Genetic Evaluation of Susceptibility to- and Recoverability from Mastitis in Dairy Cows

Berihu Gebremedhin Welderufael
Centre for Veterinary Medicine and Animal Science (VHC)
Department of Animal Breeding and Genetics
Uppsala

and
Centre for Quantitative Genetics and Genomics
Department of Molecular Biology and Genetics
Aarhus University, Tjele, Denmark

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2017

Acta Universitatis agriculturae Sueciae
2017:71

Cover: posterior density of heritabilities for susceptibility to- and recoverability from mastitis (from this thesis, photo by the author)

The research compiled in this thesis is from a collaborative PhD project work between SLU and Aarhus University, Denmark. The PhD project was part of the Erasmus Mundus joint doctorate programme 'EGS-ABG: European Graduate School in Animal Breeding and Genetics '.

ISSN 1652-6880
ISBN (print version) 978-91-7760-032-9
ISBN (electronic version) 978-91-7760-033-6

Print: SLU Service/Repro, Uppsala 2017

Genetic Evaluation of Susceptibility to- and Recoverability from Mastitis in Dairy Cows

Abstract

Mastitis has been the focus of many dairy cattle research projects over the last decades. However, in the genetic evaluation of udder health, only susceptibility to mastitis has been considered, leaving aside the other aspect of the disease - the recoverability. The aim of this thesis was to improve the genetic evaluation of udder health by introducing a new approach and models that can make use of the information contained in both directions: susceptibility to- and recoverability from mastitis.

In paper I, extensive simulation analyses were performed to develop a bivariate model for joint genetic evaluations of susceptibility to- and recoverability from mastitis. In paper II, the bivariate model with an added time function as well as several systematic effects was applied to real data to estimate genetic parameters in Danish Holstein cows. In paper III, genome-wide association studies were conducted to identify associated single-nucleotide polymorphisms and thereof candidate genes. In paper IV, a dynamic health classification, which takes severity of possible infection into account was introduced to further improve the genetic evaluation of mastitis.

Findings in paper I demonstrated that both traits can be modelled jointly and genetic parameters could be correctly reproduced. In paper II, we detected presence of genetic variation that resulted to heritability (ranging from 0.06 to 0.08) of similar size for both traits. The between trait genetic correlation was -0.83. Despite the strong negative genetic correlation, association signals in paper III did not overlap, suggesting that the traits are at least partially regulated by different genes. Complexity of the traits was manifested with the absence of strong association signals. In paper IV, considerable genetic variation was detected for cows' presence in health classes defined for longer periods, whereas the variations in health classes defined for short-term and sudden changes (e.g., acute) were mostly attributed to environmental factors. Although susceptibility to- and recoverability from mastitis are strongly negatively correlated, recoverability which is as heritable as susceptibility could be considered a new trait for selection. Evaluating and modelling the ability of animals to overcome infection could be of specific benefit in situations of high disease incidence.

Keywords: bivariate model, dairy cow, genetic evaluation, mastitis, recoverability, susceptibility

Author's address: Berihu Gebremedhin Welderufael, SLU, Department of Animal Breeding and Genetics,
P.O. Box 7023, 750 07 Uppsala, Sweden
E-mail: Berihu.Welderufael@ slu.se

Dedication

My mother, Dayet!

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Welderufael, B. G., D. J. de Koning, L. L. G. Janss, J. Franzén, and W. F. Fikse (2016). Simultaneous genetic evaluation of simulated mastitis susceptibility and recovery ability using a bivariate threshold sire model. *Acta Agriculturae Scandinavica, Section A — Animal Science* 66(3):125-134.

II Welderufael, B. G., L. L. G. Janss, D. J. de Koning, L. P. Sorensen, P. Lovendahl, and W. F. Fikse (2017). Bivariate threshold models for genetic evaluation of susceptibility to and ability to recover from mastitis in Danish Holstein cows. *Journal of Dairy Science* 100(6), 4706-4720.

III Welderufael, B.G., P. Løvendahl, D.J. de Koning, L.L.G. Janss and W.F. Fikse. Genome-wide association study for susceptibility to- and recovery from mastitis in Danish Holstein cows (submitted to journal: *Frontiers in Genetics* section *Livestock Genomics*).

IV Sørensen, L.P., M. Bjerring, B.G. Welderufael, D.J. de Koning, L.L.G. Janss and P. Løvendahl. Dynamic udder health classification scheme in dairy cows using online cell count (manuscript).

Papers I-II are reproduced with the permission of the publishers.

The contribution of B.G.W to the papers included in this thesis was as follows:

I Performed the final statistical data analyses and wrote the first draft of the paper. Analysed intermediate results with significant inputs from Freddy Fikse in data simulation. Wrote the final manuscript with regular input from the co-authors.

II Performed the final statistical analyses with significant inputs from Freddy Fikse in data editing. Was responsible for writing and completing the manuscript with regular input from the co-authors.

III Performed the statistical analyses. Wrote the manuscript with regular input from the co-authors.

IV Derived new traits from the model with multiple disease classes. Performed the genetic analyses and contributed to the writing of the manuscript.

Abbreviations

AI-REML	Average information restricted maximum likelihood
AMS	Automatic milking systems (machines)
CM	Clinical mastitis
DH	Transitions from diseased state to healthy state
DIC	Deviance information criterion
DIM	Days in milk
EBV	Estimated breeding value
EMR	Elevated mastitis risk
GEBV	Genomic estimated breeding value
GWAS	Genome-wide association study
HD	Transitions from healthy state to diseased state
HPD	Highest posterior density
MCMC	Markov chain Monte Carlo
OCC	Online somatic cell count
OR	Odds ratio
PE	Permanent environmental
QTL	Quantitative trait loci
SCC	Somatic cell count
SCM	Subclinical mastitis
SCS	Somatic cell score
SD	Standard deviation
SE	Standard error
SNP	Single-nucleotide polymorphism
TBV	True breeding value
US	United States of America
VMS	Voluntary milking system
WIM	Weeks in milk

1 Introduction

Modern dairy farms own high milk producing cows shaped by centuries of selection and breeding for improved milk production. Advances in many scientific disciplines, mainly quantitative genetics, have made a significant contribution to the recorded genetic improvement of dairy cattle productivity (Gianola & Rosa, 2015; Brotherstone & Goddard, 2005). The increase in milk yield of cows in the UK by 1200kg per lactation in 12 years (1988 ─ 2000) is a typical example of a successful application of quantitative genetics in dairy cattle (Brotherstone & Goddard, 2005). Similar trend of genetic response in production traits and milk yield per cow has been observed in all major dairying countries (Brotherstone & Goddard, 2005). However, it did not take long for scientists to discover unwanted consequences of selection for improved milk production. Declining fertility, declining longevity and deteriorating udder health are among the most unwanted consequences of selection for increased milk yield (Oltenacu & Broom, 2010). Consequently, selection indices have evolved to a more balanced breeding goal including fertility, longevity, conformation traits, and udder health (Miglior *et al.*, 2005).

Mastitis is the most relevant trait regarding udder health (Campbell & Marshall, 2016; Govignon-Gion *et al.*, 2016). The word mastitis comes from the Greek words *mastos* and *itis* meaning “breast” and “inflammation”, respectively (Campbell & Marshall, 2016). Bovine mastitis, therefore, refers to inflammation of the mammary gland of the cow (Campbell & Marshall, 2016). Currently, mastitis is one of the most important traits in the breeding goals for dairy cattle. This is because mastitis is a common disease causing large economic losses and problems in quality of milk and dairy products worldwide (Hogeveen *et al.*, 2011; Halasa *et al.*, 2007). The need for breeding dairy cows resistant to mastitis is, therefore, drawing the attention of modern dairy cattle breeding programmes worldwide. Developing better models for genetic evaluation of udder health is a prerequisite for the implementation of a

successful breeding programme to breed cows with improved resistance to mastitis.

In the genetic evaluation of udder health, susceptibility to mastitis has been the focus of many dairy cattle research projects (Govignon-Gion *et al.*, 2012; Carlén, 2008; de Haas, 2003), disregarding the other aspect of the disease - the recoverability. Because mastitis is a common and unavoidable problem, evaluation of the ability of cows to recover should, therefore, be of interest. Introduction of the recovery aspect as a new trait in the analyses is expected to enhance the genetic evaluation of udder health by capturing more information from both directions of the disease, contracting and recovery (Franzén *et al.*, 2012). In the studies compiled in this thesis, a great emphasis was given to introduce and include the recovery aspect in the modelling and analyses of mastitis to improve udder health in dairy cattle.

By analysing the changes that each cow exhibits during lactation, we have developed models that can make better use of available information in disease data with mastitis focus. Models that can make use of additional and time dependent information are valuable to the dairy cattle systems as well as for researchers and producers working on milk production systems. The additional and time dependent information will result in higher accuracy of selection by accounting all variations a cow may exhibits over the course of a lactation, in both directions of the disease (susceptibility and recoverability). This thesis has introduced a novel approach to include and model the recoverability in the genetic analyses of mastitis to improve udder health. In situations with high disease incidences, like mastitis which can have detrimental effects, evaluating the capacity of cows for recovery could be of specific benefit to the dairy industry.

