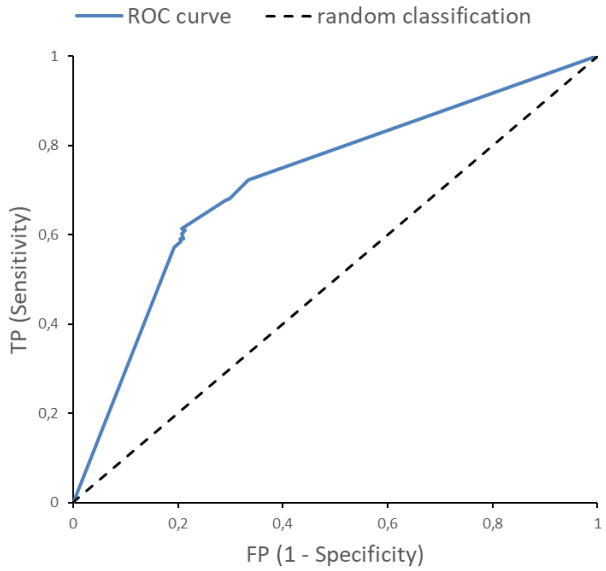


Figure 5: ROC diagram example

Table 6: Classification of AUC values according to Lantz (2015)

Classification	AUC value
Outstanding	0.9 – 1.0
Excellent/ good	0.8 – 0.9
Acceptable/ fair	0.7 – 0.8
Poor	0.6 – 0.7
No discrimination	0.5 – 0.6

We set up 7 models, each of the three variables alone (MIR, SCS and DSCC) and all combinations out of the three for time window -/+ 21 days (definition of mastitis and healthy). Furthermore, we calculated the prediction models with narrowing time windows (-/+ 14, -/+7). Thus, number of mastitis and healthy cases changed and sensitivity and specificity. We also split the results of each of the three time windows in shorter time windows. That changed the sensitivity but not the specificity, because the time window was only narrowed for validation and thus number of mastitis cases changed, but healthy cases remained unchanged.

3 Results

3.1 Analyses of SCC and DSCC in healthy reference

The results of the t-test on SCS and DSCC by parity are shown in Table 7. LS means of SCS were significantly different between lactations. The results showed that there is a significant increase of SCS and thus SCC by parity. LS means of DSCC were also significantly different between lactations, except between lactation 2 and 3. The results showed a DSCC range of 73 % in first lactation and 71 % in fifth and higher lactations.

Table 7: Differences of SCS and DSCC per parity with t-test (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Lactation	SCS LS means	SCC back transf., cells/ml	DSCC LS means, %
1	1.78 **a**	43,043	72.69 **a**
2	2.03 **b**	50,889	68.26 **b**
3	2.34 **c**	63,234	68.45 **bc**
4	2.55 **d**	73,135	69.32 **d**
5+	2.90 **e**	93,080	71.07 **e**

LS means with same letter are not significantly different.

Table 8 shows the results of t-test on SCS and DSCC by third of lactation. LS means of SCS and DSCC were significantly different between thirds of lactation. The results showed that SCC increases with DIM and DSCC decreases with DIM.

Table 8: Differences of SCS and DSCC per third of lactation with t-test (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Third of lactation	SCS LS means	SCC back transf., cells/ml	DSCC LS means, %
1	1.86 **a**	45,400	73.30 **a**
2	2.24 **b**	58,907	70.29 **b**
3	2.74 **c**	83,627	67.84 **c**

LS means with same letter are not significantly different.

Figure 6 shows a boxplot diagram of SCC and DSCC by parity. It can be seen that SCC mean and median increased with parity. Furthermore, the wide data range of SCC can be observed, which explains the need to transform the SCC to SCS. DSCC mean and median decreased until the third lactation and then increased.

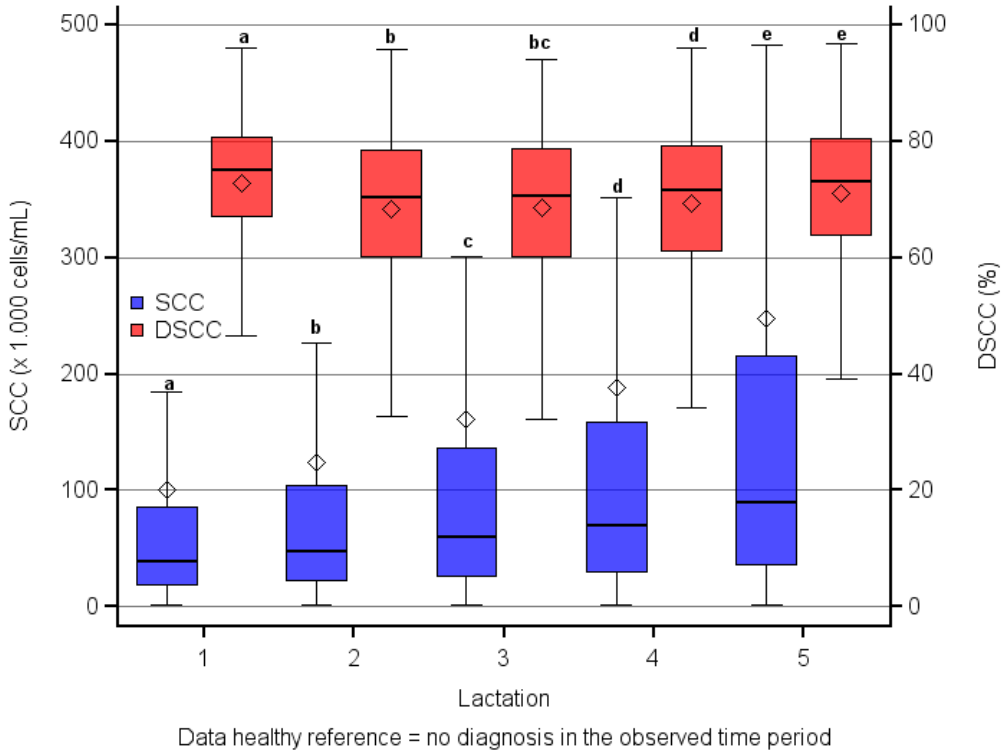


Figure 6: Boxplot SCC and DSCC by parity. The box indicates the interquartile range, the bar inside the box is the median and the diamond is the mean. The whiskers are drawn from the box to the most extreme point that is less than or equal to 1.5 times the interquartile range. Means with same letter are not significantly different.

In Figure 7 SCC and DSCC medians are plotted by DIM. For DSCC we only used data > 0. Otherwise, the median would be falsified, because DSCC could only be analysed when SCC is higher than 52,000 cells/mL and as seen in Figure 7, median SCC was rising by more than 52,000 cells/mL until second third of lactation. The figure shows that, after a short phase in early lactation, SCC increased and DSCC decreased with days in milk.

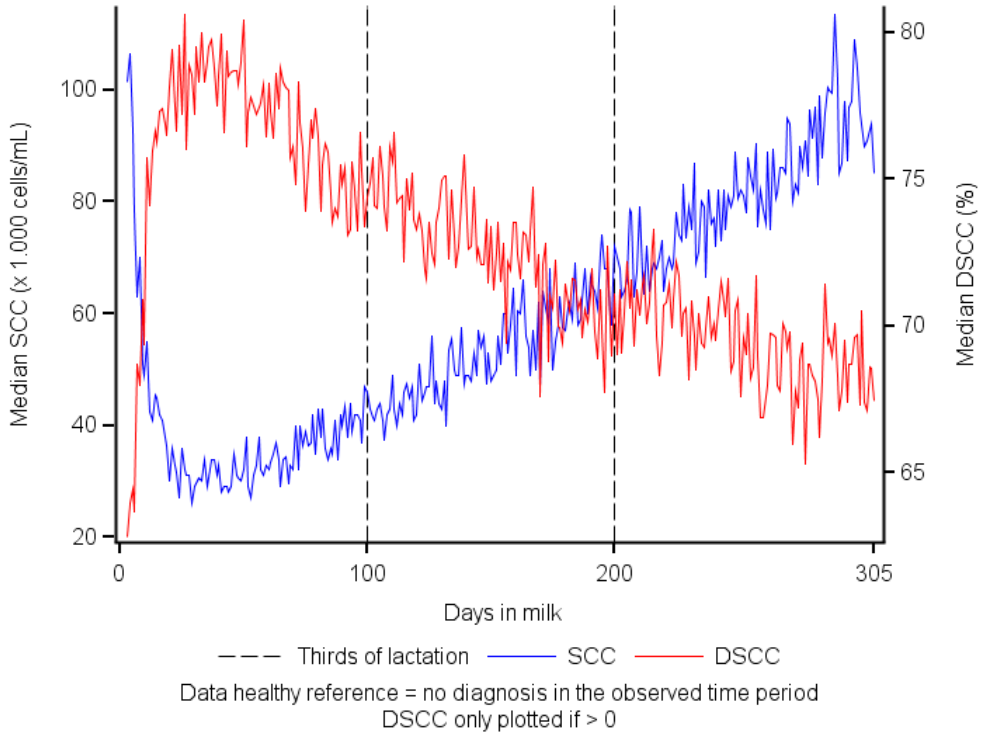


Figure 7: Median SCC and DSCC by days in milk

3.2 SCC and DSCC before/after diagnosis

Figure 8 shows the development of SCC median before and after mastitis diagnosis. We plotted acute and chronic mastitis events separately. The line for acute mastitis shows a rapid increase of SCC ten days before diagnosis with a peak at the time of diagnosis. After diagnosis, a rapid decrease could be observed back to a normal SCC level within 21 days. SCC median of chronic mastitis events showed a up and down 50 to 21 days before diagnosis. From 21 days before the median SCC increased to the peak at 6 days before diagnosis and then decreased with up and downs until 30 days after diagnosis. The regression line was plotted for combined acute and chronic mastitis and showed an increase of SCC until 7 days before diagnosis with a decrease afterwards.

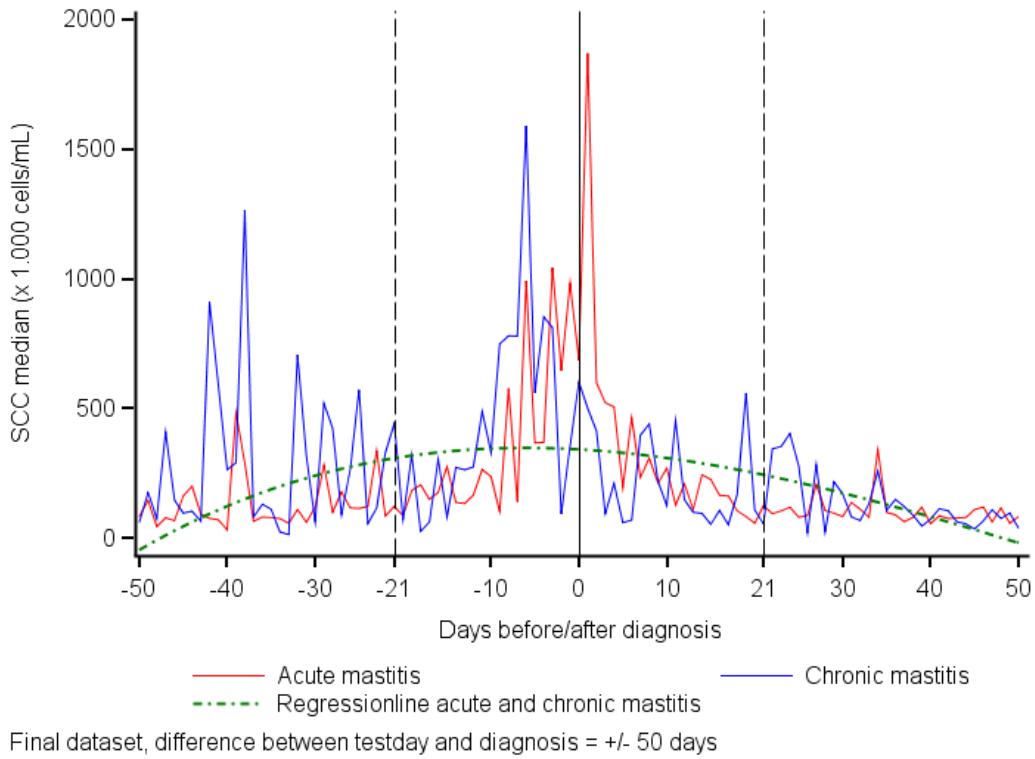


Figure 8: Development of SCC before and after Diagnosis

The development of DSCC median before and after diagnosis can be seen in Figure 9. We also plotted acute and chronic mastitis events in separate graphs and thus with separate regression lines. For acute mastitis, a permanent increase of DSCC could be observed from 50 days before diagnosis until day of diagnosis with a decrease after diagnosis. According to the regression line the DSCC for chronic mastitis events also showed an increase of median before and a decrease after diagnosis but with peaks at high and low levels.

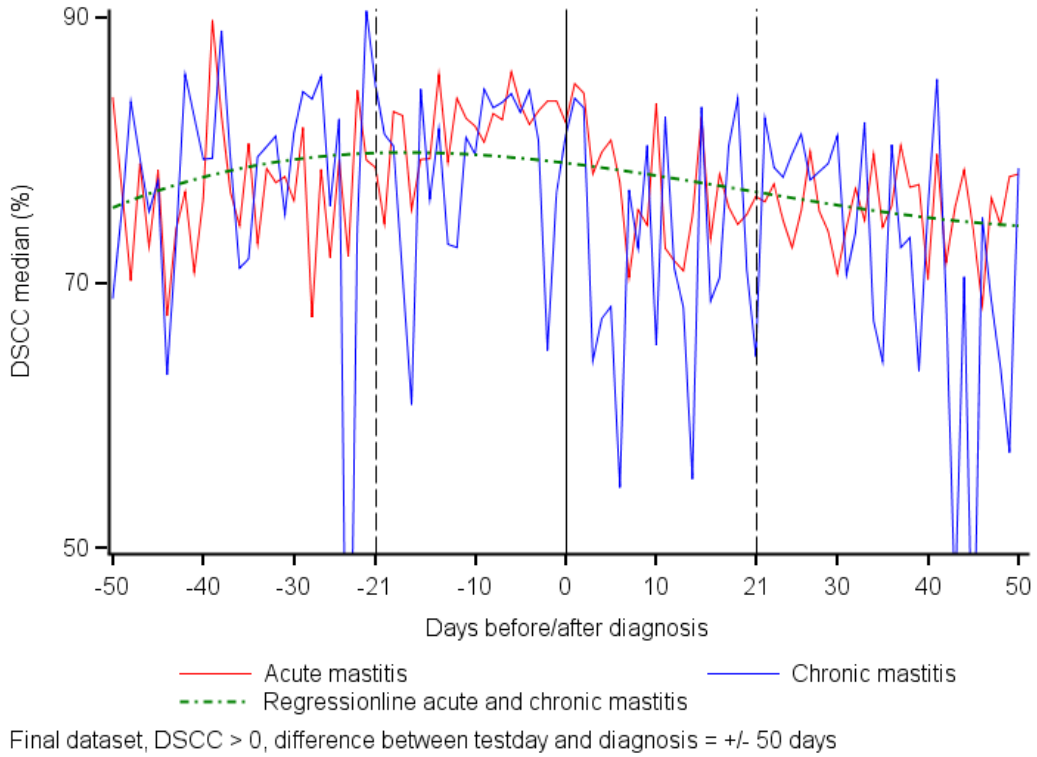


Figure 9: Development of DSCC before and after Diagnosis

3.3 Association of SCC and DSCC with mastitis incidences

The overall F-test was significant for all models, explained in chapter 2.2.3, with a p-value < 0.0001. This means that at least one of the variables (SCC, DSCC and interaction between SCC and DSCC) was significantly associated with diagnosis of mastitis. In Table 9 the output of SAS proc glm (SAS Institute Inc. 2018) for the model with acute and chronic mastitis with definition of mastitis as diagnoses with linked test days in the range of 21 days before and after diagnosis is shown. The tables for other time windows and acute and chronic mastitis separately analysed are listed in the appendix. All outputs, except one, showed that SCC, considered in addition to DSCC and the interaction of SCC and DSCC was significantly associated with diagnosis of mastitis. The output of the model with only chronic diagnoses and linked test days in the range of 7 days before and after diagnosis showed no significance of SCC, with a p-value of 0.0836. DSCC as additional variable to SCC and the interaction between SCC and DSCC showed no significance in any model. The interaction between SCC and DSCC in addition to the other variables was significant in all, except one, models. The excepted model was again the one with only chronic diagnoses and -/+ 7 days difference between diagnosis and test day.