2 General background

2.1 Mastitis in dairy cattle

Mastitis in dairy cattle (bovine mastitis) is defined as an inflammatory reaction of the mammary gland to various pathogenic infections (International Dairy Federation, 1987). It is a very common disease in the dairy industry. Mastitis can be either clinical or subclinical (Viguier *et al.*, 2009). Clinical mastitis (CM) is accompanied by visible symptoms such as swelling, redness, abnormal secretion in the infected mammary gland and other systemic effects on the cow (Wu *et al.*, 2007). On the other hand, subclinical mastitis (SCM) has no visible signs of infections and hence difficult to detect. SCM is generally monitored and characterised by an increase in somatic cell count (Schukken *et al.*, 2003). SCM is the most prevalent form of mastitis affecting between 20 and 40% of cows in commercial dairy herds in many countries (Ramirez *et al.*, 2014).

Mastitis could be further classified on the basis of severity and duration of infection as acute and chronic. Acute mastitis is characterised by a sudden onset whereas chronic mastitis is characterised by an inflammatory process that lasts for months (International Dairy Federation, 1987). Chronic mastitis is typically characterised by no obvious clinical signs but may show periodical clinical symptoms (International Dairy Federation, 2011). If achieved, objective classification of mastitis could help to implement efficient mastitis control programmes and decision making. For example, identification of cows with chronic mastitis and timely decision (may be culling) could be necessary to maintain herd health as cows with chronic mastitis act as a reservoir of infection for the healthy cows in a herd and hence, are a headache in every dairy farm.

2.1.1 Importance of mastitis

Economically, mastitis is the most expensive disease in the dairy industry incurring huge cost from declining in milk yield, increased treatment costs, discarded milk, increased culling, associated cow replacement rates, and sometimes financial penalties for exceeding legal milk quality limits (Bennedsgaard et al., 2003). Affecting one third of all dairy cows, estimates show that mastitis costs over $1.8 billion a year in the US alone (Schroeder, 2012). Recently, Rollin *et al.* (2015) estimated a cost of $444 for each incident of CM that occurs during the first 30 days of lactation for a representative of a typical large US dairy herd. In a herd-simulation model, Hagnestam-Nielsen and Ostergaard (2009) estimated a maximum loss of €97 per cow/year due to CM in Sweden. Nielsen and Emanuelson (2013) reported that incidence of CM has not changed significantly over the last two decades. Although more difficult to quantify, SCM accounts for almost two-third of economic loss (Halasa *et al.*, 2007; Seegers *et al.*, 2003). In a study to estimate (milk) production loss due to SCM measured via somatic cell count increase in Dutch dairy cows, Halasa *et al.* (2009) found that the greater the somatic cell count increases above 100,000 cells/mL, the greater production losses. Primiparous and multiparous cows were predicted to lose 0.31 and 0.58 kg of milk/day, respectively, at somatic cell count of 200,000 cells/mL (Halasa *et al.*, 2009). The Swedish Board of Agriculture (http://www.jordbruksverket.se/) put mastitis on the top of the list of diseases in dairy cattle in 2013/14 across all herd sizes (Table 1) and across all breed types (Table 2) in Sweden.

Table 1. *Occurrence of disease in dairy cows by herd size and disease, cases of disease per 100 animals included in the control 2013/14*

Disease	Average herd size in the control year								
	- 24.9	25 to 49.9	50 to 74.9	75 to 99.9	100 to 149.9	150 to 199.9	200 to 299.9	300-	Overall
Calving difficulty	0.5	0.5	0.5	0.3	0.3	0.4	0.3	0.3	0.4
Milk fever	3.6	3.3	3.1	2.8	2.6	2.4	2.7	2.4	2.8
Retained placenta	0.6	0.6	0.6	0.5	0.7	0.7	0.5	0.6	0.6
ketosis	2.2	1.5	1.1	0.6	0.5	0.4	0.4	0.4	0.8
mastitis	11.1	10.9	9.7	9.9	10.3	10.8	10.9	14.4	10.8
Udder injuries	0.3	0.3	0.2	0.2	0.1	0.1	0.1	0.1	0.2
Other diseases	10.3	10.8	11.0	10.8	9.8	10.7	10.4	12.9	10.8

Source: Swedish Board of Agriculture (http://www.jordbruksverket.se/)

Table 2. *Occurrence of disease in dairy cows by breed and disease, cases of disease per 100 animals included in the control 2013/14*

Disease	Breed					
	Swedish Red and White	Swedish Friesian	Swedish Polled	Swedish Jersey	other breeds	Overall
Calving difficulty	0.3	0.4	0.2	0.3	0.3	0.4
Milk fever	2.4	3.1	2.6	5.4	2.5	2.8
Retained placenta	0.6	0.6	0.4	0.3	0.5	0.6
ketosis	0.8	0.7	2.6	1.0	0.6	0.8
mastitis	9.5	11.8	10.7	10.0	8.9	10.7
Udder injuries	0.2	0.2	0.3	0.1	0.1	0.2
Other diseases	*9.3*	*12.0*	*8.6*	*12.7*	*8.8*	*10.7*

Source: Swedish Board of Agriculture (http://www.jordbruksverket.se/)

The importance of mastitis goes beyond economics. Seriously affecting animal well-being, mastitis is a primary reason for culling or death of dairy cattle (Grohn *et al.*, 1998). Furthermore, mastitis is associated with the use of antibiotics, which is a public concern (Hansson & Lagerkvist, 2014). Because of concern about overuse of antibiotics that risks the development of antibiotics resistance bacteria, Swedish legislation stipulates that antibiotic therapy may only be initiated after a diagnosis has been made by a veterinarian (Hansson & Lagerkvist, 2014; Nielsen, 2009).

2.1.2 Pathogens of mastitis

Mastitis is often caused by invading pathogens. Multiple microorganisms (e.g., bacteria, viruses, mycoplasma, yeasts and algae) have been implicated as pathogens of bovine mastitis (Watts, 1988). However, just a few species of bacteria are the main pathogens for mastitis (Kuang *et al.*, 2009). Those mastitis-causing pathogens are commonly categorised as contagious and environmental based on their primary source and mode of transmission. The sources of contagious mastitis are infected cows and transmission is hence from cow to cow whereas the primary source of environmental pathogens is the surroundings in which a cow lives (Hogan & Smith, 1987).

The magnitude and increase of somatic cell counts in response to infection varies according to the pathogen involved (Djabri *et al.*, 2002). Depending on the magnitude of inflammatory response, pathogens can be grouped as major or minor pathogens. Major pathogens group include *Staphylococcus aureus*, *Streptococci* (*agalactiae*, *dysgalactiae*, *uberis*), *Escherichia coli* and *Klebsiella spp*. (Kuang *et al.*, 2009; Djabri *et al.*, 2002) . These major pathogens are usually implicated with CM resulting changes of milk composition (Djabri *et al.*, 2002). On the other hand, *Staphylococci* other than *S. aureus* (*S.*

chromogenes, S. hyicus, S. epidermidis, and S. xylosus), and *Corynebacterium bovis* are grouped as minor pathogens (Djabri *et al.*, 2002). These pathogens are associated with a modest infection (Djabri *et al.*, 2002).

Mastitis is a complex and multifactorial disease and hence, for pathogens to enter the udder and cause an infection, depends on a multitude of other factors (Schroeder, 2012). Those factors can be grouped as the individual genotype, the pathogens involved, the environment (housing, bedding, hygiene, climate, milking machines, feeding) and interactions among these (Schroeder, 2012; Oviedo-Boyso *et al.*, 2007) (Figure 1).

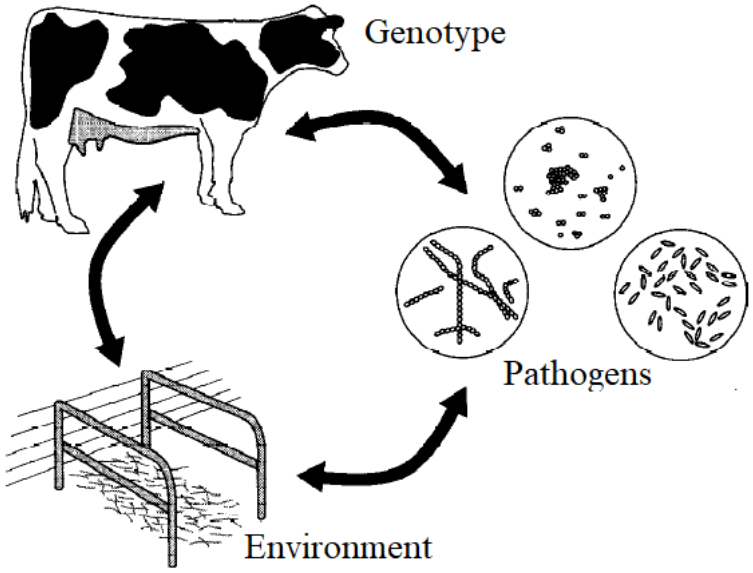

Figure 1. The factors involved in the development of mastitis: the pathogens, the cow as a host, and the environment, which can influence both the cow and the pathogens, adapted from Schroeder (2012).