Table 9: Association of SCC and DSCC with acute and chronic mastitis (Difference between test day and diagnosis = -/+ 21 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	30.231255	0.581370	46.19	<0.0001
Error	109031	1372.257863	0.012586		
Corrected Total	109083	1402.489118			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.021555	861.2115	0.112187	0.013027		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	2.00639164	0.25079895	19.93	<0.0001
DSCC	6	0.03276524	0.00546087	0.43	0.8567
DSCC*SCC	38	1.04591101	0.02752397	2.19	<0.0001

Figure 10 shows the association of SCC and DSCC with acute and chronic mastitis incidences. Mastitis is defined as diagnosis with linked test days 21 days before and after diagnosis. The figure shows that in higher SCC or DSCC classes also the LS means of diagnoses were higher, which describes the association of SCC and DSCC with mastitis. The plots from the other models are shown in the appendix.

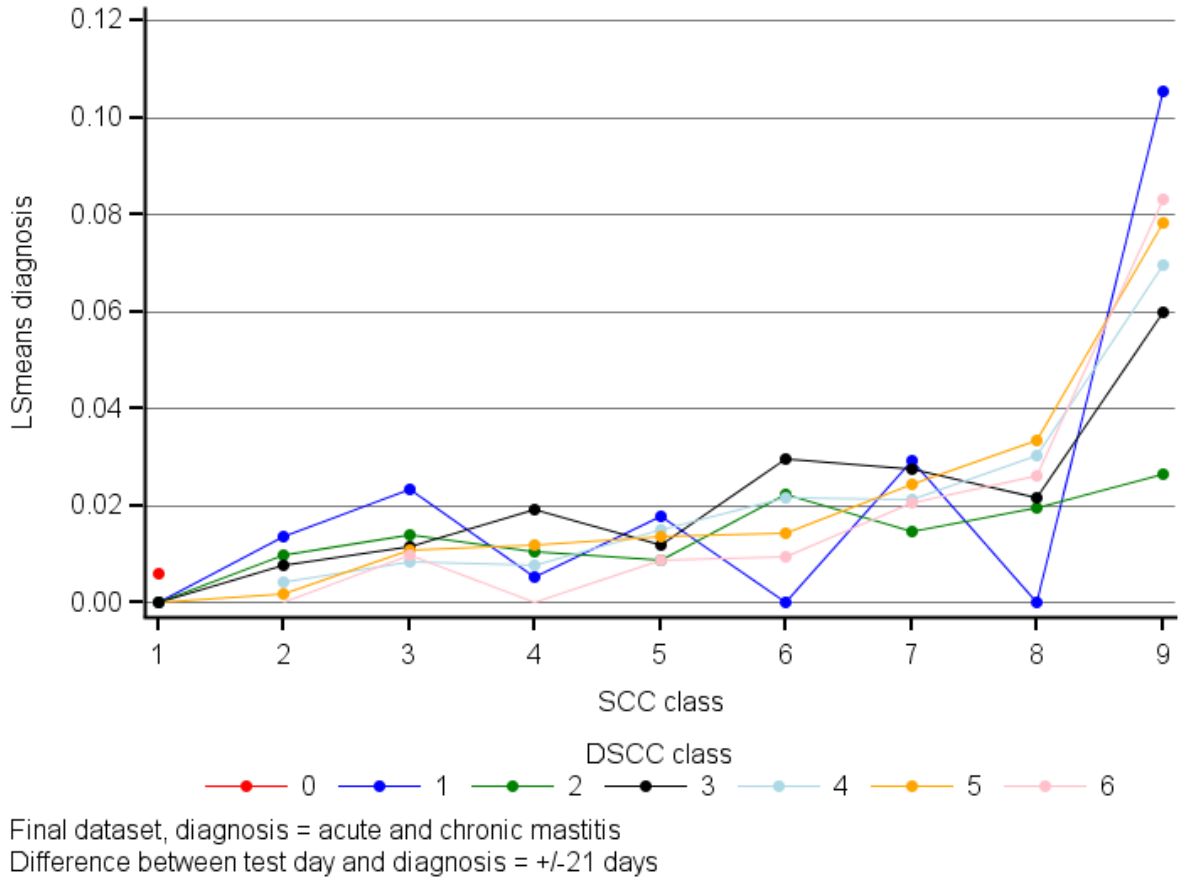


Figure 10: Association of SCC and DSCC with acute and chronic mastitis (Difference between test day and diagnosis = -/+ 21 days)

3.4 Comparison of UHG and mastitis diagnoses

The results of the comparison of UHG and diagnoses can be seen in Table 10. The highest sensitivity with 0.757 was achieved with UHG original and -/+ 7 days difference between diagnosis and test day. The lowest specificity with 0.607 was also achieved with this model. The highest specificity with 0.830 and simultaneously the lowest sensitivity with 0.519 was achieved with the UHG modified and -/+ 21 days difference between diagnosis and test day. The balanced accuracy increased with narrowing time windows and reached the maximum of 0.740 with UHG modified and -/+ 7 days difference between diagnosis and test day.

Table 10: Comparison of UHG models according to sensitivity, specificity and balanced accuracy in different time windows. UHG original: records with SCC < 200,000 cells/mL and DSCC < 65 % are considered as healthy; UHG modified: records with SCC < 200,000 cells/mL are considered healthy, no distinction according DSCC is made.

Diff. diagnosis and test day	Model	Sensitivity	Specificity	Balanced accuracy
-/+ 21	UHG original	0.653	0.609	0.631
	UHG modified	0.519	0.830	0.675
-/+ 14	UHG original	0.692	0.608	0.650
	UHG modified	0.570	0.829	0.700
-/+ 7	UHG original	0.757	0.607	0.682
	UHG modified	0.653	0.828	0.740

The highest value of the column is marked green and the lowest is marked red.

Figure 11 shows the SCC and DSCC values of test days with linked diagnoses (-/+ 21 days) and test days with no adjacent diagnoses classified as healthy. With the two marked threshold lines of SCC and DSCC the plotted data was classified into UHG as described in the legend of the figure. Figures of narrowed time windows (-/+ 14, -/+ 7) are listed in the appendix.

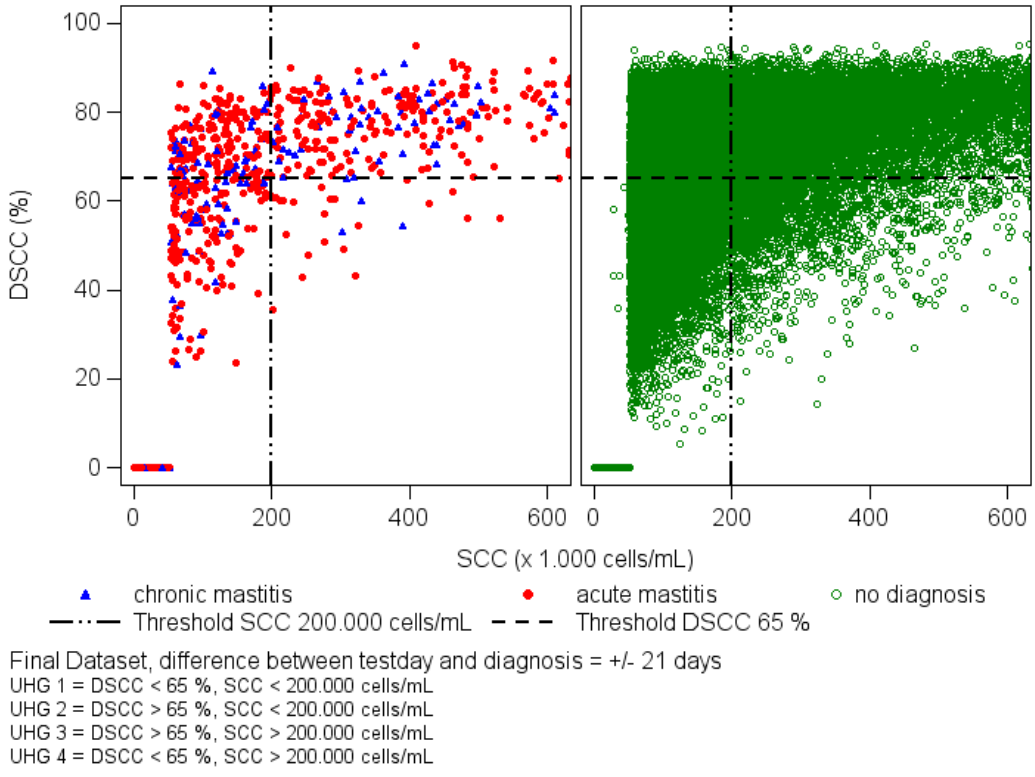


Figure 11: UHG model -/+ 21 days difference between diagnosis and test day

The investigation of thresholds showed that the amount of test days with SCC > 200,000 cells/mL increased by third of lactation and parity in both datasets (final dataset, healthy reference) as presented in Figure 12 and Figure 13. Table 11 shows that the increase by third of lactation and parity was significant. The proportion of test days with DSCC > 65 % decreased from 1st to 3rd lactation and then increased in both

datasets. The difference between 2nd and 3rd lactation was not significant in both datasets as shown in Table 11, all other differences were significant. Within thirds of lactation the proportion of DSCC > 65 % was decreasing significantly from first to last third in both datasets. The proportion of mastitis diagnoses with adjacent test day records in the range of 21 days before and after was increasing with parity. The difference between 1st and 2nd lactation was not significant but all others were. Mastitis diagnoses were decreasing significantly from first to last third of lactation. Considering the incidence of mastitis at any time of lactation, proportions increased significantly with parity and third of lactation.

Table 11: Differences of percentage of SCC > 200,000 cells/mL, DSCC > 65 % and diagnoses by parity and thirds of lactation with t-test (significance at p-value ≤ 0.05, values with same letter are not significantly different)

	Healthy reference		Final dataset			
Lact.	**Percentage SCC ≥ 200,000 cells/mL**	**Percentage DSCC ≥ 65 %**	**Percentage SCC ≥ 200,000 cells/mL**	**Percentage DSCC ≥ 65 %**	**Percentage diagnoses**	**Percentage diagnoses in lactation**
1	8.79 **a**	79.43 **a**	9.60 **a**	79.40 **a**	0.87 **a**	6.23 **a**
2	12.00 **b**	64.68 **b**	12.99 **b**	65.34 **b**	0.88 **ab**	7.50 **b**
3	16.67 **c**	64.77 **bc**	17.88 **c**	65.23 **bc**	1.30 **c**	9.59 **c**
4	20.12 **d**	67.11 **d**	21.53 **d**	67.53 **d**	1.62 **d**	11.95 **d**
5	26.78 **e**	72.60 **e**	28.61 **e**	73.18 **e**	2.04 **e**	14.29 **e**
Third of Lact.						
1	13.98 **a**	79.33 **a**	15.63 **a**	79.74 **a**	2.00 **a**	8.42 **a**
2	14.90 **b**	71.53 **b**	16.42 **b**	71.81 **b**	1.02 **b**	9.58 **b**
3	19.67 **c**	63.02 **c**	20.81 **c**	63.32 **c**	0.84 **c**	10.98 **c**

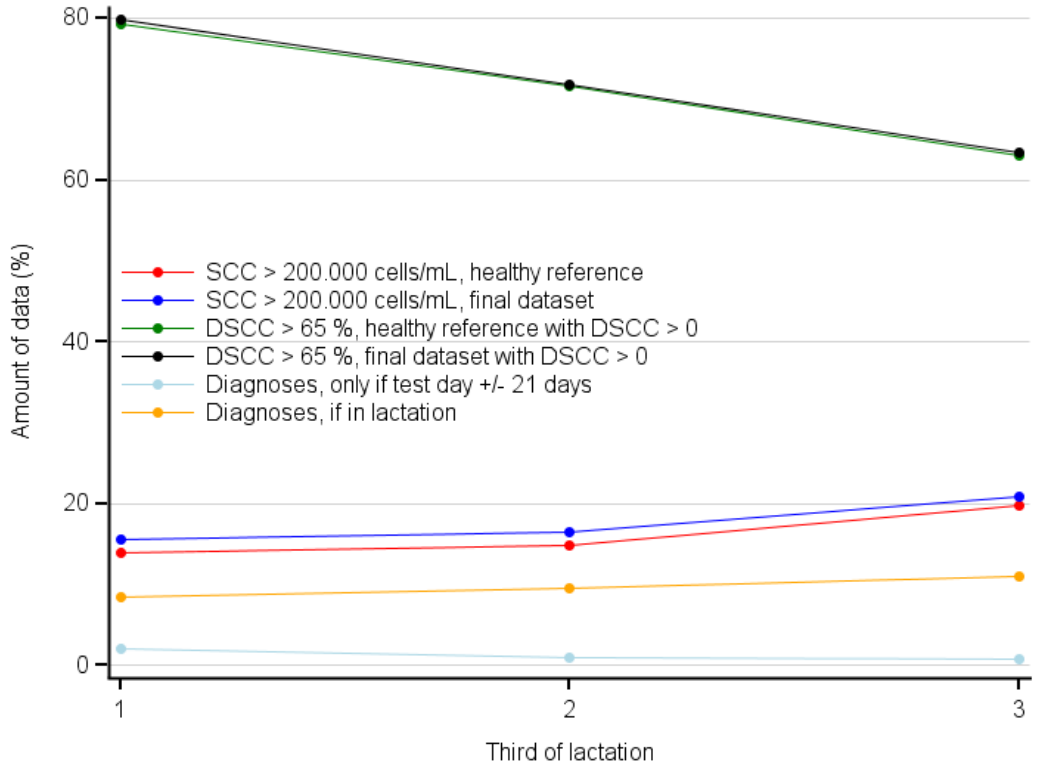


Figure 12: Number of test day data with SCC > 200,000 cells/mL, DSCC > 65 % and diagnoses by third of lactation