2.2 Somatic cell count as indicator for mastitis

Somatic cell count (SCC) is the total number of cells normally present in a millilitre of (cow) milk. SCC consists of various cell types with relative proportion depending on the health status of the udder (Nielsen, 2009). In healthy mammary glands, leukocytes (white blood cells) account for the major proportion of SCC (Oviedo-Boyso *et al.*, 2007). These white blood cells are primarily macrophages and lymphocytes with a small proportion of neutrophils and epithelial cells (Kehrli & Shuster, 1994). Infection of mammary glands with mastitis pathogens results in a considerable increase in SCC that are primarily neutrophils (Oviedo-Boyso *et al.*, 2007; Kehrli & Shuster, 1994).

The fact that SCC increases as part of the cow's immune response and the observed high genetic correlation (0.97) (Lund *et al.*, 1994) between SCC and mastitis, makes SCC a widely used and feasible indicator for mastitis.

Most genetic evaluations of mastitis are performed based on either the analyses of recorded treatments of CM or measurements of SCC. Routine genetic evaluations on the bases of CM records have been implemented in the Nordic countries and recently in France (Govignon-Gion *et al.*, 2016) and Canada (Jamrozik *et al.*, 2013). Selection would be more efficient, if genetic evaluations are based on occurrences of CM. However, routine recording of occurrences of CM at farm or cow level is not easily available in many countries because of difficulties associated to its detection (Carlén *et al.*, 2006). With advancements in farming equipment in general and in farms using automatic milking systems (AMS) in particular, SCC records are easily accessible at large scale and with almost no cost (e.g., through the online cell counter fitted in the AMS for continuous monitoring of cow udder health). For this reason, SCC or log-transformed SCC (somatic cell score, SCS) are still used as major phenotypic measures in genetic evaluations to improve udder health of dairy cows (Sørensen *et al.*, 2009). The use of SCC in genetic evaluations of dairy cattle has a long history. In the US during late 1970 and early 1980s, dairy herd breeding programmes began to implement SCC measurements for assessment of mastitis cases (Shook & Schutz, 1994). Such historical and wide acceptance of SCC as a proxy for mastitis is due to its ease of recording and high genetic correlation with mastitis (Gernand & Konig, 2014; Lund *et al.*, 1994). Use of SCC which is easily accessible is further justified by the 2 to 3 times higher heritability of SCC than CM and by the fact that SCC is an indicator of not only CM but also SCM (reviewed by Dekkers *et al.* (1998)).

Misclassifications are unavoidable when SCC is used as an indicator for mastitis. If the boundary between healthy and diseased is too low, high random fluctuations around "normal" SCC levels will lead to falsely classified cases of mastitis (Franzén *et al.*, 2012). Bishop and Woolliams (2010) demonstrated that non-genetic factors such as imperfect sensitivity and specificity of diagnosis are likely to impact genetic parameters for disease traits. The choice of SCC threshold can affect the model's sensitivity and specificity to detect mastitis cases. The SCC that are normally present in a health cow is about 70,000 cells/mL (Djabri *et al.*, 2002) but are dependent on cow factor such as age, breed, stage of lactation, and milk yield (Hiitio *et al.*, 2017; Nyman *et al.*, 2014; Dohoo *et al.*, 2011). Several literatures (Nyman *et al.*, 2014; Schroeder, 2012) propose a threshold of 200,000 cells/mL as optimal cut-off point to determine whether a cow is infected with mastitis.

2.3 Breeding against mastitis

Mastitis is a frequent and costly disease in dairy cattle, but the genetic variability seems to represents a small proportion of the total variance, indicating that genetic progress could be slow. Literature reviews (Rupp & Boichard, 2003; Heringstad *et al.*, 2000) show that SCC and CM have moderate and low heritability, respectively. The low heritability is partially due to the complex nature of the trait and failure of conventional models to capture the available genetic variation (Lipschutz-Powell *et al.*, 2012), otherwise the role of genetics is not negligible. Despite its low heritability, mastitis exhibits genetic variation giving opportunity for efficient genetic improvement of resistance to mastitis in dairy cattle (Govignon-Gion *et al.*, 2016). "From economic and genetic analyses, and according to welfare and food safety considerations and to breeders and consumer's concern, there is more and more evidence that mastitis should be included in breeding objective of dairy cattle (Rupp & Boichard, 2003)".

Selection and breeding against mastitis is performed directly using actual CM records and/or indirectly using indicator traits. In the Scandinavian countries, breeding schemes were redefined in the 90's and have been selecting directly against CM for over 35 years (Osteras *et al.*, 2007; Rupp & Boichard, 2003). More recently, France (Govignon-Gion *et al.*, 2016) and Canada (Jamrozik *et al.*, 2013) have introduced routine national genetic and genomic evaluations for direct selection against CM. However, in most other countries selection against mastitis is performed indirectly through SCC (Miglior *et al.*, 2005). The heritability of SCC (h^2 = 0.17, Jamrozik and Schaeffer (2012)) is higher than the heritability of CM (h^2 = 0.06 to 0.12, Heringstad *et al.* (2000)), giving opportunity for more effective genetic progress than direct selection for CM that requires accurate recording systems. Because many countries record mastitis cases on a voluntary basis the reliability of the genetic estimates for CM are questionable (Thompson-Crispi *et al.*, 2014). The reliability of genetic estimates for CM and other producer-recorded health event data, however, can be improved through genomic selection (Parker Gaddis *et al.*, 2014).

Genomic selection refers to selection methods based on genetic merits directly estimated from genetic markers spanning the genome (Goddard, 2009; Meuwissen *et al.*, 2001). In dairy cattle, this involves shifting from the traditional (progeny testing proven bull-based) breeding programme to genomic estimated breeding value (GEBV-based) breeding programme. These GEBV are calculated from markers spanning the whole genome to the level that every quantitative trait loci (QTL) is in linkage disequilibrium (LD) with at least one marker (Goddard, 2009; Meuwissen *et al.*, 2001). The genetic marker effects are estimated from a subset of individuals from a population with both

phenotypes and genotypes information known called reference population. These marker effects are then used to select candidates with genotypes information but not necessarily phenotypes information (Boichard *et al.*, 2016), highlighting the attractiveness of genomic selection as a tool to improve the genetics of traits difficult or expensive to measure including mastitis. The absence of sufficient recording of CM at farm or cow level because of difficulties associated to its detection has been a limiting factor in breeding against mastitis. The dairy cattle breeding industry is rapidly shifting to genomic selection (Boichard *et al.*, 2012; Hayes *et al.*, 2009). The possibilities to accurately predict the genetic merit of bulls early in life and the increase in prediction reliabilities as well as the economic benefit from these developments are main factors for the rapid adoption of genomic selection by the dairy cattle breeding industry (Hutchison *et al.*, 2014; Schaeffer, 2006). As early as 2012, the use of genomically tested young bulls had reached 50.9% in the US Holstein herds (Hutchison *et al.*, 2014). Traditional breeding programme takes 64 months to complete progeny testing activities for young bulls (Schaeffer, 2006), whereas in the GEBV-based breeding programme a young bull is ready to get its genomic evaluation right after its birth and can be used to sire the next generation once it reached sexual maturity, usually around 12 months. Higher rate of genetic gain is expected from the use of genomically tested young bulls as a result of short generation interval and high reliability. Literature reviews (Hayes *et al.*, 2009) show an increase of 2 to 20 % in reliabilities of GEBV compared to EBV for bull calves with no daughter records i.e. bull calves with only parental average EBV.

Nowadays, genome-wide association studies (GWAS) are the most preferable methods to locate genomic variants or markers associated with complex disease phenotypes. GWAS utilise information on genetic markers like single nucleotide polymorphisms (SNPs) to determine association with a trait of interest assuming that the genetic marker is in LD with, or close to, a causative mutation (Goddard & Hayes, 2009; Hirschhorn & Daly, 2005). In cattle, GWAS have been performed to evaluate marker or SNP association with mastitis related traits and many quantitative trait loci (QTLs) have been reported. QTLs have been reported for SCS on *Bos taurus* autosome (BTA) 6, 13, 14 and 20 in Nordic Holstein cattle (Sahana *et al.*, 2013); on BTA 6, 10, 15, and 20 in Irish Holstein-Friesian cattle (Meredith *et al.*, 2012); genetic variants (a total of 171 significant SNPs) in 24 chromosomes in Valdostana Red Pied cattle breed (Strillacci *et al.*, 2014); and on BTA 6, 13, 19 and X in German Holstein cows (Abdel-Shafy *et al.*, 2014). A review (Sender *et al.*, 2013) indicated that QTLs have been found on almost all chromosomes.