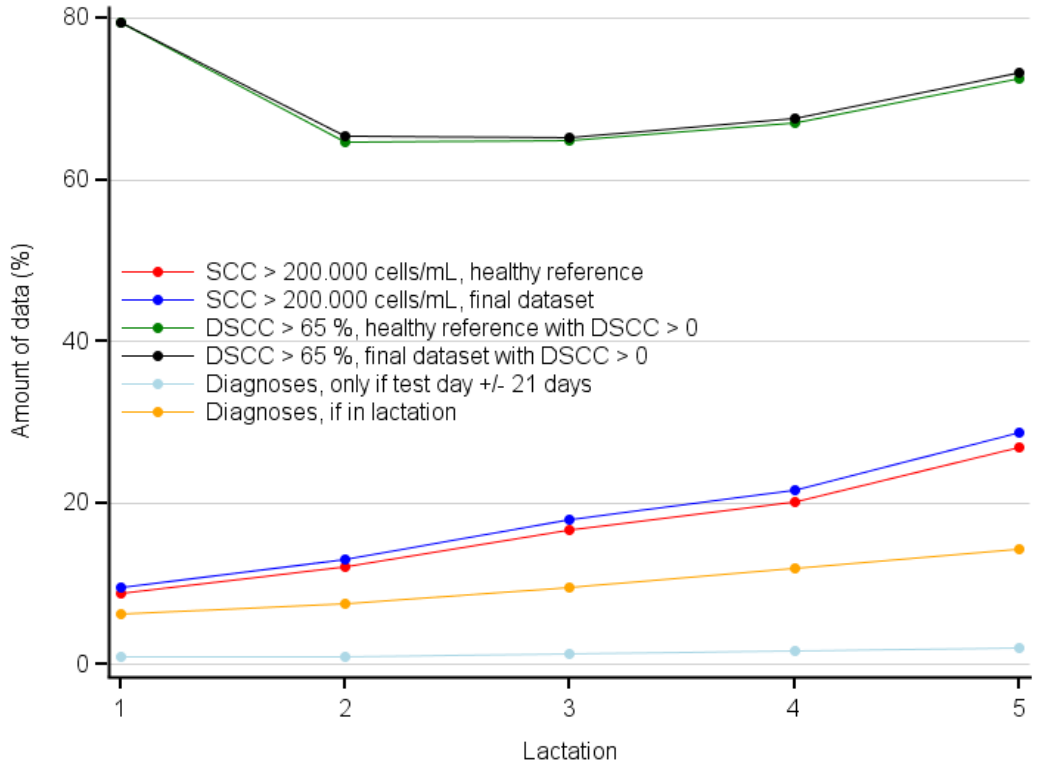


Figure 13: Number of test day data with SCC > 200,000 cells/mL, DSCC > 65 % and diagnoses by lactation

3.5 Results of PLSDA prediction model

The results of the PLSDA prediction model in calibration (train) dataset are shown in Table 12. The highest value for sensitivity (0.714) was reached with DSCC alone as predictor variable. Furthermore, also the lowest specificity (0.551) and lowest balanced accuracy (0.632) were reached with DSCC alone. The lowest sensitivity (0.594) was achieved with SCS + DSCC as predictor variable. Highest specificity (0.774) and highest balanced accuracy (0.715) could be reached with MIR + SCS + DSCC as predictor variables.

Table 12: Results of calibration (train) for different predictor variables (time window: -/+ 21 days), t-test for significance with Bonferroni correction (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Predictor variable	Sens.	Spec.	Bal. Acc.
MIR	0.649^{a}	0.694^{a}	0.671^{ad}
SCS	0.626^{b}	0.708^{a}	0.667^{a}
DSCC	0.714^{c}	0.551^{b}	0.632^{b}
MIR+SCS	0.657^{a}	0.765^{c}	0.711^{ce}
MIR+DSCC	0.679^{d}	0.703^{a}	0.691^{cd}
SCS+DSCC	0.594^{e}	0.761^{c}	0.677^{ad}
MIR+SCS+DSCC	0.656^{a}	0.774^{c}	0.715^{e}

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the column is marked green and the lowest is marked red.

Table 13 shows the results of validation (test) dataset for predicting mastitis without differentiation in acute and chronic cases. The results of the total dataset in validation (test) (-/+ 21 days) were additionally split into shorter time windows before/after diagnosis. Thereby the number of mastitis cases were changed but not the number of healthy cases. Thus, the specificity of different time windows did not change but the sensitivity and therefore also balanced accuracy and AUC changed.

Results for total dataset (- 21 to + 21) showed that the highest sensitivity (0.710) could be reached with DSCC but also the lowest specificity (0.564). Lowest sensitivity and yet highest specificity were achieved by SCS + DSCC. MIR + SCS showed the highest balanced accuracy (0.697) and AUC value (0.757). MIR as predictor value alone showed the lowest balanced accuracy (0.633) and AUC value (0.675).

When considering the split into shorter time windows, the highest sensitivity (0.808) was reached with SCS as predictor variable in time window + 0 to + 7. MIR + SCS reached the highest balanced accuracy (0.767) and AUC value (0.841) in time window + 0 to + 7.

Table 13: Results of validation (test) for different predictor variables for acute and chronic mastitis (time window: -/+ 21) t-test for significance with Bonferroni correction (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Predictor variable	Sens.	Spec.	Bal. Acc.	AUC	Time window
MIR	0.598^{ac}	0.668^{a}	0.633^{a}	0.675^{a}	**-21 to +21 (total)**
SCS	0.624^{ac}	0.725^{bd}	0.675^{bd}	0.737^{b}	
DSCC	0.710^{b}	**0.564^{c}**	0.637^{a}	0.690^{c}	
MIR+SCS	0.630^{ac}	0.764^{b}	0.697^{c}	0.757^{d}	
MIR+DSCC	0.642^{a}	0.691^{ad}	0.666^{d}	0.720^{e}	
SCS+DSCC	0.584^{c}	**0.776^{b}**	0.680^{b}	0.742^{b}	
MIR+SCS+DSCC	0.624^{ac}	0.763^{b}	0.694^{c}	0.755^{d}	
MIR	0.514^{a}	0.668^{a}	0.591^{a}	**0.614^{a}**	**-21 to -15**
SCS	0.539^{a}	0.725^{bd}	0.632^{bc}	0.703^{b}	
DSCC	0.644^{b}	**0.564^{c}**	0.604^{ac}	0.673^{b}	
MIR+SCS	0.544^{a}	0.764^{b}	0.654^{b}	0.704^{b}	
MIR+DSCC	0.582^{ab}	0.691^{ad}	0.637^{bc}	0.668^{b}	
SCS+DSCC	0.509^{a}	**0.776^{b}**	0.643^{bc}	0.708^{b}	
MIR+SCS+DSCC	0.542^{a}	0.763^{b}	0.653^{b}	0.708^{b}	
MIR	0.571^{a}	0.668^{a}	0.619^{a}	0.658^{a}	**-14 to -8**
SCS	0.632^{ab}	0.725^{bd}	0.679^{b}	0.738^{b}	
DSCC	0.707^{b}	**0.564^{c}**	0.635^{a}	0.702^{b}	
MIR+SCS	0.600^{a}	0.764^{b}	0.682^{b}	0.739^{b}	
MIR+DSCC	0.617^{ab}	0.691^{ad}	0.654^{ab}	0.702^{b}	
SCS+DSCC	0.591^{a}	**0.776^{b}**	0.684^{b}	0.742^{b}	
MIR+SCS+DSCC	0.611^{a}	0.763^{b}	0.687^{b}	0.739^{b}	

Predictor variable	Sens.	Spec.	Bal. Acc.	AUC	Time window
MIR	0.649^{a}	0.668^{a}	0.658^{a}	0.709^{a}	
SCS	0.738^{bc}	0.725^{bd}	0.731^{b}	0.785^{b}	
DSCC	0.782^{b}	**0.564^{c}**	0.673^{a}	0.735^{a}	
MIR+SCS	0.700^{ac}	0.764^{b}	0.732^{b}	0.800^{b}	**-7 to -0**
MIR+DSCC	0.710^{abc}	0.691^{ad}	0.701^{c}	0.772^{b}	
SCS+DSCC	0.702^{ac}	**0.776^{b}**	0.739^{b}	0.782^{b}	
MIR+SCS+DSCC	0.696^{ac}	0.763^{b}	0.730^{b}	0.792^{b}	
MIR	0.700^{a}	0.668^{a}	0.684^{a}	0.749^{a}	
SCS	**0.808^{b}**	0.725^{bd}	0.766^{b}	0.825^{bc}	
DSCC	0.803^{b}	**0.564^{c}**	0.684^{a}	0.743^{a}	
MIR+SCS	0.769^{bc}	0.764^{b}	**0.767^{b}**	**0.841^{b}**	**+0 to +7**
MIR+DSCC	0.763^{bc}	0.691^{ad}	0.727^{c}	0.801^{c}	
SCS+DSCC	0.752^{c}	**0.776^{b}**	0.764^{b}	0.833^{b}	
MIR+SCS+DSCC	0.764^{bc}	0.763^{b}	0.764^{b}	0.839^{b}	
MIR	0.585^{a}	0.668^{a}	0.626^{a}	0.663^{a}	
SCS	0.596^{a}	0.725^{bd}	0.661^{b}	0.737^{bd}	
DSCC	0.726^{b}	**0.564^{c}**	0.645^{ab}	0.678^{ac}	
MIR+SCS	0.628^{a}	0.764^{b}	0.696^{c}	0.748^{b}	**+8 to +14**
MIR+DSCC	0.632^{a}	0.691^{ad}	0.662^{b}	0.708^{cd}	
SCS+DSCC	0.557^{a}	**0.776^{b}**	0.667^{bd}	0.744^{b}	
MIR+SCS+DSCC	0.619^{a}	0.763^{b}	0.691^{cd}	0.750^{b}	

Predictor variable	Sens.	Spec.	Bal. Acc.	AUC	Time window
MIR	0.583^{a}	0.668^{a}	0.625^{ad}	0.669^{ad}	
SCS	0.467^{bc}	0.725^{bd}	0.596^{b}	0.650^{ab}	
DSCC	0.603^{a}	**0.564^{c}**	**0.584^{b}**	0.625^{b}	
MIR+SCS	0.550^{a}	0.764^{b}	0.657^{c}	0.719^{c}	**+15 to +21**
MIR+DSCC	0.574^{a}	0.691^{ad}	0.632^{cd}	0.686^{cd}	
SCS+DSCC	**0.426^{b}**	**0.776^{b}**	0.601^{ab}	0.657^{abd}	
MIR+SCS+DSCC	0.531^{ac}	0.763^{b}	0.647^{cd}	0.715^{c}	

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the time window is marked green and the lowest is marked red respectively.
The highest and lowest values of the column are marked in bold letters.

Figure 14 shows AUC values and balanced accuracies of prediction models in different time windows. MIR + SCS, MIR + SCS + DSCC, SCS + DSCC and SCS showed nearly the same development with an increase up to time window + 0 to + 7 and a decrease afterwards. All variations of predictor variables including MIR showed a flatter decrease between the last two time windows (+ 8 to + 14 and + 15 to + 21). The highest AUC level at any time window was always reached by MIR + SCS as predictor variables.

Furthermore, we also set up two models with a changed definition of mastitis and healthy cases (-/+ 14, -/+ 7) in calibration and validation. These plots are listed in the appendix.

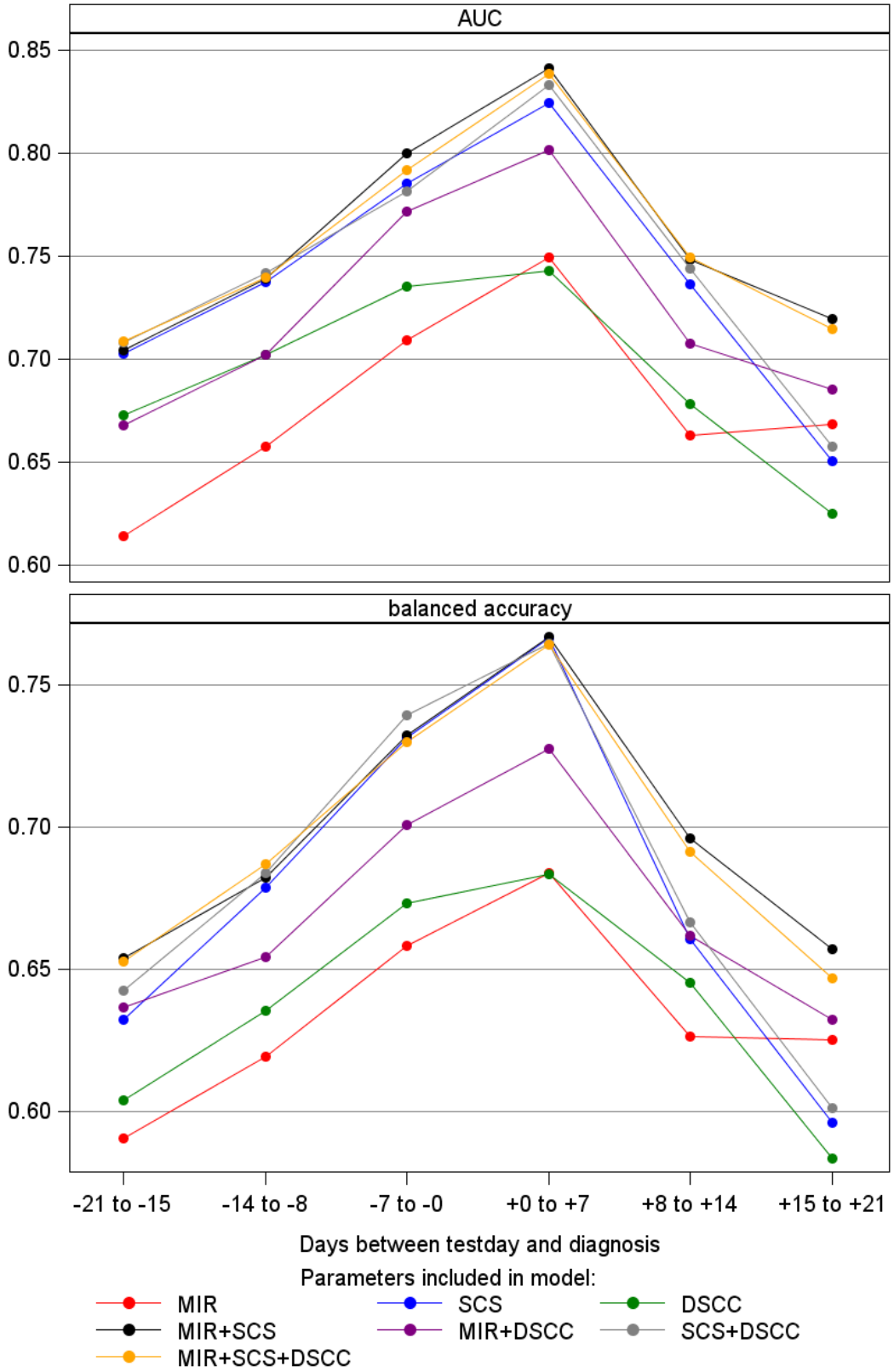


Figure 14: Prediction model acute and chronic mastitis (AUC, balanced accuracy, time window: -/+ 21 days)

The results of calibration of the two models with shorter time windows are shown in Table 14 and Table 15. Again, DSCC reached the highest sensitivities (0.755 and 0.794). It is also markable that with shorter time window in calibration, the sensitivity of DSCC as predictor variable increased. Yet, DSCC also showed the lowest specificity and balanced accuracy in both data sets (-/+ 14 and -/+ 7). Highest balanced accuracy was reached with MIR + SCS + DSCC in model with -/+ 14 days and with MIR + SCS in model -/+ 7 days.