2.3.1 Genetic evaluation of udder health

In current genetic evaluations, only the susceptibility to mastitis is taken into account. A method developed by Franzén *et al.* (2012) included both the disease susceptibility and the recovery process by modelling transitions to and from states of infection. The model was built on the idea that measurements are taken at regular intervals (say, weekly) and that for each measurement an individual cow is classified into one of two possible states (healthy or diseased), whereupon transition probabilities between these states are analysed. This method models transitions to and from states of infection, i.e. both the disease susceptibility and the recovery process are considered, enhancing the genetic evaluation of mastitis. The SCC-based analysis by Franzén et al. (2012), however, ignored possible genetic correlation between susceptibility and recoverability from mastitis. Though both susceptibility to- and recoverability from mastitis were considered, the traits were analysed separately with a single trait model. A simple product-moment correlation between estimated breeding values failed to reproduce different values of simulated genetic correlations between mastitis susceptibility and recoverability. Literature on estimation of genetic correlation between mastitis susceptibility to- and recoverability from mastitis is not available. To be able to investigate whether a genetic correlation exists between mastitis susceptibility and recoverability, a bivariate model had to be developed. Such bivariate methods and models, if applied to real data could enhances the genetic evaluation of mastitis by the ability to capture genetic variation not only for susceptibility to mastitis but also for recovery from mastitis.

Since mastitis is a relatively frequent and unavoidable problem, evaluation of capacity for ability to recover should, therefore, be of interest. The introduction of the recovery aspect as a new trait in the analyses enhance the genetic evaluation of udder health by capturing as much genetic information as possible from the entire disease course. The methods and studies compiled in this thesis allows a simultaneous evaluation of the genetic potential of cows for both susceptibility to- and recoverability from mastitis. In the initial stage of the PhD project, simulation studies were carried out to develop models for simultaneous genetic evaluation of susceptibility to- and recovery from mastitis, followed by real data analyses and estimation of genetic parameters. Genes associated with capacity of animals for recovery were searched and towards the end of the PhD studies a more realistic and dynamic health classification which take severity of possible infection into account was introduced for future genetic evaluation of udder health, specifically mastitis.

2.3.2 Contribution of this thesis to both animal science and the dairy industry

In situations with high mastitis prevalence, evaluating the capacity of cows for recovery could be of specific benefit to the dairy industry. Generally, the dairy industry may benefit from the new approach and models presented in this thesis as follows:

1. Use of estimated breeding values (EBVs) not only for susceptibility but also for recoverability and therefore increased genetic progress per generation as a result of using more information (more accurate estimates).
2. Use of genetic marker assisted selection not only for susceptibility but also for recoverability.
3. Differentiate between different patterns of SCC and assign cows to more realistic and dynamic health classes. This may enable a refinement of the EBVs and/or the breeding goal.
4. Assess individual cows for their susceptibility against mastitis as well as their ability to recover, enabling more informed decisions about culling strategies.
5. Contribute to a better understanding of the dynamics of SCC and mastitis.

Furthermore, the approach and models could be applied to other disease traits and species. Since phenotypic recording is expected to become more sophisticated and automated (e.g., SCC records from AMS), approaches developed in this study can help to exploit such information and model disease data over time. The new approach and models are tailored to monitor health status of animals over time that enables us to make selection decisions for an animal with fast recoverability or even a resistant animal.

3 Aims of the thesis

The aim of the thesis was to improve the genetic evaluation of udder health by introducing a new approach and thereof develop models that can make use of the information contained in the entire disease course. In today's genetic evaluation of udder health, only susceptibility to contract mastitis is taken into account. The thesis aimed to enhance the genetic evaluation of udder health by adding genetics of recoverability in the analyses. Specifically the aim of this thesis was to introduce new models that captures as much genetic information as possible that each cow exhibits during lactation. The objectives of the studies compiled in this thesis were, therefore, to:

(Paper I)

- develop methods and models for a joint genetic evaluation of susceptibility to- and recoverability from mastitis.
- evaluate the effect of daughter group sizes and level of mastitis incidence on breeding value accuracies.

(Paper II)

- evaluate whether the developed methods and models in paper I is identifiable and can be fitted to real data.
- estimate genetic parameters of susceptibility to- and recoverability from mastitis in Danish Holstein cows using the methods and models developed in paper I.

(Paper III)

- identify new or confirm previously identified variants and regions of the genome associated with susceptibility to mastitis.
- identify variants and regions of the genome associated with recovery from mastitis.

(Paper IV)

- improve mastitis health classification by assigning cows to more realistic and dynamic health classes .

- estimate genetic parameters among the newly defined health classes.

4 Summary of investigations

This thesis consists of four studies presented in four papers. In paper I, extensive simulation analyses were performed to develop methods and models for simultaneous genetic evaluation of susceptibility to- and recoverability from mastitis. The bivariate model was evaluated for its performance and findings suggested that it could be used to analyse both traits and genetic parameters could be correctly reproduced. In paper II, the bivariate model developed through simulation with added systematic effects and a function of time was applied to real data. Findings in paper II indicated presence of genetic variation in both traits with heritabilities of similar size. Therefore, in paper III, we performed GWAS to find positions on the genome that affect susceptibility to and/or recoverability from mastitis. Though findings from paper II showed that the traits are strongly negatively correlated, association signals for susceptibility to mastitis were mapped in different locations from associations for recoverability from mastitis, implying that the two traits are at least partially regulated by different genes. Despite the absence of strong association signals, several SNPs were identified within and nearby genes annotated in immunity and wound healing. In the last stage of the PhD project (paper IV) a new and more dynamic health classification which take severity of possible infection into account was introduced as a means to further improve the genetic evaluation of udder health in dairy cattle.

4.1 Material and methods

Often, mastitis is seen as a categorical or binary trait, reflecting presence or absence of mastitis within a defined time interval (Vazquez *et al.*, 2012). In genetic evaluations, this all-or-none trait definition may not fully utilise all information available in the data, for instance the time it takes to recover or different levels of infection (Vazquez *et al.*, 2009; Carlén *et al.*, 2005). The

methods and models developed and used in this thesis have tried to address this issue by adopting the concept of a transition model to define traits of interest (susceptibility to- and recoverability from mastitis). The transition model helps to capture the variations an individual cow may exhibit in the entire disease course- contracting and recovery. During a specified period of time, a cow is assumed to move between or within two states – healthy (**H**) and diseased (**D**). A cow may return to the same state more than once, meaning repeated disease cases are acknowledged by the model (Franzén *et al.*, 2012). For a healthy cow, there is a risk of becoming infected and for a mastitic cow there is a possibility of recovering, which is conceptualised into probabilities of mastitis and recovery (Franzén *et al.*, 2012). For each cow, several transitions from H to D, denoted as **HD**, and from D to H, denoted as **DH**, can occur within a lactation. Franzén *et al.* (2012) depicted such occurrences in a transition probability matrix, T_i , for cow *i*, as follows:

$$T_i = \begin{bmatrix} 1- \pi_i^{HD} & \pi_i^{HD} \\ \pi_i^{DH} & 1-\pi_i^{DH} \end{bmatrix}$$

where $1-\pi_i^{HD}$= Probability of remaining in the H state for cow *i*
π_i^{HD}= Probability of moving from H to D state for cow *i*
π_i^{DH}= Probability of moving from D to H state for cow *i*
$1-\pi_i^{DH}$= Probability of remaining in the D state for cow *i*.

In the above transition probability matrix, the first row consists the probabilities of being in either of both states at time *t+1*for cow *i* healthy at time *t*, and the second row consists probabilities of being in either of both states at *t+1* for cow *i* diseased at time *t*. In practise, the transition matrix is desired to have high values of $1-\pi_i^{HD}$ (probability of remaining in the H state for cow *i*) and π_i^{DH}(probability of moving from D to H state i.e. fast recovery if cow *i* had moved from H to D state), and consequently low values of π_i^{HD} (probability of moving from H to D state) and $1-\pi_i^{DH}$ (probability of remaining in D state) (Franzén *et al.*, 2012).

Thus, for each cow and lactation the sequence of H's and D's, indicating whether or not a cow had mastitis on subsequent test weeks, was converted into a new sequence of weekly transitions indicators: 0 if a cow remains in the same state and 1 if the cow changes state. This resulted into two series of transitions: one for healthy to diseased (HD, to define susceptibility to mastitis) and the other for diseased to healthy (DH, to define recoverability from mastitis). Only diseased cows can recover, and only healthy cows can get diseased. A cow without a case of classified mastitis did therefore not have any data for transitions in the D to H direction. This resulted in a much smaller

dataset for DH in comparison to HD. Hereafter, HD and DH refers to susceptibility to- and recoverability from mastitis, respectively. This concept of traits definition is generally applicable to papers I - III.

4.1.1 Simultaneous genetic evaluation of simulated mastitis susceptibility and recoverability using a bivariate threshold sire model (paper I)

The aim of paper I was two-fold: 1) to develop methods and models for a joint genetic evaluation of HD and DH and 2) to evaluate effect of daughter group size and level of mastitis incidence on breeding values for HD and DH.

Data simulation and statistical model

Simulation software by Carlén *et al.* (2006), further developed by Franzén *et al.* (2012), was used to generate SCC observations for individual cows on the basis of simulated mastitis cases. Weekly SCC observations (generated from the simulated mastitis status) were used to assess whether a cow was in an H or D state. If a cow's observed weekly SCC exceeded a predefined boundary, the cow was considered mastitic and in the D state; otherwise the cow was in the H state i.e. below the boundary (Figure 2).