Table 14: Results of calibration (train) for different predictor variables (time window: -/+ 14 days), t-test for significance with Bonferroni correction (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Predictor variable	Sens.	Spec.	Bal.Acc.
MIR	0.667^{ae}	0.729^{ad}	0.698^{a}
SCS	0.661^{a}	0.735^{a}	0.698^{a}
DSCC	0.755^{b}	0.540^{b}	0.647^{b}
MIR+SCS	0.691^{ce}	0.783^{c}	0.737^{c}
MIR+DSCC	0.711^{c}	0.707^{d}	0.709^{a}
SCS+DSCC	0.624^{d}	0.783^{c}	0.703^{a}
MIR+SCS+DSCC	0.688^{ce}	0.805^{c}	0.746^{c}

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the column is marked green and the lowest is marked red

Table 15: Results of calibration (train) for different predictor variables (time window: -/+ 7 days), t-test for significance with Bonferroni correction (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Predictor variable	Sens.	Spec.	Bal.Acc.
MIR	0.686^{a}	0.740^{a}	0.713^{a}
SCS	0.722^{acd}	0.772^{ad}	0.747^{b}
DSCC	0.794^{b}	0.563^{b}	0.678^{c}
MIR+SCS	0.726^{cd}	0.822^{c}	0.774^{d}
MIR+DSCC	0.753^{c}	0.744^{ad}	0.748^{b}
SCS+DSCC	0.706^{ad}	0.791^{cd}	0.748^{b}
MIR+SCS+DSCC	0.742^{cd}	0.789^{cd}	0.766^{bd}

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the column is marked green and the lowest is marked red

The results for validation in shorter time windows are shown in the appendix. The distribution of highest and lowest values was nearly the same as for the -/+ 21 time window. DSCC alone reached the highest sensitivity but lowest specificity. Predictor variables which additionally included MIR always reached the highest AUC level. Generally, the values of sensitivity, specificity, balanced accuracy and AUC were increasing with shorter time windows as was also shown for the calibration set.

4 Discussion

4.1 Development of SCC and DSCC over lactation stages and lactations

The results of analysis of SCC and DSCC show that SCC is increasing with parity and DIM in healthy reference. These results are similar to those in recent studies of Kirkeby et al. (2020) and Schwarz et al. (2020a). Our results for development of DSCC show a significant decrease by DIM and a decrease up to 2nd and 3rd lactation and a small increase up to 5th and higher lactations. These developments match closely with those of Kirkeby et al. (2020) but not with those of Schwarz et al. (2020a) who found an increase of DSCC by parity and DIM. This might be because the authors also took DSCC values with 0 % into account and we only took DSCC values > 0 into account. As mentioned above, DSCC can only be analysed if SCC is ≥ 52,000 cells/mL. However, SCC is increasing with DIM and parity and thus the amount of test day records with SCC ≥ 52,000 cells/mL. This leads to more DSCC values > 0 and furthermore to an increase of DSCC by DIM and parity.

Results of development of SCC and DSCC before acute mastitis diagnosis show a steady increase until day of diagnosis and a decrease afterwards. SCC and DSCC values for chronic mastitis diagnosis show an up and down until diagnosis and afterwards. Reason could be the low number of only 268 chronic mastitis diagnoses but also the fact, that a chronic mastitis infection can be prevalent over a longer period (weeks) (Winter and Zehle 2009) with fluctuating SCC and DSCC.

4.2 Association of SCC and DSCC with mastitis incidences

Results of association model show that there is a significant correlation of at least one effect (SCC, DSCC or interaction between SCC and DSCC) and mastitis incidences in all time windows and when splitting in acute and chronic mastitis. These results match with those of other studies (Kirkeby et al. 2020; Schwarz et al. 2020b) which also found an interaction between SCC, DSCC and mastitis incidence.

R-Square values were low and ranged from 0.004 (-/+ 7 days, chronic mastitis) to 0.022 (-/+ 21 days acute and chronic mastitis). These low values were calculated for the whole model. Probably one reason for these low values is that the date of mastitis diagnosis is not the date of adjacent milk recording, considered diseased in the current analyses. Also, mastitis is a 0/1 variable while SCC and DSCC are continuous variables.

4.3 Use of udder health groups for predicting mastitis

Results of UHG show that with narrowing time windows, sensitivity is increasing and specificity is staying at the same level in both models (original and modified) respectively. Reason for the higher sensitivity are the increasing SCC and DSCC values when narrowing the time window as seen in Figure 8 and Figure 9. Thus, the number of diagnoses above the thresholds is higher. The specificity is staying at the same level because the true healthy cases do not change remarkably. Whole model performance, expressed through balanced accuracy, increased with narrowing time window. Best balanced accuracy was achieved with -/+ 7 days and modified UHG model.

We used 200,000 cells/mL as SCC and 65 % as DSCC thresholds within this study. Highest sensitivity (0.757) was achieved with time window -/+ 7 days and original UHG

model. Highest specificity (0.830) was achieved with time window -/+ 21 days and modified UHG model. A comparable study (same thresholds) found a much higher sensitivity with 0.923 but a lower specificity with 0.660 (Schwarz et al. 2020b). In this study IMI was detected by bacterial culture (BC). Each cow in the study was sampled once a month and milk was analysed for SCC, DSCC, BC and other indicators. This might be the reason for higher sensitivity because the test day recording was linked with BC result from the same time. In our case we linked diagnoses with adjacent test days because we did not work with BC results. These diagnoses needed not necessarily be validated by a BC. Thus missed test days of cows carrying the pathogen but not being diagnosed. Schwarz et al. (2020b) categorised an udder as infected if a pathogen was detected.

According to Kirkeby et al. (2020) and Schwarz et al. (2020a), SCC and DSCC values are changing with DIM and parity. On these grounds we analysed the development of thresholds during DIM and parity. Results show that in final data set and healthy reference the number of cows with SCC > 200,000 cells/mL increased by DIM and parity significantly. Furthermore, the number of diagnoses from GMON also increased by DIM and parity but as shown in Figure 12 and Figure 13 the number of cows with SCC > 200,000 cells/mL increased disproportionately compared to diagnoses. Our analyses also show that the number of cows with DSCC > 65 % decreased by DIM in both datasets significantly. From 1st to 2nd lactation the number of cows with DSCC > 65 % decreased significantly and from 3rd to 5+ and more lactations increased significantly. These analyses show a changing of SCC and DSCC also in healthy cows by DIM and parity and could also be seen as a reason for low sensitivity values in UHG models. Zecconi et al. (2019) found different DSCC thresholds for parity and DIM, which confirm our results on changing DSCC values not only by IMI. It has to be mentioned that within this study mastitis was defined as a milk sample with > 200,000 cells/mL, which may not be highly reliable. Further research with verified diagnoses, as used in our study, or BC results will be necessary.

4.4 Comparison of prediction models

In this study we used the same model as used by Rienesl et al. (2019). The main aim within these calculations was to analyse whether the existing model with MIR + SCS could be improved with DSCC values. Our results have shown that models with DSCC alone have significantly the highest sensitivity in total dataset (-/+ 21 days) and in models with shorter time windows (-/+ 14, -/+7 days). On the other hand, DSCC also significantly showed lowest specificity in all three time windows. This led to comparatively low AUC levels, from 0.690 to 0.730. A prediction model with DSCC alone as variable would not be reliable. In comparison to UHG model, parity and breed were considered as fixed effects and had an impact on the prediction. DIM as an additional fixed effect could improve the model as shown in the study of Zecconi et al. (2019).

Figure 15 shows development of models with time window -/+ 21 days (calibration and validation) and splitting in shorter time windows. When expanding MIR + SCS model with DSCC, development of sensitivity was nearly the same and no improvement could be seen. Also, specificity was staying at same level and thus AUC could not be improved by adding DSCC. The statistical test for differences in means showed that there was no significant difference between the two models.

Highest model accuracies for each split time window (as marked in table Table 13) were largely not significant and thus will not be discussed.

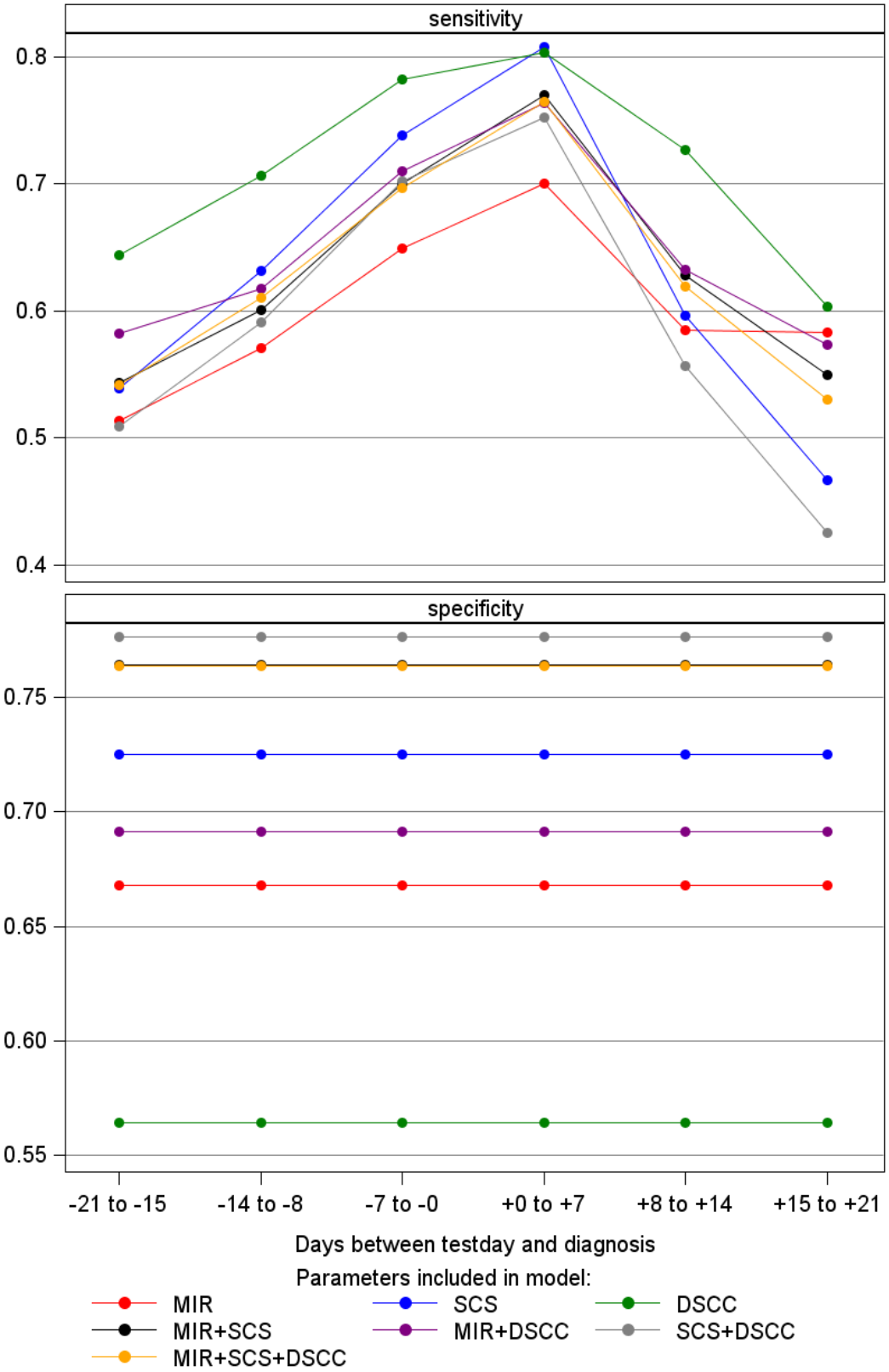


Figure 15: Prediction model acute and chronic mastitis (sensitivity, specificity, time window: -/+ 21 days)

When narrowing time windows in calibration and validation predictor variables of all models improved. Reason therefore could be the shorter interval between test day and diagnosis. As mentioned above Schwarz et al. (2020b) used test day record and BC from the same time and achieved results with a high sensitivity. These results could verify a model for shorter time windows, but such model would not be practicable. Most milk recording systems test cows 9 to 11 times per year and BC is not part of the standard milk examination. Therefore, we used the time window of 6 weeks (-/+ 21 days) for our main calculations to set up a practicable prediction model.

Further research which includes BC in addition to diagnoses and considers development of SCC and DSCC by DIM and parity will be needed to set up a reliable prediction model with DSCC as additional indicator.

4.5 Comparison of UHG and MIR data prediction model

Table 16: Comparison of UHG and MIR data prediction model

Predictor variable	Sens.	Spec.	Bal. Acc.	Time window
UHG original	0,653	0,609	**0,631**	**-/+ 21**
UHG modified	**0,519**	**0,830**	0,675	
MIR + SCS + DSCC	0,624	0,763	0,694	
MIR + SCS	0,630	0,764	0,697	
UHG original	0,692	0,608	0,650	**-/+ 14**
UHG modified	0,570	0,829	0,700	
MIR + SCS + DSCC	0,647	0,776	0,712	
MIR + SCS	0,651	0,768	0,710	
UHG original	**0,757**	**0,607**	0,682	**-/+ 7**
UHG modified	0,653	0,828	**0,740**	
MIR + SCS + DSCC	0,708	0,763	0,736	
MIR + SCS	0,708	0,713	0,710	

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the time window is marked green and the lowest is marked red respectively.
The highest and lowest values of the column are marked in bold letters.

Table 16 shows the comparison of UHG and MIR data prediction model regarding sensitivity, specificity and balanced accuracy. Overall model accuracies of both methods seem to be nearly identical. In shortest and widest time window UHG method showed highest and lowest model accuracies. Regarding balanced accuracy both methods showed similar values. In time window -/+ 21 days model accuracies of UHG show that there were also huge differences regarding sensitivity and specificity. Further investigation on thresholds of UHG model regarding parity and DIM will be needed.

5 Conclusion

This thesis evaluated the association of SCC and DSCC with mastitis diagnoses and the capability of SCC, DSCC, MIR spectra data and combination of the variables to predict mastitis via different methods. Results show that there is an association between SCC and DSCC with mastitis events. For using UHG as practicable management tool for farmers, further research on adapting thresholds for days in milk, parity and milk yield is needed. Results also have shown that the current mastitis prediction model (SCS + MIR) could not be improved by adding DSCC as an additional predictor variable. Also, further research will be needed to improve the prediction model by including DIM as additional fixed effect.

6 References

Aghamohammadi M., Haine D., Kelton D.F., Barkema H.W., Hogeveen H., Keefe G.P. and Dufour S. (2018). Herd-Level Mastitis-Associated Costs on Canadian Dairy Farms. Frontiers in veterinary science, 5, 100

Blowey R.W. and Edmondson P. (2010). Mastitis control in dairy herds. CABI, Cambridge, Mass, 2nd ed.

Bobbo T., Penasa M. and Cassandro M. (2019). Short communication: Genetic aspects of milk differential somatic cell count in Holstein cows: A preliminary analysis. Journal of dairy science, 102, 4275–4279

Bobbo T., Penasa M. and Cassandro M. (2020). Combining total and differential somatic cell count to better assess the association of udder health status with milk yield, composition and coagulation properties in cattle. Italian Journal of Animal Science, 19, 697–703

Bonfatti V., Di Martino G. and Carnier P. (2011). Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. Journal of dairy science, 94, 5776–5785

Bonfatti V., Ho P.N. and Pryce J.E. (2020). Usefulness of milk mid-infrared spectroscopy for predicting lameness score in dairy cows. Journal of dairy science, 103, 2534–2544

Bradley A. (2002). Bovine mastitis: an evolving disease. Veterinary journal (London, England : 1997), 164, 116–128

Burvenich C., Spiegeleer B. de, Winter P.M. and Zehle H.-H. (2009). Somatische Zellen und Zellzahlen. In: Praktischer Leitfaden Mastitis – Vorgehen beim Einzeltier und im Bestand ; 91 Tabellen (Eds. P.M. Winter and C. Burvenich), Parey, Stuttgart, 17–24

Cervinkova D., Vlkova H., Borodacova I., Makovcova J., Babak V., Lorencova A., Vrtkova I., Marosevic D. and Jaglic Z. (2013). Prevalence of mastitis pathogens in milk from clinically healthy cows. Veterinární Medicína, 58, 567–575

Damm M., Holm C., Blaabjerg M., Bro M.N. and Schwarz D. (2017). Differential somatic cell count-A novel method for routine mastitis screening in the frame of Dairy Herd Improvement testing programs. Journal of dairy science, 100, 4926–4940

Egger-Danner C. (2018). Eutergesundheit in Österreich – Status Quo. In: Tagunsband ZAR-Seminar 2018 – Brennpunkt Eutergesundheit und Antibiotikaeinsatz (Ed. Zentrale Arbeitsgemeinschaft österreichischer Rinderzüchter (ZAR)), Wien, 8–19

Fürst C., Dodenhoff J., Egger-Danner C., Emmerling R., Hamann H., Krogmeier D. and Schwarzenbacher H. (2019). Zuchtwertschätzung beim Rind Grundlagen, Methoden und Interpretationen

Garcia A. (2004). Contagious vs. Environmental Mastitis. Extension Extra, 126

Gengler N., Tijani A., Wiggans G.R. and Misztal I. (1999). Estimation of (Co)variance Function Coefficients for Test Day Yield with a Expectation-Maximization Restricted Maximum Likelihood Algorithm. Journal of dairy science, 82, 1849.e1–1849.e23

Grelet C., Bastin C., Gelé M., Davière J.-B., Johan M., Werner A., Reding R., Fernandez Pierna J.A., Colinet F.G., Dardenne P., Gengler N., Soyeurt H. and Dehareng F. (2016). Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. Journal of dairy science, 99, 4816–4825

Grelet C., Fernández Pierna J.A., Dardenne P., Baeten V. and Dehareng F. (2015). Standardization of milk mid-infrared spectra from a European dairy network. Journal of dairy science, 98, 2150–2160

Halasa T., Huijps K., Østerås O. and Hogeveen H. (2007). Economic effects of bovine mastitis and mastitis management: a review. The veterinary quarterly, 29, 18–31

Heikkilä A.-M., Nousiainen J.I. and Pyörälä S. (2012). Costs of clinical mastitis with special reference to premature culling. Journal of dairy science, 95, 139–150

Hogeveen H. and Winter P.M. (2009). Ökonomie der Mastitis. In: Praktischer Leitfaden Mastitis – Vorgehen beim Einzeltier und im Bestand ; 91 Tabellen (Eds. P.M. Winter and C. Burvenich), Parey, Stuttgart, 230–235

Huijps K., Lam T.J. and Hogeveen H. (2008). Costs of mastitis: facts and perception. The Journal of dairy research, 75, 113–120

International Dairy Federation - IDF (2013). Guidelines for the use and interpretation of bovine milk somatic cell counts (SCC) in the dairy industry. FIL/IDF, Brussels

International Organization for Standardization - ISO (2013). Milk and liquide milk products – Guidelines for the application of mid-infrared spectrometry, Switzerland

Kirkeby C., Toft N., Schwarz D., Farre M., Nielsen S.S., Zervens L., Hechinger S. and Halasa T. (2020). Differential somatic cell count as an additional indicator for intramammary infections in dairy cows. Journal of dairy science, 103, 1759–1775

Lainé A., Bastin C., Grelet C., Hammami H., Colinet F.G., Dale L.M., Gillon A., Vandenplas J., Dehareng F. and Gengler N. (2017). Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. Journal of dairy science, 100, 2863–2876

Lantz B. (2015). Machine learning with R Disciver how to build machine leaning algorithms, prepare data, and dig deep into data prediction techniques with R. Pact Publ, Birmingham, 2nd ed.

Leitner G., Shoshani E., Krifucks O., Chaffer M. and Saran A. (2000). Milk leucocyte population patterns in bovine udder infection of different aetiology. Journal of veterinary medicine. B, Infectious diseases and veterinary public health, 47, 581–589

Marchi M. de, Toffanin V., Cassandro M. and Penasa M. (2014). Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. Journal of dairy science, 97, 1171–1186

Nickerson S.C. (1989). Immunological Aspects of Mammary Involution. Journal of dairy science, 72, 1665–1678

Paape M.J., Wergin W.P., Guidry A.J. and Pearson R.E. (1979). Leukocytes–Second Line of Defense Against Invading Mastitis Pathogens. Journal of dairy science, 62, 135–153

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org

Rienesl L., Khayatzadeh N., Köck A., Dale L., Werner A., Grelet C., Gengler N., Auer F.-J., Egger-Danner C., Massart X. and Sölkner J. (2019). Mastitis Detection from Milk Mid-Infrared (MIR) Spectroscopy in Dairy Cows. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67, 1221–1226

Rienesl L., Pfeiffer P., Khayatzadeh N., Köck A., Dale L., Werner A., Grelet C., Gengler N., Auer F.-J., Egger-Danner C., Leblois J. and Sölkner J. (2020). Prediction of pregnancy state from milk mid-infrared (MIR) spectroscopy in dairy cows. Acta fytotechnica et zootechnica, 23, 224–232

Roos A.P.W. de, van den Bijgaart H.J.C.M., Hørlyk J. and Jong G. de (2007). Screening for subclinical ketosis in dairy cattle by Fourier transform infrared spectrometry. Journal of dairy science, 90, 1761–1766

Rutten M.J.M., Bovenhuis H., Hettinga K.A., van Valenberg H.J.F. and van Arendonk J.A.M. (2009). Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. Journal of dairy science, 92, 6202–6209

SAS Institute Inc. (2018). SAS Software 9.4, Cary, NC, USA

Schwarz D. (s.a.a). Differential Somatic Cell Count (DSCC) a rationale for the new parameter. Foss Inc., Hilleroed

Schwarz D. (s.a.b). Differential Somatic Cell Count with the Fossomatic 7 DC a novel parameter. Foss Inc., Hilleroed

Schwarz D. (2020). Good Separation for DSCC analysis with Fossomatic. verbal message

Schwarz D., Diesterbeck U.S., Failing K., König S., Brügemann K., Zschöck M., Wolter W. and Czerny C.-P. (2010). Somatic cell counts and bacteriological status in quarter foremilk samples of cows in Hesse, Germany--a longitudinal study. Journal of dairy science, 93, 5716–5728

Schwarz D., Diesterbeck U.S., König S., Brügemann K., Schlez K., Zschöck M., Wolter W. and Czerny C.-P. (2011). Flow cytometric differential cell counts in milk for the evaluation of inflammatory reactions in clinically healthy and subclinically infected bovine mammary glands. Journal of dairy science, 94, 5033–5044

Schwarz D., Kleinhans S., Reimann G., Stückler P., Reith F., Ilves K., Pedastsaar K., Yan L., Zhang Z., Valdivieso M., Barreal M.L. and Fouz R. (2020a). Investigation of dairy cow performance in different udder health groups defined based on a combination of somatic cell count and differential somatic cell count. Preventive veterinary medicine, 183, 105123

Schwarz D., Lipkens Z., Piepers S. and Vliegher S. de (2019). Investigation of differential somatic cell count as a potential new supplementary indicator to somatic cell count for identification of intramammary infection in dairy cows at the end of the lactation period. Preventive veterinary medicine, 172, 104803

Schwarz D., Santschi D.E., Durocher J. and Lefebvre D.M. (2020b). Evaluation of the new differential somatic cell count parameter as a rapid and inexpensive

supplementary tool for udder health management through regular milk recording. Preventive veterinary medicine, 181, 105079

Sharma N., Singh N.K. and Bhadwal M.S. (2011). Relationship of Somatic Cell Count and Mastitis: An Overview. Asian-Australasian Journal of Animal Sciences, 24, 429–438

Sordillo L.M. and Nickerson S.C. (1988). Morphologic changes in the bovine mammary gland during involution and lactogenesis. American journal of veterinary research, 49, 1112–1120

Soyeurt H., Bastin C., Colinet F.G., Arnould V.M.-R., Berry D.P., Wall E., Dehareng F., Nguyen H.N., Dardenne P., Schefers J., Vandenplas J., Weigel K., Coffey M., Théron L., Detilleux J., Reding E., Gengler N. and McParland S. (2012). Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. Animal : an international journal of animal bioscience, 6, 1830–1838

Soyeurt H., Bruwier D., Romnee J.-M., Gengler N., Bertozzi C., Veselko D. and Dardenne P. (2009). Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. Journal of dairy science, 92, 2444–2454

Soyeurt H., Colinet F.G., Arnould V.M.-R., Dardenne P., Bertozzi C., Renaville R., Portetelle D. and Gengler N. (2007). Genetic variability of lactoferrin content estimated by mid-infrared spectrometry in bovine milk. Journal of dairy science, 90, 4443–4450

Toledo-Alvarado H., Vazquez A.I., Los Campos G. de, Tempelman R.J., Bittante G. and Cecchinato A. (2018). Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. Journal of dairy science, 101, 2496–2505

Vanlierde A., Vanrobays M.-L., Dehareng F., Froidmont E., Soyeurt H., McParland S., Lewis E., Deighton M.H., Grandl F., Kreuzer M., Gredler B., Dardenne P. and Gengler N. (2015). Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. Journal of dairy science, 98, 5740–5747

Winter P.M. and Zehle H.-H. (2009). Klinik der Mastitisformen. In: Praktischer Leitfaden Mastitis – Vorgehen beim Einzeltier und im Bestand ; 91 Tabellen (Eds. P.M. Winter and C. Burvenich), Parey, Stuttgart, 95–101

Zecconi A., Vairani D., Cipolla M., Rizzi N. and Zanini L. (2019). Assessment of subclinical mastitis diagnostic accuracy by differential cell count in individual cow milk. Italian Journal of Animal Science, 18, 460–465

7 Appendix

Association of SCC and DSCC with mastitis incidences

Table 17: Association of SCC and DSCC with acute mastitis (Difference between test day and diagnosis = -/+ 21 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	21.366496	0.410894	39.27	<0.0001
Error	106,973	1,119.212139	0.010463		
Corrected Total	107,025	1,140.578635			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.018733	949.4653	0.102287	0.010773		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	1.70840918	0.21355115	20.41	<0.0001
DSCC	6	0.02872992	0.00478832	0.46	0.8400
DSCC*SCC	38	0.71539747	0.01882625	1.80	0.0018

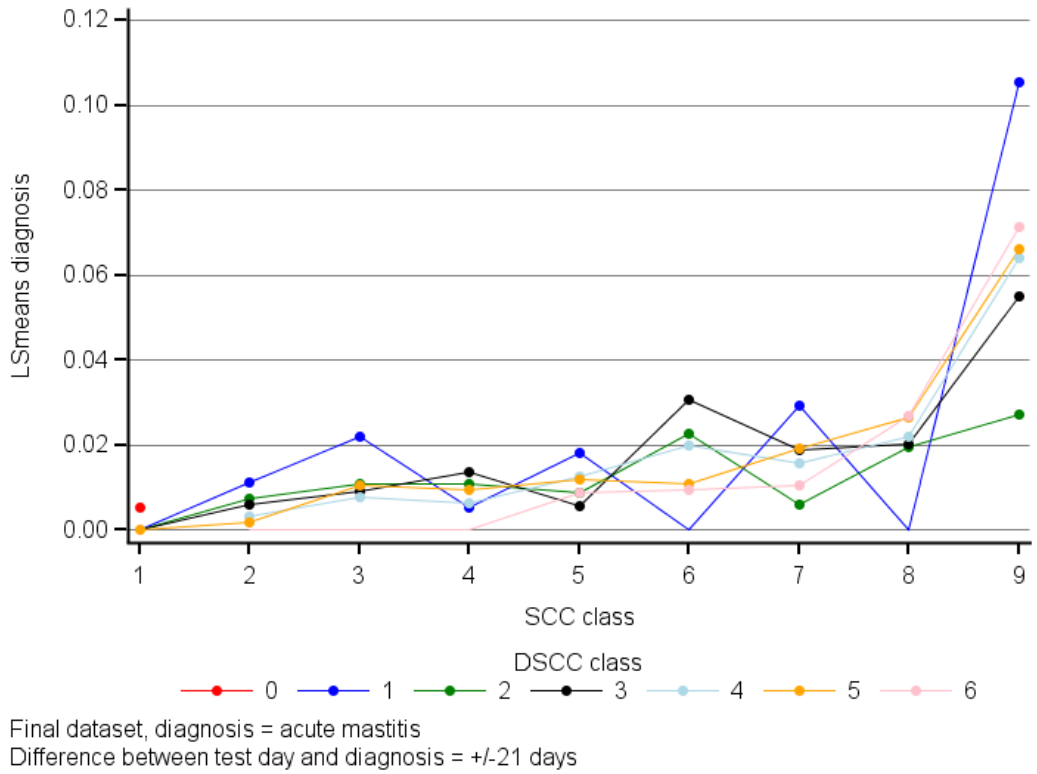


Figure 16: Association of SCC and DSCC with acute mastitis (Difference between test day and diagnosis = -/+ 21 days)

Table 18: Association of SCC and DSCC with chronic mastitis (Difference between test day and diagnosis = -/+ 21 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	1.5946255	0.0306659	11.61	<0.0001
Error	100,605	265.6918296	0.0026409		
Corrected Total	100,657	267.2864551			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.005966	1,930.158	0.051390	0.002662		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	0.04849302	0.00606163	2.30	0.0187
DSCC	6	0.00985251	0.00164208	0.62	0.7131
DSCC*SCC	38	0.24155636	0.00635675	2.41	<0.0001