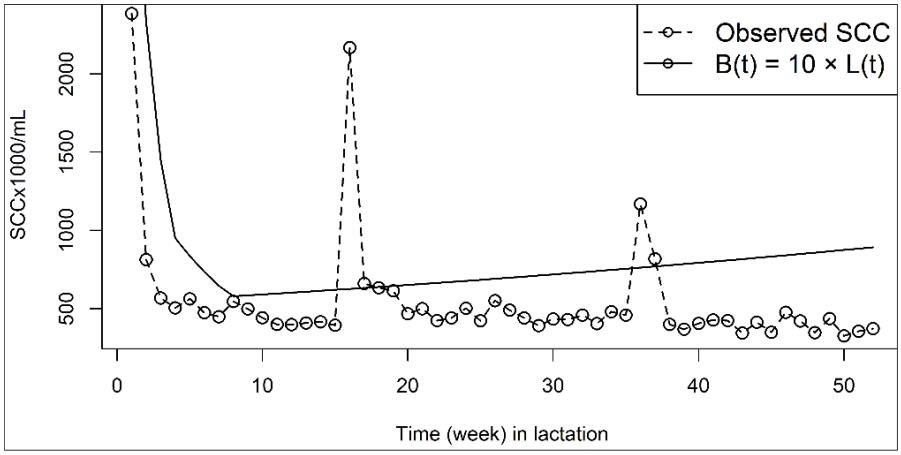

Figure 2. Somatic Cell Count (SCC) level based boundary (B(t), solid line) and observed SCC (broken line) as a function of time (weeks) in lactation for a given cow. According to the observed SCC, the cow in this figure has made a transition from Healthy (H) to Disease (D) state at week 15 and stayed diseased for 2 weeks, recovered at week 18 to stay in H state until week 35 and moved to D state in the next two weeks and finally recovered to remain in H state throughout the remaining lactation period.

Two levels of mastitis incidence were considered: scenario 1 (0.28 cases/lactation) and scenario 2 (0.95 cases/lactation). Three different genetic correlations ($r_g = 0.0$, $r_g = 0.2$ and $r_g = -0.2$) between HD and DH were

simulated. Both traits were simulated to have a within-herd heritability of 0.039. The cows were daughters of either 400 or 100 unrelated sires distributed over 1200 herds with fixed herd size of 20, which resulted into daughter group sizes of 60 and 240, respectively. All combinations of mastitis incidence, daughter group sizes, and genetic correlations were produced in 20 replicates.

Assuming SCC_{ijkt} to be the observed SCC value for cow *i*, daughter of sire *j*, from herd *k* at time *t*, for *t* = 1, 2, 3.... A binary response h_{ijkt} stated whether the t^{th} observation of SCC was below or above the boundary. All the transitions between states were recorded as a binary variable which indicates whether or not a transition took place between two consecutive observations, at time *t* and *t* +1.

$$y_{ijkt}^{HD} = \begin{cases} 1, \text{if } h_{ijkt} = 0 \text{ and } h_{ijk(t+1)} = 1 \\ 0, \text{if } h_{ijkt} = 0 \text{ and } h_{ijk(t+1)} = 0 \\ \text{missing, if } h_{ijkt} = 1 \text{ and } h_{ijk(t+1)} = 0 \text{ or } 1 \end{cases}$$

and

$$y_{ijkt}^{DH} = \begin{cases} 1, \text{if } h_{ijkt} = 1 \text{ and } h_{ijk(t+1)} = 0 \\ 0, \text{if } h_{ijkt} = 1 \text{ and } h_{ijk(t+1)} = 1 \\ \text{missing, if } h_{ijkt} = 0 \text{ and } h_{ijk(t+1)} = 0 \text{ or } 1 \end{cases}$$

for HD and DH, respectively, for *t* = 1,2.3..., $t_i - 1$.

The following bivariate threshold (probit) sire model was fitted to the series of binary transitions indicators:

$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{s} + \mathbf{e}$$

where, $\boldsymbol{\lambda}$ is a vector of the underlying liabilities linked to the transition scores (y_{ijkt}^{HD} and y_{ijkt}^{DH}) with probit function, $\boldsymbol{\mu}$ is mean of HD and DH for an average cow, $\boldsymbol{\beta}$ is vector of fixed effects (herd), **s** is a vector of random additive sire genetic effects, **e** is a vector of random residuals, **X** and **Z** are appropriate incidence matrices. The random effects were assumed to be normally distributed with zero means, additive genetic variance $var(a) = I_s\sigma_s^2$ and residual variance $var(e) = I_n\sigma_e^2$, where I_s and I_n are identity matrices with dimensions equal to the number of sires and transition records, respectively for each trait. σ_s^2 and σ_e^2 are the sire additive genetic and residual variances, respectively. The σ_e^2 was fixed to 1. The two traits were never observed simultaneously and the residual covariance between them was therfore constrained to zero.

Analysis, sampling and Bayesian inference

Bayesian analysis with the RJMC package in DMU (Madsen & Jensen, 2013) was performed to obtain the posterior distribution of all model parameters. Flat prior distributions were employed for all (co)variance components. We ran single chains of 50000 iterations with the first 10000 iterations discarded as burn-in and a sampling interval of 25 to produce a posterior distribution of sample size of 2000 from which point estimates of parameters were derived.

4.1.2 Bivariate threshold models for genetic evaluation of susceptibility to- and recoverability from mastitis in Danish Holstein cows (paper II)

The aim of paper II was two-fold: 1) to evaluate and apply the methods and models developed in paper I to real data and 2) to estimate genetic parameters of HD and DH in Danish Holstein cows.

Data

Data were extracted from a database connected to VMS milking robots (Voluntary Milking System, DeLaval International AB, Tumba, Sweden) fitted with online somatic cell count (OCC) measuring units. Because individuals with an unknown sire are not very informative (especially in a sire model), they were filtered out. Only cows from sires with five or more daughters were kept. After filtering and editing, a total of 1,791 Danish Holstein cows were used for the final analyses. Table 3 shows the number of records, sires, and cows by herd for the edited data used in the final analyses for paper II.

Table 3. *Number of transition records, sires, and cows by herd*

Herd[1]	Sires	Cows[2]	Records[3]
1	45	134	9,848
2	68	388	13,039
3	68	270	11,102
4	45	233	7,006
5	61	346	14,896
6	52	224	15,569
7	43	302	17,772
Total	382	1,897	89,232

[1]The first six herds were commercial farms. The 7th herd was a research herd

[2]For cows changing herds during the lactation (1,897-1791=106 cows), only the records from the first herd were kept

[3]Records were made for weekly transitions between assumed states of mastitis and non-mastitis; 0 if the cow stayed within the same state during the whole week, and 1 if the cow changed state.

Converting OCC to EMR and state transitions

The OCC were converted to elevated mastitis risk- EMR to define a cow's health status on a continuous scale according to the procedure outlined by Sørensen *et al.* (2016). The EMR predicts the risk of a cow having mastitis: values close to zero indicate low risk of mastitis and values close to one indicate high risk of mastitis. The EMR values were used to determine health states (H or D): EMR values above a certain threshold (EMR = 0.6) (Figure 3) were assumed to indicate that the cow had mastitis. If there were multiple observations per day, the highest EMR value was used to create a sequence of transitions. Multiple records of elevated cell counts (EMR values above the threshold) within seven consecutive days for an individual cow were assumed to describe the same case. For each cow and lactation, the sequence of health states was converted to weekly state transitions: 0 if the cow stayed within the same state during the whole week, and 1 if the cow changed state. A time variable was added to indicate the length of each episode. The time indicator counted the weekly intervals until a transition occurred. Because only a few cows had high OCC for a long time, there were very few data points for DH following long mastitis episodes. The time indicator was therefore log-transformed, to avoid substantial influence of these data points in the analyses.

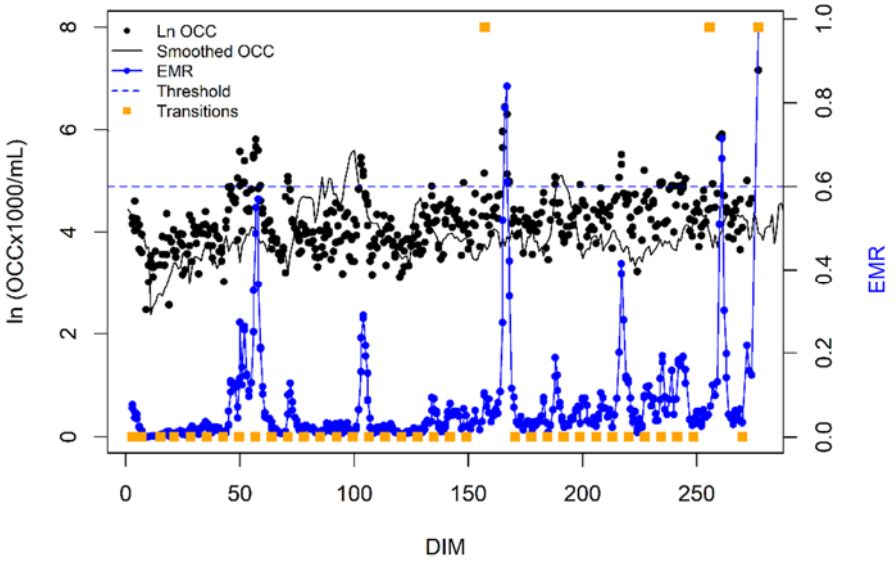

Figure 3. Converting OCC (online somatic cell count) to EMR and state transitions. The smoothed (Smoothed OCC) and ln-transformed OCC from every milking of a cow were combined into elevated mastitis risk (EMR). From the sequence of risk indicator, EMR based health states ('mastitis' if EMR > 0.6), two series of transitions were created: one to model susceptibility to mastitis and the other to model recoverability from mastitis.