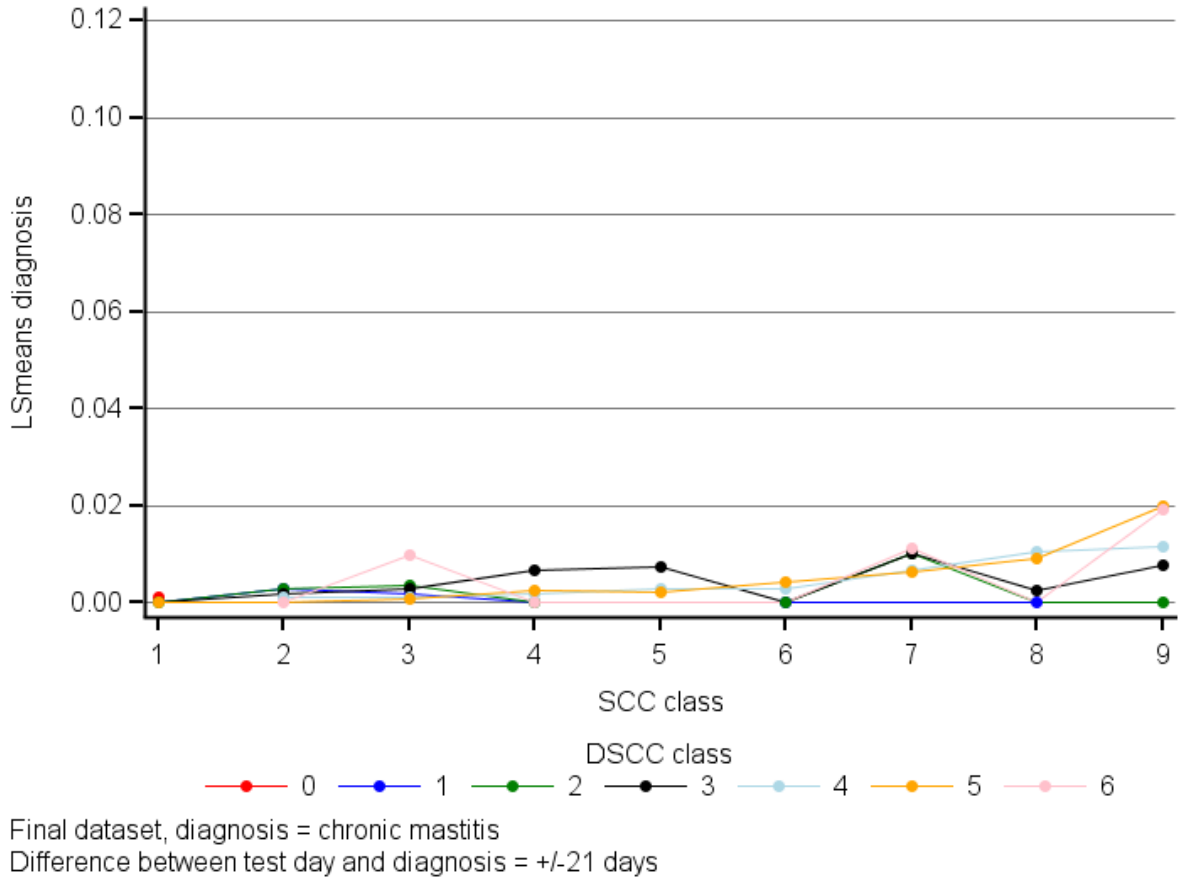


Figure 17: Association of SCC and DSCC with chronic mastitis (Difference between test day and diagnosis = -/+ 21 days)

Table 19: Association of SCC and DSCC with acute and chronic mastitis (Difference between test day and diagnosis = -/+ 14 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	19.4494138	0.3740272	43.48	<0.0001
Error	109,031	937.9961144	0.0086030		
Corrected Total	109,083	957.4455282			

R-Square	Coeff Var	Root MSE	diagnosis Mean
0.020314	1,047.392	0.092752	0.008856

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SCC	8	1.49290554	0.18661319	21.69	<0.0001
DSCC	6	0.03398739	0.00566457	0.66	0.6834
DSCC*SCC	38	0.56916623	0.01497806	1.74	0.0031

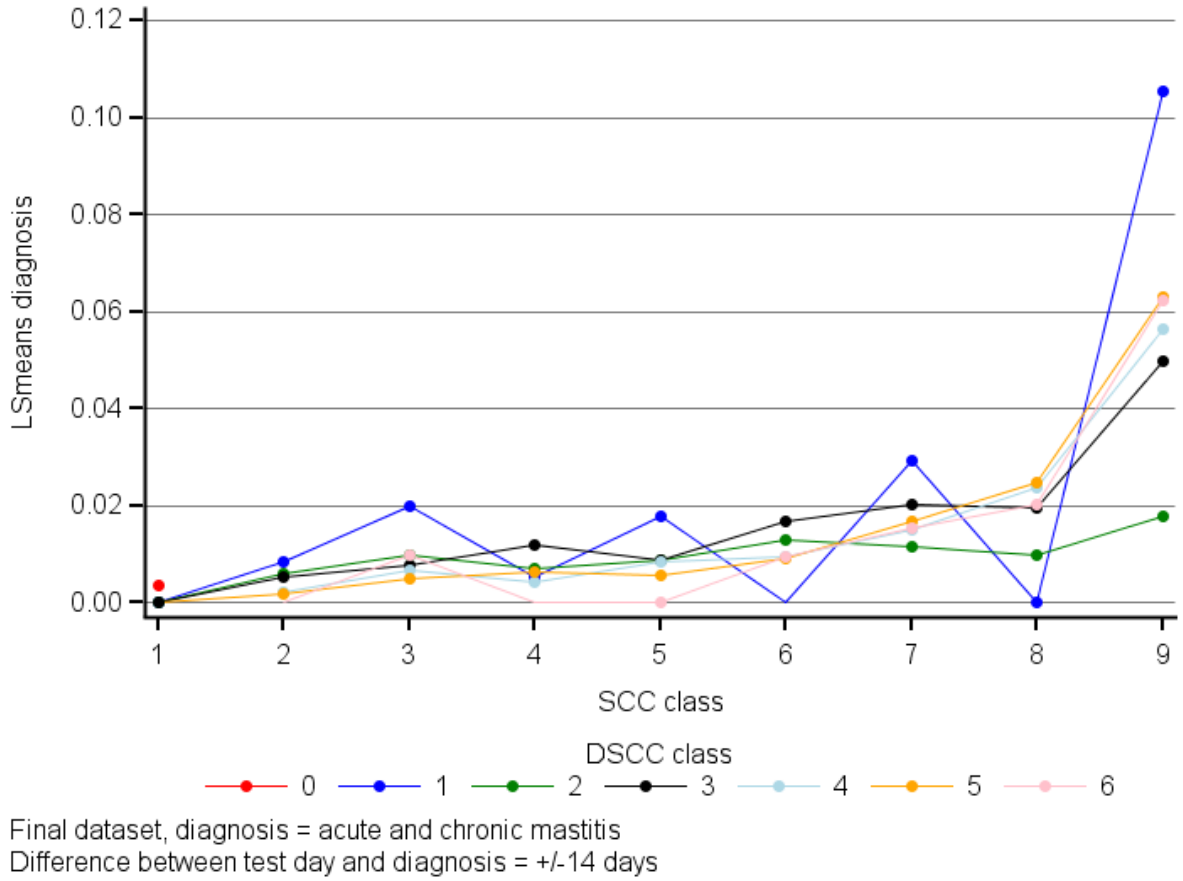


Figure 18: Association of SCC and DSCC with acute and chronic mastitis (Difference between test day and diagnosis = -/+ 14 days)

Table 20: Association of SCC and DSCC with acute mastitis (Difference between test day and diagnosis = -/+ 14 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	13.6367737	0.2622456	36.64	<0.0001
Error	106,973	765.6055132	0.0071570		
Corrected Total	107,025	779.2422869			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.017500	1,153.414	0.084599	0.007335		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	1.25577589	0.15697199	21.93	<0.0001
DSCC	6	0.03805521	0.00634254	0.89	0.5038
DSCC*SCC	38	0.39964251	0.01051691	1.47	0.0310

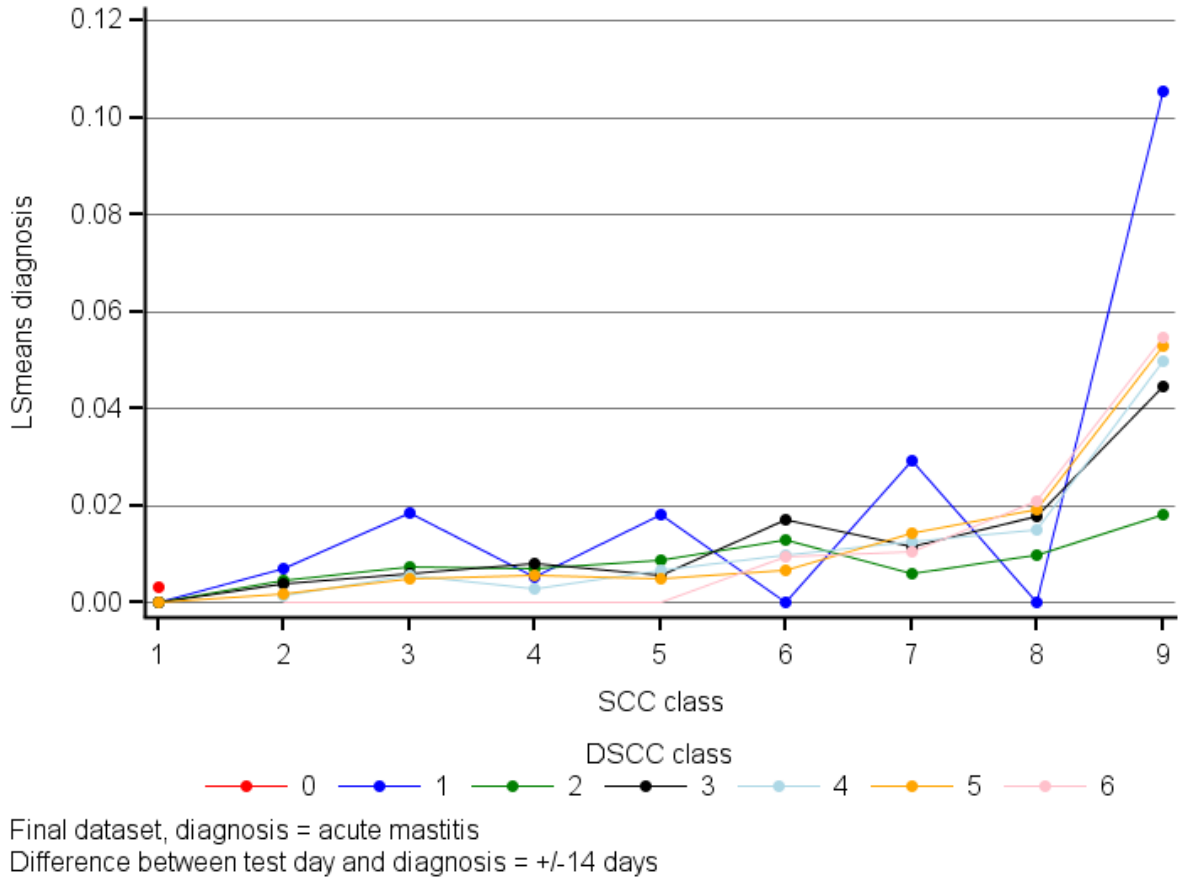


Figure 19: Association of SCC and DSCC with acute mastitis (Difference between test day and diagnosis = -/+ 14 days)

Table 21: Association of SCC and DSCC with chronic mastitis (Difference between test day and diagnosis = -/+ 14 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	1.0421997	0.0200423	11.22	<0.0001
Error	100,605	179.6323319	0.0017855		
Corrected Total	100,657	180.6745316			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.005768	2,349.915	0.042255	0.001798		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	0.03342247	0.00417781	2.34	0.0164
DSCC	6	0.00732553	0.00122092	0.68	0.6628
DSCC*SCC	38	0.15065671	0.00396465	2.22	<0.0001

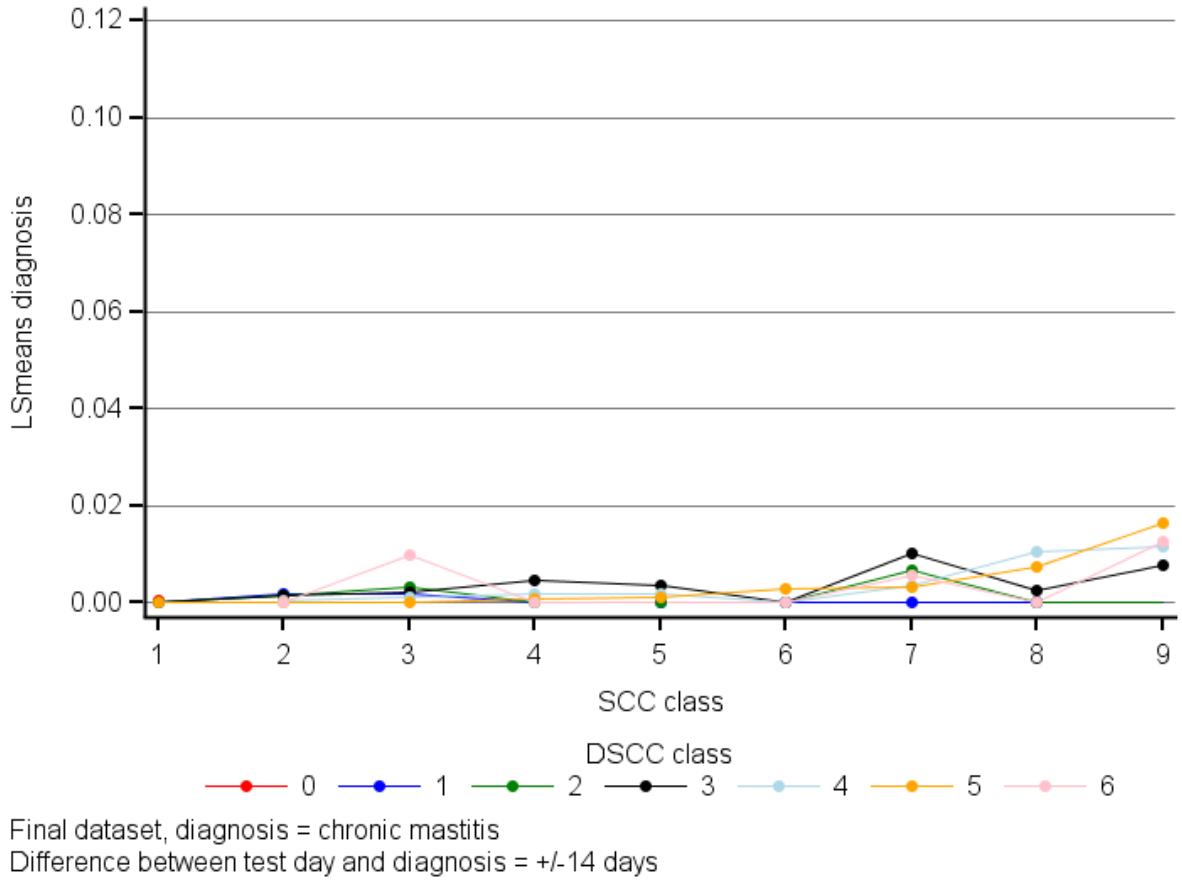


Figure 20: Association of SCC and DSCC with chronic mastitis (Difference between test day and diagnosis = -/+ 14 days)

Table 22: Association of SCC and DSCC with acute and chronic mastitis (Difference between test day and diagnosis = -/+ 7 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	6.5911514	0.1267529	31.37	<0.0001
Error	109,031	440.5607224	0.0040407		
Corrected Total	109,083	447.1518738			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.014740	1,544.339	0.063566	0.004116		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	0.49763684	0.06220460	15.39	<0.0001
DSCC	6	0.02551363	0.00425227	1.05	0.3889
DSCC*SCC	38	0.29684248	0.00781164	1.93	0.0005

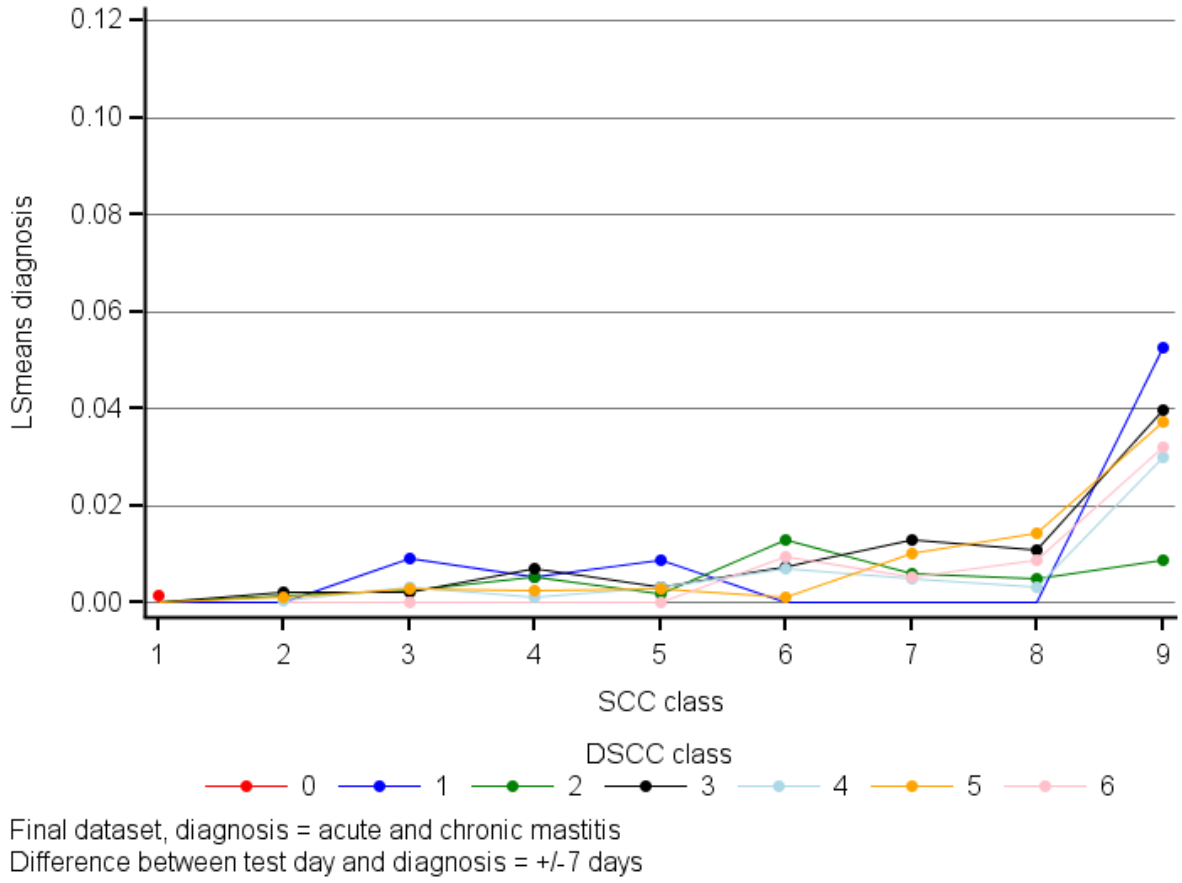


Figure 21: Association of SCC and DSCC with acute and chronic mastitis (Difference between test day and diagnosis = -/+ 7 days)

Table 23: Association of SCC and DSCC with acute mastitis (Difference between test day and diagnosis = -/+ 7 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	4.6497052	0.0894174	26.64	<0.0001
Error	106,973	359.1055038	0.0033570		
Corrected Total	107,025	363.7552090			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.012783	1,698.910	0.057939	0.003410		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	0.41261239	0.05157655	15.36	<0.0001
DSCC	6	0.01842713	0.00307119	0.91	0.4828
DSCC*SCC	38	0.24121577	0.00634778	1.89	0.0007

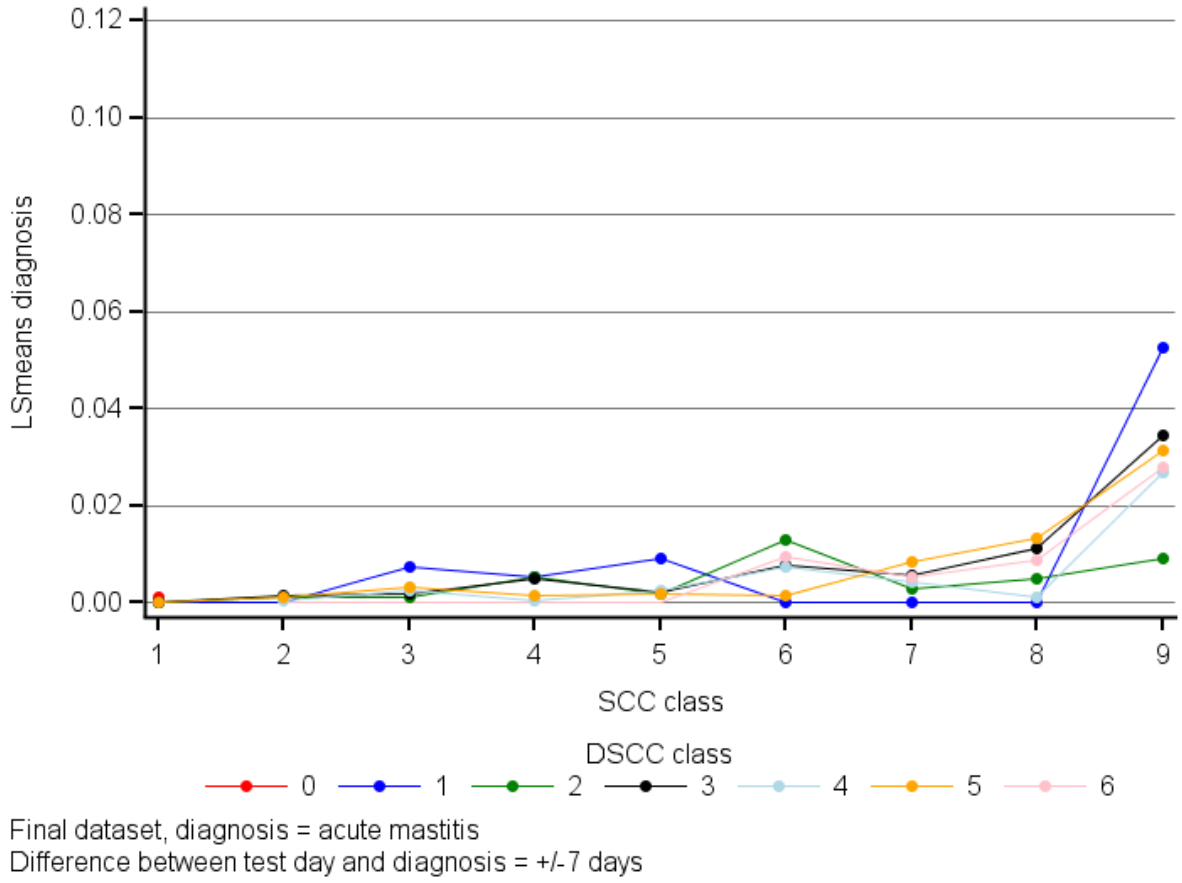


Figure 22: Association of SCC and DSCC with acute mastitis (Difference between test day and diagnosis = -/+ 7 days)

Table 24: Association of SCC and DSCC with chronic mastitis (Difference between test day and diagnosis = -/+ 7 days)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	52	0.34006300	0.00653967	7.87	<0.0001
Error	100,605	83.58983825	0.00083087		
Corrected Total	100,657	83.92990125			
R-Square	**Coeff Var**	**Root MSE**	**diagnosis Mean**		
0.004052	3,454.108	0.028825	0.000835		
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
SCC	8	0.01157498	0.00144687	1.74	0.0836
DSCC	6	0.00199234	0.00033206	0.40	0.8797
DSCC*SCC	38	0.04341330	0.00114246	1.38	0.0618

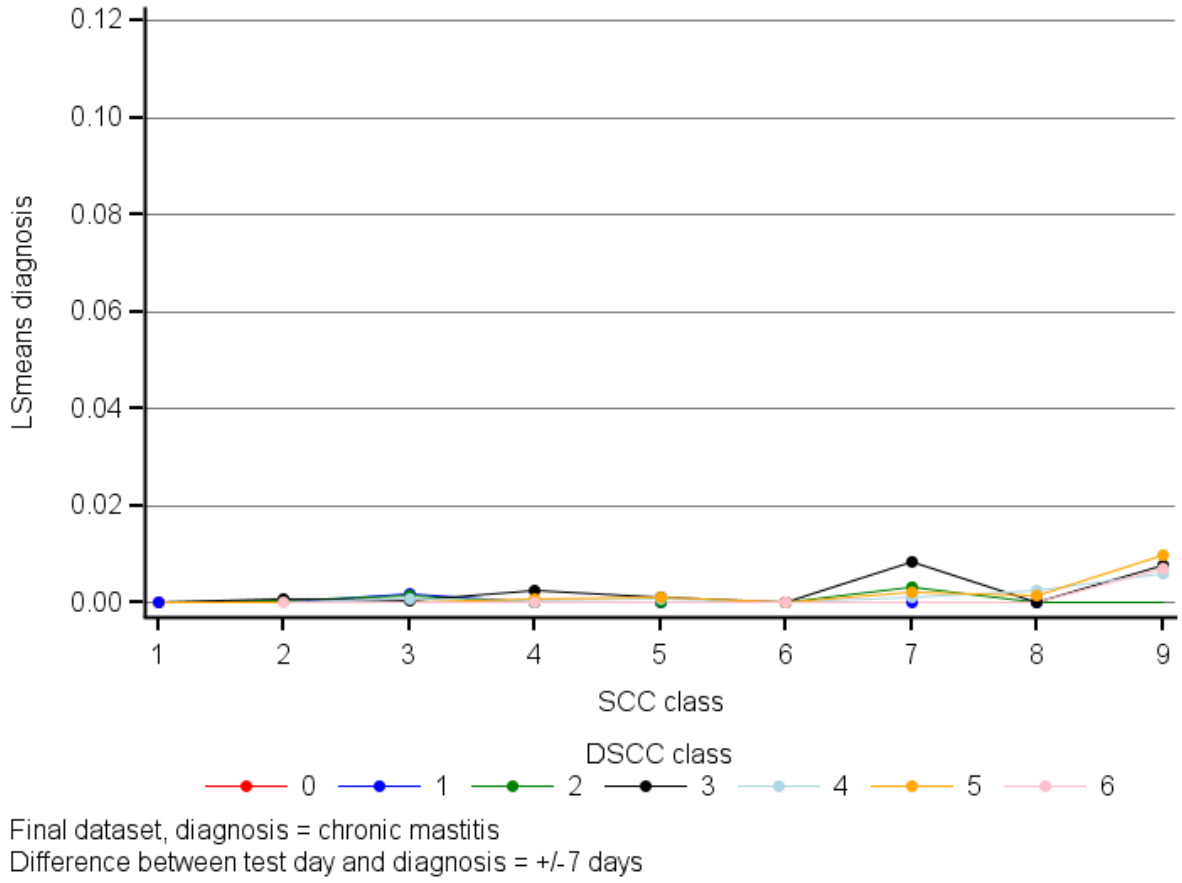
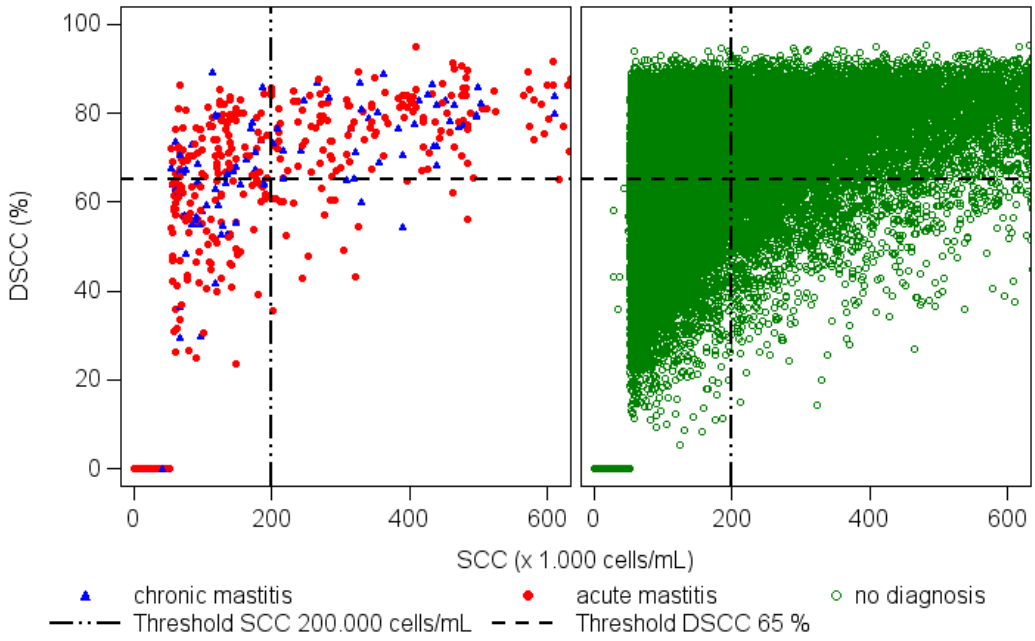