Analysis of transitions

In addition to the bivariate threshold sire model (paper I), the transitions were analysed with a threshold animal model for comparison. The observed transitions were modelled as a linear combination of systematic effects and of a function of time. The days in milk (DIM) over the lactation curve [f(DIM)] was modelled by a combination of Legendre polynomials and a Wilmink term ($\exp^{-0.05\times DIM}$) (Wilmink, 1987), to reflect that susceptibility and recoverability are not constant during the lactation. Changes of risk during each episode [f(time)] were modelled with Legendre polynomials. This time effect reflects that the recovery rate during the first week after getting infected may be different from the recovery rate in, say, the third week post-infection, and that a healthy cow may have different risk the first week after recovery from a previous episode compared to a cow having been healthy for many weeks. After comparing models using deviance information criterions (DIC), the traits were analysed with the following threshold model:

$$\boldsymbol{\lambda} = \mathbf{Xb} + \mathbf{Z_1 h} + \mathbf{Z_2 c} + \mathbf{Z_3 p} + \mathbf{Z_3 a} + \mathbf{e}$$

where $\boldsymbol{\lambda}$ was a vector of unobserved liabilities to linked to the observed transition scores (**y**); **b** was a vector of all fixed effects (herd and parity) and covariates (regression coefficients of a second order polynomial on DIM plus a Wilmink term and regression coefficients of a third order polynomial on time); **h**, **c**, **p,** and **a** were vectors of random effects of herd-test-week, cow-parity interaction, cow permanent environmental and animal additive genetic (in animal model) or sire additive genetic (in sire model) effects, respectively; **X,** $\mathbf{Z_1}$, $\mathbf{Z_2}$, and $\mathbf{Z_3}$ were respective incidence matrices, **e** was a vector of the residual effects. The vectors of random effects (**h**, **c**, **p**, **a,** and **e)** were assumed to be normally distributed: $\mathbf{h} \sim N(0, \sigma^2_{htw}\ \mathbf{I})$, $\mathbf{c} \sim N(0, \sigma^2_{cp}\ \mathbf{I})$, $\mathbf{p} \sim N(0, \sigma^2_{pe}\ \mathbf{I})$, $a \sim N(0, \sigma^2_a\ \mathbf{A})$, and $\mathbf{e} \sim N(0, \sigma^2_e\ \mathbf{I})$ where, **I** are identity matrices of appropriate size. **A** is the additive genetic relationship matrix. The σ^2_{htw}, σ^2_{cp}, σ^2_{pe}, σ^2_{a} and σ^2_e were variances for herd-test-week, cow-parity interaction, cow permanent environmental, sire or animal additive genetic, and residual effects, respectively.

Odds ratios (**OR**) were calculated to demonstrate the influence of fixed effects on the two traits. We calculated OR for parity and herd relative to parity one and herd one considered as reference classes for parity and herd, respectively. An OR > 1 indicates that the specified factor has an increased risk than the reference value (Pantoja *et al.*, 2009; Green *et al.*, 2007).

4.1.3 Genome-wide association study for susceptibility to- and recoverability from mastitis in Danish Holstein Cows (paper III)

Paper III was motivated by the observed high negative genetic correlation and presence of genetic variation in both traits that resulted in heritabilities of similar size (paper II). The aim of paper III was three-fold: 1) to identify new or confirm previously identified regions of the genome for their association with susceptibility to mastitis and 2) to identify variants and regions of the genome associated with recoverability from mastitis and 3) to find common genes that affect both traits.

Phenotypic data and phenotypic statistical analyses

The data analysed in paper II were used as raw phenotype data. Phenotypic statistical analyses were performed and the phenotypes were adjusted for the different fixed and random effects stated in paper II. The phenotype data were analysed with a threshold model using the statistical software package MCMCglmm (Hadfield, 2010) in R (R Core Team, 2016). The random cow plus cow*parity effects retrieved from each Markov chain Monte Carlo (MCMC) iteration of the phenotypic analyses were used as the dependent variable in the association analysis.

Genotype data and quality control

The raw data contained 1957 cows genotyped with the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) for a total of 46931 autosomal SNP markers. After matching the phenotype to the genotype data, only 997 cows with phenotypes remained in the genotype data. Quality control (QC) for markers was done using PLINK (Purcell *et al.*, 2007). SNPs with more than 10% missing genotypes, SNPs with minor allele frequency (< 1%) and cows with less than 90% genotype call rate were excluded. In order to visualise possible population stratifications, multidimensional scaling (MDS) plots of an identity-by-state (IBS) matrix was generated containing first two MDS components of the underlying genetic variation using PLINK (Purcell *et al.*, 2007). After QC and removal of four cows (one for having less than 90% genotype data and three other cows for being genetic outliers), 39378 SNPs and 993 cows were used for final association analyses.

Association analysis and significance test

Each SNP was fitted as a covariate in a single SNP association analysis. To account for shared genetic effects of related individuals, pedigree-based polygenic effect was included by fitting individual animal as a random effect in the model. To account for the confounding effect of population structure, the

first top five eigenvectors were included as covariates. The single SNP regression analysis was performed using the following statistical model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Xb} + \mathbf{S}\alpha + \mathbf{Za} + \mathbf{e},$$

where **y** was a vector of corrected phenotypes; **1** was a vector of ones; μ was the general mean; **X** was a matrix containing the five eigenvectors derived and fitted as covariates, and **b** was a vector of associated effects; **S** was vector of SNP genotypes coded as 0, 1, or 2 for genotype copies of one of the alleles at each locus; **α** was the allele substitution effect; **a** was a vector of random additive polygenic effects, which was assumed to follow multivariate normal distribution, $\mathbf{a} \sim N(0, A\sigma_a^2)$ where **A** is the pedigree-based additive genetic relationship matrix; **Z** was an incidence matrix relating elements of the vector of additive polygenic values **a** to individual phenotypes; **e** was a vector of random residuals and the residuals were assumed to follow multivariate normal distribution, $e \sim N(0, W\sigma_a^2)$, where W was a diagonal matrix with diagonal element of $1/(PSD)^2$; σ_a^2 and σ_e^2 were the additive polygenic and residual variances, respectively. The weight $=1/(PSD)^2$, where PSD is the posterior standard deviation of the cow plus cow*parity effects, is used to account for the different residual variances due to parities.

Association analyses were carried out on four datasets (each parity separately and all parities together) to investigate parity specific associations. The analysis was conducted using the statistical software package DMU (Madsen & Jensen, 2013). The null hypothesis H_0: **b**=0 was tested with a t-test. A somewhat liberal threshold of p-value < 10^{-4} was used to declare significant SNP by trait associations. Genes either harbouring or nearby to significantly associated SNP variants were identified using the Variant Effect Predictor (McLaren *et al.*, 2016) and the "ensemble" gene annotation system (https://www.ensembl.org/index.html) in general (Aken *et al.*, 2016).

4.1.4 Dynamic udder health classification using online cell count (paper IV)

In paper IV, a more realistic and dynamic udder health classification defined according to assumed disease severity was introduced. The adoption of AMS by dairy farms presents opportunities to cheaply and easily collect data on a daily basis, or even at individual milkings. Paper IV aimed to further improve the genetic evaluation of mastitis by better capturing the changes in OCC that each cow exhibits in every milking. Accordingly, six mutually exclusive udder health classes with increasing severity were defined. The newly defined mutually exclusive health classes and the criteria use to define them are presented in Table 4.

Table 4. *Criteria for dynamic udder health classification and newly defined health classes*[*]

Health class no.	Health class name	criteria
1	*Very healthy*	EMR ≤ 0.04 and smoothed OCC value ≤ 50,000 cells/mL
2	*Normal healthy*	0.04 < EMR ≤ 0.20
3	*Less healthy (short term)*	0.20 < EMR < 0.60
4	*Short-term sick*	EMR ≥ 0.60 unless acute sick
5	*Acute sick*[**]	First time EMR ≥ 0.60
6	*Long term sick*	This class utilises fluctuating OCC patterns which may indicate presence of pathogens expressing a cyclic shredding pattern

[*]In cases where a milking failed to produce a valid OCC measurement (Sørensen et al., 2016), the previous udder health class assignment was retained. [**]To distinguish between 2 separate acute mastitis cases at least 8 days with EMR < 0.60 was required. Cows only stay in this class until next milking.