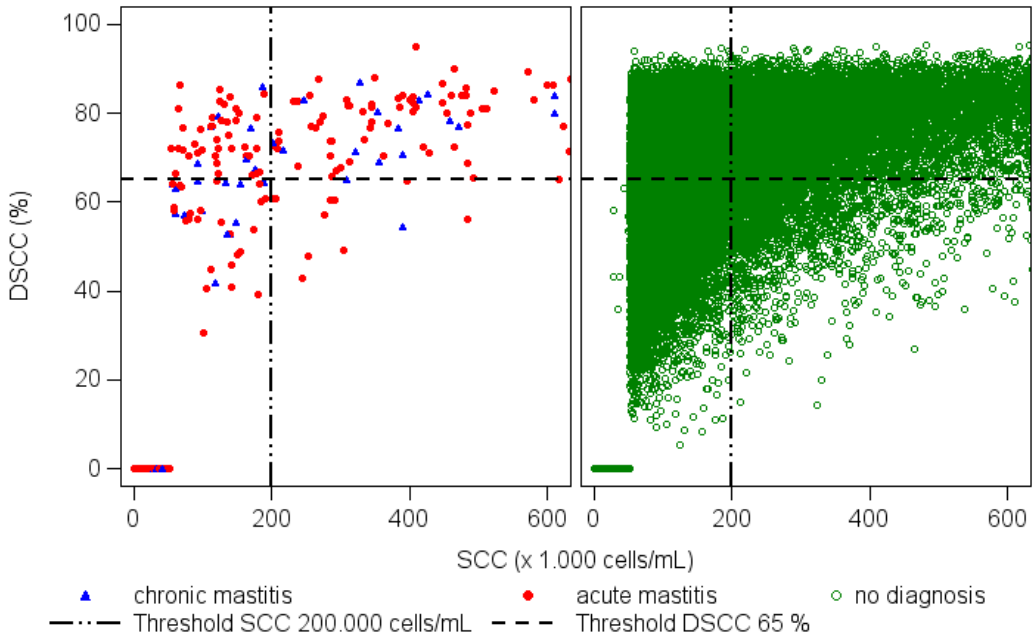
Figure 23: Association of SCC and DSCC with chronic mastitis (Difference between test day and diagnosis = -/+ 7 days)

UHG models



Final Dataset, difference between testday and diagnosis = +/- 14 days
UHG 1 = DSCC < 65 %, SCC < 200.000 cells/mL
UHG 2 = DSCC > 65 %, SCC < 200.000 cells/mL
UHG 3 = DSCC > 65 %, SCC > 200.000 cells/mL
UHG 4 = DSCC < 65 %, SCC > 200.000 cells/mL

Figure 24: UHG model -/+ 14 days difference between diagnosis and test day



Final Dataset, difference between testday and diagnosis = +/- 7 days
UHG 1 = DSCC < 65 %, SCC < 200.000 cells/mL
UHG 2 = DSCC > 65 %, SCC < 200.000 cells/mL
UHG 3 = DSCC > 65 %, SCC > 200.000 cells/mL
UHG 4 = DSCC < 65 %, SCC > 200.000 cells/mL

Figure 25: UHG model -/+ 7 days difference between diagnosis and test day

Prediction model

Table 25: Results of validation (test) for different predictor variables for acute and chronic mastitis (time window: -/+ 14), t-test for significance with Bonferroni correction (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Predictor variable	Sens.	Spec.	Bal.Acc.	AUC	Time window
MIR	0.589[a]	0.681[a]	0.635[a]	0.686[a]	
SCS	0.662[ab]	0.745[bd]	0.703[b]	0.764[b]	
DSCC	0.760[c]	**0.550**[c]	0.655[c]	0.709[c]	
MIR+SCS	0.651[ab]	0.768[b]	0.710[b]	0.773[b]	**-14 to +14 (total)**
MIR+DSCC	0.667[b]	0.692[ad]	0.680[d]	0.741[d]	
SCS+DSCC	0.623[ab]	**0.792**[b]	0.708[b]	0.767[b]	
MIR+SCS+DSCC	0.647[ab]	0.776[b]	0.712[b]	0.772[b]	
MIR	**0.536**	0.681	**0.608**	**0.650**	
SCS	0.626	0.745	0.685	0.737	
DSCC	0.724	**0.550**	0.637	0.703	
MIR+SCS	0.609	0.768	0.688	0.740	**-14 to -8**
MIR+DSCC	0.619	0.692	0.655	0.710	
SCS+DSCC	0.583	**0.792**	0.687	0.740	
MIR+SCS+DSCC	0.604	0.776	0.690	0.744	
MIR	0.610	0.681	0.645	0.698	
SCS	0.717	0.745	0.731	0.785	
DSCC	0.800	**0.550**	0.675	0.744	
MIR+SCS	0.687	0.768	0.728	0.790	**-7 to -0**
MIR+DSCC	0.696	0.692	0.694	0.761	
SCS+DSCC	0.684	**0.792**	0.738	0.779	
MIR+SCS+DSCC	0.667	0.776	0.722	0.781	
MIR	0.677	0.681	0.679	0.742	
SCS	0.775	0.745	**0.760**	0.815	
DSCC	**0.812**	**0.550**	0.681	0.739	
MIR+SCS	0.740	0.768	0.754	**0.828**	**+0 to +7**
MIR+DSCC	0.768	0.692	0.730	0.797	
SCS+DSCC	0.715	**0.792**	0.754	0.821	
MIR+SCS+DSCC	0.732	0.776	0.754	0.825	
MIR	0.569	0.681	0.625	0.676	**+8 to +14**

SCS	0.570	0.745	0.657	0.733
DSCC	0.720	**0.550**	0.635	0.669
MIR+SCS	0.596	0.768	0.682	0.750
MIR+DSCC	0.622	0.692	0.657	0.716
SCS+DSCC	0.539	**0.792**	0.666	0.741
MIR+SCS+DSCC	0.605	0.776	0.691	0.753

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the time window is marked green and the lowest is marked red respectively
The highest and lowest values of the column are marked in bold letters

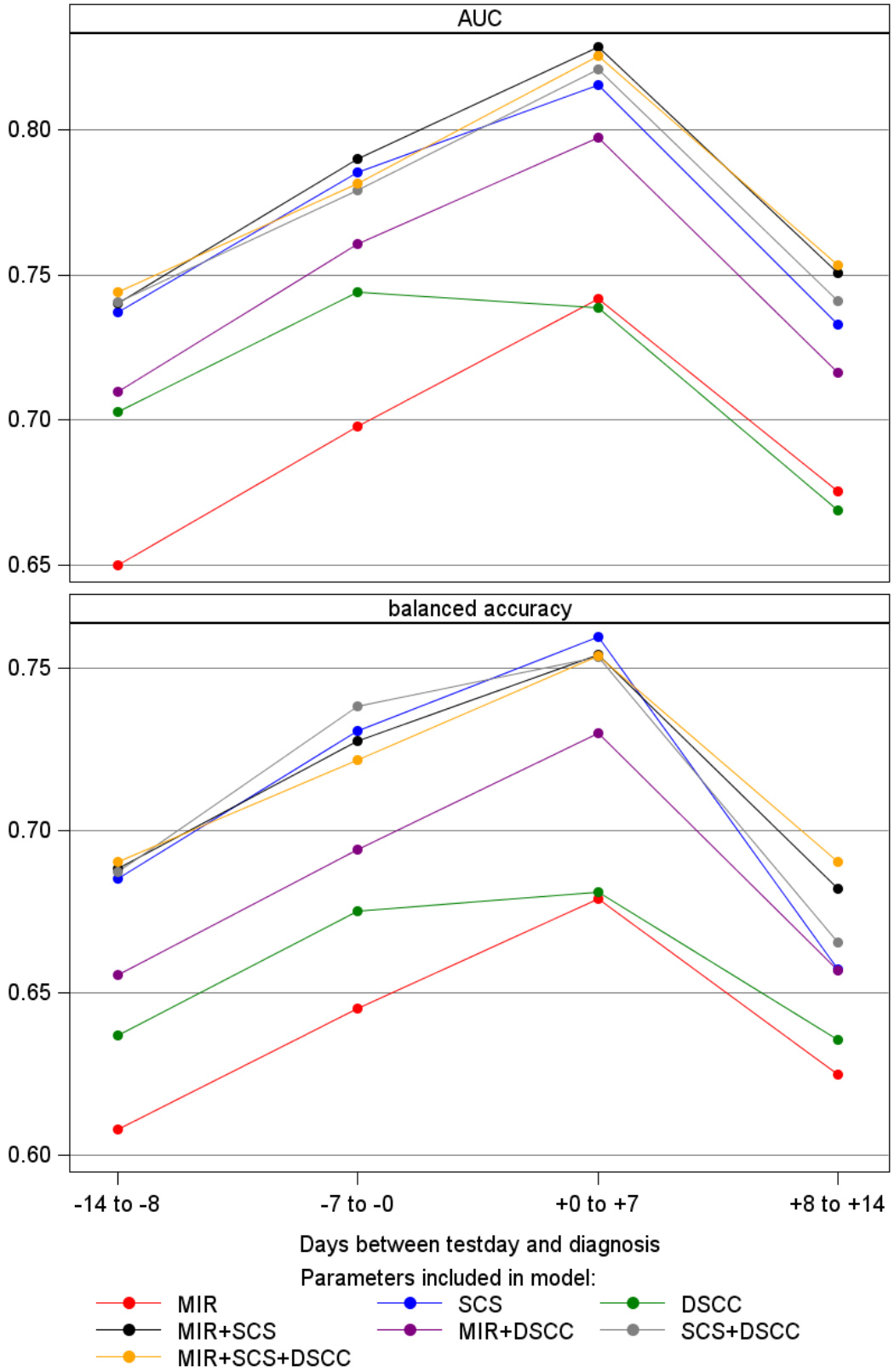


Figure 26: Prediction model acute and chronic mastitis (AUC, balanced accuracy, time window: -/+ 14 days)

Table 26: Results of validation (test) for different predictor variables for acute and chronic mastitis (time window: -/+ 7), t-test for significance with Bonferroni correction (significance at p-value ≤ 0.05, values with same letter are not significantly different)

Predictor variable	Sens.	Spec.	Bal.Acc.	AUC	Time window
MIR	0.617[a]	0.695[a]	0.656[a]	0.706[a]	**-7 to +7 (total)**
SCS	0.720[b]	0.769[bd]	0.744[b]	0.794[b]	
DSCC	0.792[c]	**0.562[c]**	0.677[a]	0.730[c]	
MIR+SCS	0.687[b]	**0.801[b]**	0.744[b]	0.804[b]	
MIR+DSCC	0.708[b]	0.713[ad]	0.710[c]	0.770[d]	
SCS+DSCC	0.697[b]	0.777[bd]	0.737[b]	0.793[bd]	
MIR+SCS+DSCC	0.708[b]	0.763[bd]	0.736[b]	0.797[b]	
MIR	**0.593**	0.695	**0.644**	**0.688**	**-7 to -0**
SCS	0.705	0.769	0.737	0.776	
DSCC	0.786	**0.562**	0.674	0.731	
MIR+SCS	0.667	**0.801**	0.734	0.784	
MIR+DSCC	0.687	0.713	0.700	0.755	
SCS+DSCC	0.683	0.777	0.730	0.773	
MIR+SCS+DSCC	0.683	0.763	0.723	0.779	
MIR	0.664	0.695	0.679	0.736	**+0 to +7**
SCS	0.754	0.769	0.761	0.824	
DSCC	**0.809**	**0.562**	0.686	0.738	
MIR+SCS	0.730	**0.801**	**0.765**	**0.836**	
MIR+DSCC	0.754	0.713	0.734	0.799	
SCS+DSCC	0.721	0.777	0.749	0.826	
MIR+SCS+DSCC	0.752	0.763	0.757	0.828	

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy
The highest value of the time window is marked green and the lowest is marked red respectively.
The highest and lowest values of the column are marked in bold letters.

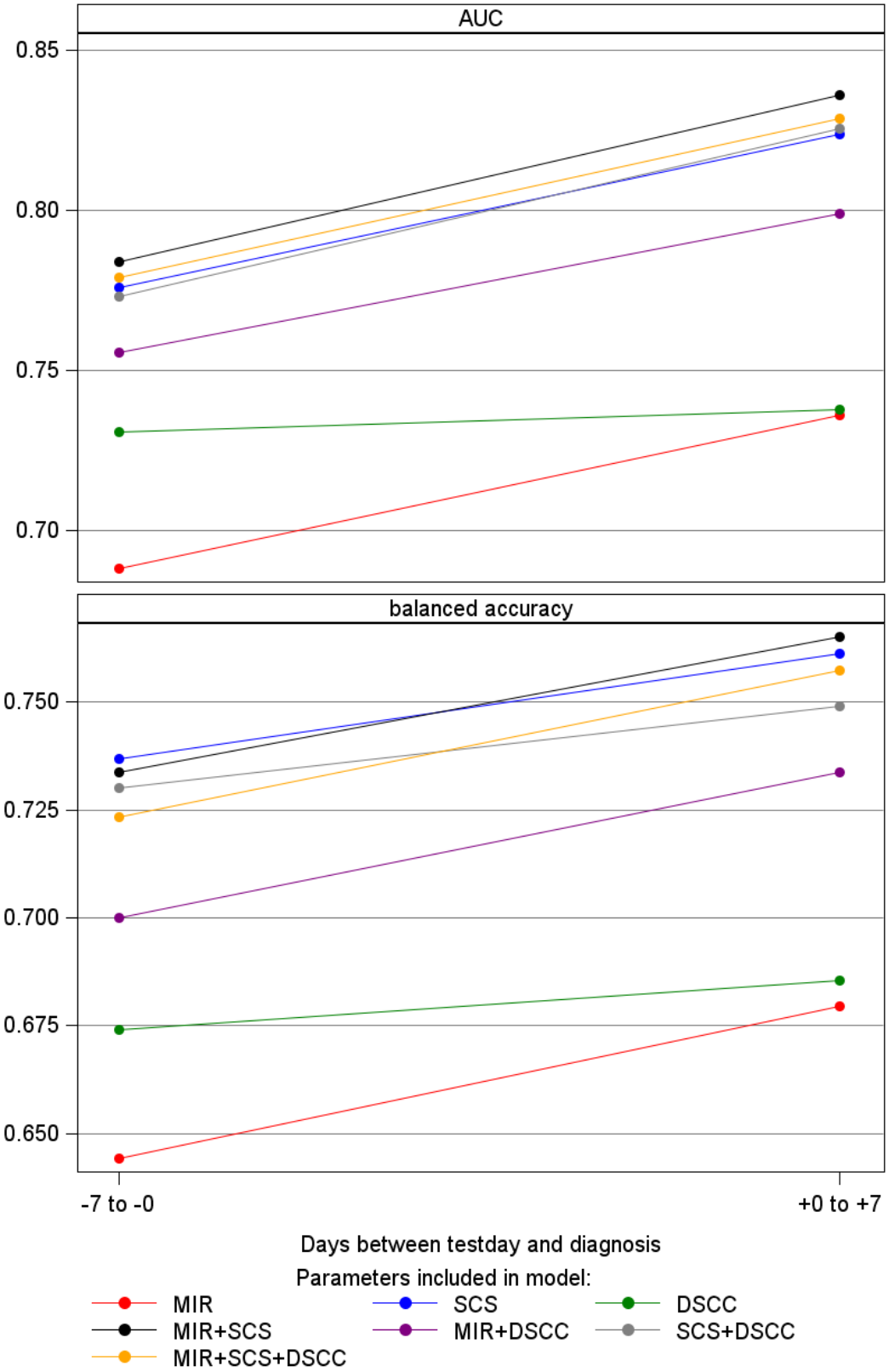


Figure 27: Prediction model acute and chronic mastitis (AUC, balanced accuracy, time window: -/+ 7 days)