Genetic analyses were performed on data restricted to parities 1-5. Cows with an unknown sire were filtered out. Lactations were divided into 3 periods of 100 days each with DIM=5-104, 105-204, and 205-305. Number of milkings in each health class were used as a response variable in the genetic analyses. The number of milkings is a proxy for the amount of time a cow is in a certain health class and as such the genetic parameters reflect the genetic disposition of a cow to stay healthy (many milkings in lower classes) or to recover when ill (fewer milking in higher classes). Cows with less than 30 milkings in a period were not considered in the genetic analyses. For normalization of data, the number of milkings were log-transformed.

Statistical analysis of log-transformed milkings

The log-transformed milkings in each health class were analysed with a linear animal and sire models:

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{Z_1 u} + \boldsymbol{Z_2 a} + \boldsymbol{Z_2 pe} + \mathbf{e}$$

Where; **y** is the vector of the log-transformed milkings in each health class; **b** is the vector of fixed effect of herd; **u** is a vector of random effects of cow-lactation- period; **pe** is a vector of random effects of permanent environment of cow; **a** is a vector of additive genetic (animal model) or sire genetic effect (sire model); **e** is a vector of random residuals; $\boldsymbol{X}$, $\boldsymbol{Z_1}$, and $\boldsymbol{Z_2}$ are incidence matrices associating **b**, **u**, and **pe** and **a** with **y**. The vectors of random effects (**u**, **pe**, **a**, and **e**) were assumed to be normally distributed with $\mathbf{u} \sim N(0, \sigma_u^2\mathbf{I})$, $\mathbf{pe} \sim N(0, \sigma_{pe}^2\mathbf{I})$, $\mathbf{a} \sim N(0, \sigma_a^2\mathbf{A})$ and $\mathbf{e} \sim N(0, \sigma_e^2\boldsymbol{I})$, where σ_u^2, σ_{pe}^2, σ_a^2 and σ_e^2 were variances for cow-lactation-period, permanent environment, additive (animal model) or sire (sire model) genetic, and residual effects, respectively. I_u, I_{pe}

and I_e were identity matrices. **A** was the additive genetic relationship matrix from a pedigree consisting of 22073 animals.

The analyses were performed using the average information restricted maximum likelihood (AI-REML algorithm in DMU (Madsen & Jensen, 2013). Heritabilities were estimated with univariate models (each health class analysed separately), whereas the random effects correlations among health classes were derived from bivariate models (two health classes analysed at a time).

4.2 Main findings

In this section of the thesis we present a condensed version of the main findings of each study. The reader is referred to each paper for detailed and complete descriptions of results in each study.

4.2.1 Simultaneous genetic evaluation of simulated mastitis susceptibility and recoverability using a bivariate threshold sire model (paper I)

Genetic correlation and heritability

Estimates of genetic correlations and heritabilities were mostly in agreement with the simulated (true) values. The mean (of 20 replicates) estimate of genetic correlation deviated by -0.055, 0.001, and 0.008 from the alternative simulated values of $r_g = 0.0$, $r_g = 0.2$, and $r_g = -0.2$, respectively, over all daughter group sizes and mastitis incidences. However, estimates were characterized by a relatively high degree of uncertainty for the alternative with small daughter group size and low mastitis incidence. Example, the length of the 95% highest posterior density (HPD) width in scenario 1 for daughter group size of 60 was larger than 1 (paper I).

The mean of estimated heritabilities for HD (ranging from 0.035 to 0.044) were consistent with the true value (0.039). On the other hand, the mean of estimated heritabilities for DH (ranging from 0.036 to 0.066) were upwardly biased. However, the posterior modes of heritabilities for both HD and DH were closer to the true value (0.039). It is clearly shown that posterior mode predicts better than posterior mean.

Accuracies of breeding values

Defined as the Pearson correlation between sire TBV and EBV, the mean of accuracies of breeding values for HD, in scenario 1, was 0.57 for the daughter group size of 60. For the daughter group size of 240 the accuracies ranged from 0.80 to 0.83. Similarly, in scenario 2, the accuracy of breeding values markedly improved from 0.69 for the daughter group size of 60 to a minimum of 0.88 for

the daughter group size of 240. The accuracies of breeding values for DH were, in all cases, low compared to those for HD and this is because of little information about recoverability due to low mastitis incidence. Accuracies of breeding values for DH in scenario 1 ranged from 0.26 to 0.27 for the daughter group size of 60 and from 0.49 to 0.50 for the daughter group size of 240. In scenario 2, the accuracy improved from a minimum value of 0.44 in the daughter group size of 60 to 0.71 in the daughter group size of 240. Effect of level of mastitis incidence on accuracy of breeding value for both HD and DH was as important as the number of daughters per sire. As expected, over all cases, the more daughters per sire combined with more mastitis incidence, the more accurate EBVs were observed.

4.2.2 Bivariate threshold models for genetic evaluation of susceptibility to- and recoverability from mastitis in Danish Holstein cows (paper II)

Descriptive statistics

Defined by transitions from healthy to diseased status, on average, mastitis incidence was 6% per lactation. The average recovery rate over three lactation periods was 45%. As expected, risk of getting diseased was lowest in lactation 1 and highest in lactation 3.

This trend of disease risk was confirmed with the calculated odds ratios presented in Figure 4. Compared to the reference class of OR = 1 (parity 1), there was higher risk of mastitis for later parities. Later parities were also associated with low recovery rates.

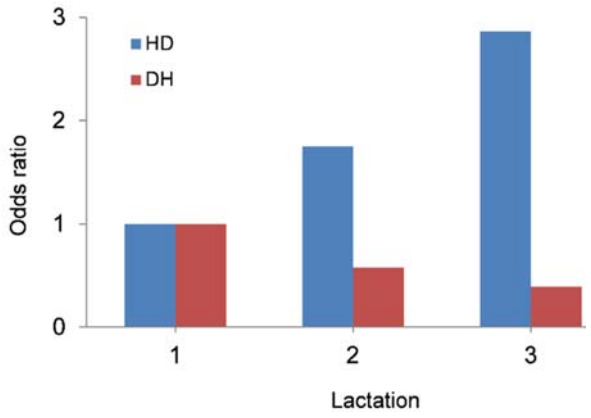

Figure 4. Odds ratio of lactation effect (lactation 1 as reference) for susceptibility (HD) to- and recoverability (DH) from mastitis

Similar patterns across and within lactations were reported by Gernand *et al.* (2012) for different dairy cow disease traits including mastitis. Our results confirm that mastitis incidence is most frequent in the first weeks of lactation

(Nyman *et al.*, 2007; de Haas *et al.*, 2002a) and differ between lactations (Sargeant *et al.*, 1998).

Genetic parameters

There was considerable genetic variation among cows. In general, the posterior means of the variances for HD were higher than for DH except for the herd-test-week variance. Higher herd-test-week variance for DH than for HD indicates that treatment methods and herd management effects would make a bigger difference for DH than for HD. However, the posterior mean of heritabilities and repeatabilities for HD and DH were of similar size in the animal and sire models. Findings indicate presence of additive genetic variation in both traits resulted in heritabilities of similar size ranging from 0.06 to 0.08 (Table 5).

The posterior mean of genetic correlations between the traits were negative for all animal related random effects. This indicates that cows that easily contract mastitis will not recover easily. The correlation between HD and DH for the herd-test-week effect included zero in the 95% HPD intervals in both the sire and animal models (Table 5). Whereas high susceptibility was associated with slow recovery at animal level, a close to zero correlation for the herd-test-week effect (r_{htw} = 0.05, PSD = 0.16) suggested this to be different at herd (management) level: recoverability could be fast even though susceptibility was high at herd level.

Table 5. *Posterior mean (standard deviation) and 95% highest posterior densities (HPD) of genetic parameters for mastitis susceptibility (HD) and recoverability (DH) from sire and animal models*

Parameter	Sire Model			Animal Model		
	Mean	HPD		Mean	HPD	
		lower	upper		lower	upper
r_s or r_g	-0.83 (0.13)	-0.99	-0.54	-0.90 (0.08)	-1.00	-0.75
r_{pe}	-0.90 (0.07)	-1.00	-0.76	-0.80 (0.21)	-1.00	-0.41
r_{cp}	-0.30 (0.10)	-0.50	-0.10	-0.31 (0.11)	-0.51	-0.10
r_{htw}	0.05 (0.16)	-0.26	0.38	0.05 (0.16)	-0.26	0.38
$heritability_{HD}$	0.07 (0.03)	0.03	0.13	0.07 (0.02)	0.04	0.10
$heritability_{DH}$	0.08 (0.03)	0.03	0.14	0.06 (0.02)	0.02	0.09
$repeatability_{HD}$	0.11 (0.01)	0.08	0.13	0.11 (0.01)	0.09	0.14
$repeatability_{DH}$	0.09 (0.02)	0.06	0.13	0.10 (0.02)	0.06	0.13

Correlations: r_s or r_g, r_{pe}, r_{htw} are correlations between mastitis susceptibility (HD) and recoverability (DH) for sire or animal additive genetic, cow permanent environmental, cow-parity interaction, and herd-test-week, respectively.

4.2.3 Genome-wide association study for susceptibility to- and recoverability from mastitis in Danish Holstein cows (paper III)

Complexity of the traits was manifested with the absence of strong association signals suggesting that numerous genes with small effects could be involved in both directions of the disease. However, depending on the level of significance, several SNP by trait associations were identified. The number of significant SNPs at different p-value cut-off levels is presented in Table 6.

Table 6. *Number of associated SNP variants at different nominal significance levels*

Trait[1]	p-value < 1e-04	p-value < 0.001	p-value < 0.01	p -value< 0.05
HD	2	44	424	2144
DH	4	46	353	1862
P1HD	4	36	398	2154
P1DH	2	36	374	1933
P2HD	3	44	419	2065
P2DH	5	50	403	1959
P3HD	5	38	468	2262
P3DH	4	25	231	1512
Total	29	319	3070	15891

[1]Trait: HD and DH represents susceptibility to- and recovery from mastitis all parities analysed together; P1HD, P2HD and P3HD = susceptibility to mastitis in parity 1, 2 and 3, respectively; P1DH, P2DH and P3DH = recovery from mastitis in parity 1, 2 and 3, respectively.

The nominal significance levels of p-value < 10^{-4} was considered to determine genome-wide significant SNP-trait association. Hence, for the full data (all parties analysed together), significant SNP by trait associations were detected on BTA3 for HD and on BTA7 and 15 for DH. Parity specific significant SNP by trait association were also detected on the same chromosomes but no overlapping association signals were observed. Significant association signals in parity 1 and in the other parities (parity 2 or 3) detected on the same chromosome did not overlap, for either of the traits. For HD, the highest number of significant associations were observed on BTA7 and 13. For DH, the highest number of significant associations were observed on BTA12 and 13. In total, 29 significant SNPs (14 for HD and 15 for DH) were detected. The identified significant SNP variants were annotated within or near a region known for their role in immune systems. Only genes that harbour identified significant SNP variants in each trait are presented in Table 7.

Table 7. *Candidate genets identified by overlapping SNPs showing significant [–log10(P-value) > 4] associations with susceptibility to- (HD) and recovery (DH) from mastitis*

Trait*	BTA	SNP rsID	Position (bp)*	α	SEα	-log$_{10}$P	Overlapping candidate gene
HD	7	RS43732911	28337398	-0.24	0.06	4.09	*MARCH3*
DH	15	RS110414316	53138487	-0.09	0.02	5.13	*ATG16L2*
P1DH	15	RS42550814	4668361	0.09	0.02	4.87	*PDGFD*
P2HD	5	RS109785134	67780167	0.15	0.03	5.21	*STAB2*
P2DH	1	RS109126926	111030092	0.08	0.02	4.11	*PTX3*
P3HD	13	RS43150474	3815373	0.16	0.04	4.67	*SLX4IP*

*Trait: HD and DH represents resistance to- and recovery from mastitis all parities analysed together; P1HD, P2HD and P3HD = resistance to mastitis for parity 1, 2 and 3, respectively; P1DH, P2DH and P3DH = recovery from mastitis for parity 1, 2 and 3, respectively. The SNP positions in base pairs (bp) were based on the UMD3.1 genome assembly (ftp://ftp.cbcb.umd.edu/pub/data/assembly/Bos_taurus/Bos_taurus_UMD_3.1/)

4.2.4 Dynamic udder health classification using online cell count (paper IV)

Results indicated presence of considerable additive genetic variance in health classes defined for longer periods, whereas the variations in health classes defined for short-term and sudden changes (e.g., acute) were rather attributed to environmental factors. The estimated heritability for the acute health class was 0.00. Heritabilities were highest for the very healthy class in both the sire and animal models. Heritability of long-term sick was modest ($h^2 = 0.07$) in the sir model (Table 8) and a bit lower in the animal model (paper IV). Estimated genetic parameters for all the newly defined health classes from the sire model are presented in Table 8.

Table 8. *Heritabilities (diagonal), genetic correlations (above diagonal) and phenotypic correlations (below diagonal) of log-transformed milkings in each health class*

Health Class No.*	1	2	3	4	5	6
1	0.11	-0.86	-1.00	-0.82	-0.99	-1.00
2	-0.37	0.08	0.43	0.60	-1.00	0.88
3	-0.34	0.24	0.06	1.00	1.00	1.00
4	-0.01	-0.22	0.17	0.09	1.00	1.00
5	-0.10	0.10	0.49	0.11	0.00	-0.98
6	-0.14	-0.26	-0.06	-0.06	-0.16	0.07

Health Class No.*: 1= Very healthy, 2 = Normal healthy, 3= Less healthy (short-term), 4 = Short-term sick, 5 = Acute, 6 = Long-term sick.

Though results in Table 8 are preliminarily, the genetic correlations suggest that the health classes could be reduced to 3 i.e. 1) very healthy; 2) normal healthy; 3) diseased that includes four of the other intermediate classes (short- and long-term less healthy acute and long-term sick or persistent).

Lactation stage (period) effects on number of milkings are presented in Figure 5. The frequency of milkings belonging to the different stages of lactation followed to a certain extent similar patterns across stages, with the least numbers in the acute health class. Lactation 1 and 2 showed higher numbers in the healthy classes and lower numbers in the disease classes compared to later lactations (Figure 5).

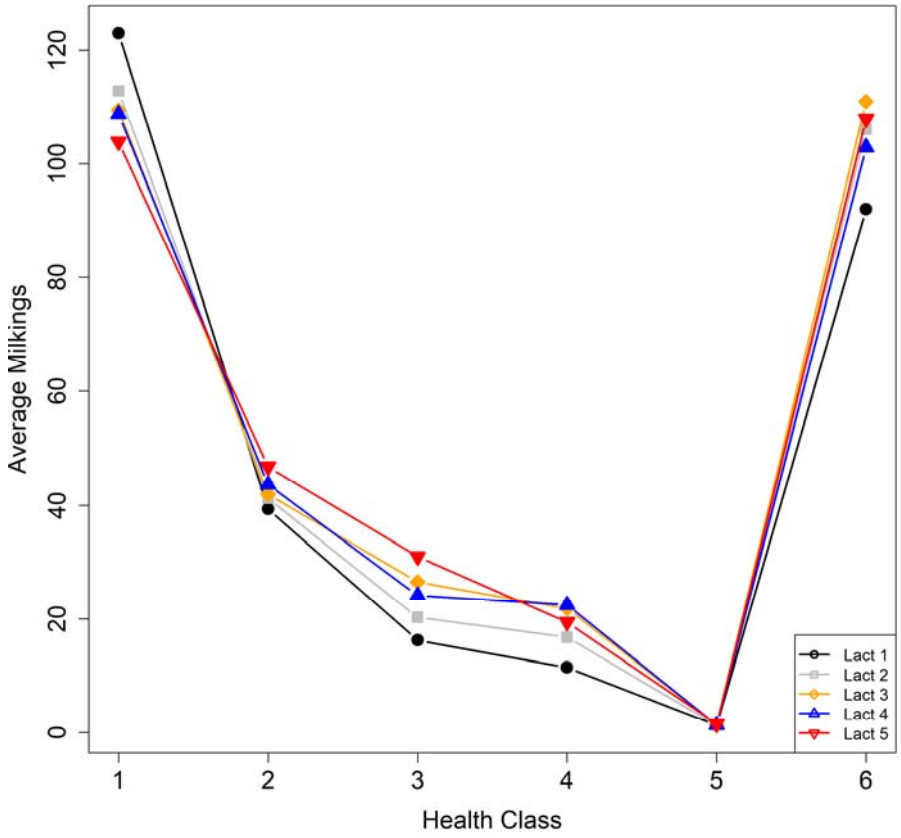

Figure 5. Lactation effect on health classes. Health Class: 1= Very healthy, 2 = Normal healthy, 3= Less healthy (Short-term), 4 = Short-term sick, 5 = Acute, 6 = Long-term sick.

5 General discussion

5.1 Joint modelling of susceptibility to- and recoverability from mastitis

The overall aim of this thesis is to improve the genetic evaluation of udder health using methods and models that can make use of the information contained in both directions of the disease: susceptibility to contract mastitis and recoverability from mastitis. With regard to introduction of recoverability or modelling both directions of mastitis, this thesis is novel in the area of genetic evaluation of udder health. In addition to the introduction of recoverability, the studies compiled in this thesis have made use of transition methods (papers I-III) and time-dependent models (papers II and III) to better use of the information contained in SCC than does the often used lactation-average models which is prone to loss of information.

Mastitis is a common disease causing large negative economic and animal welfare consequences for the dairy industry (Rollin *et al.*, 2015; Nielsen, 2009; Halasa *et al.*, 2007). For this reason, the need for better modelling is high and several dairy cattle researches have been focused on mastitis over the last decades (Govignon-Gion *et al.*, 2012; Philipsson & Lindhe, 2003). Selection against mastitis has been performed indirectly (mainly using SCC traits as indicator) and/or directly (with CM records) in many countries (Govignon-Gion *et al.*, 2016; Jamrozik *et al.*, 2013). None of these countries has yet included the recoverability of animals in the selection index for udder health, partly because of the fact that knowledge of different genetic and non-genetic parameters of a trait is a prerequisite to do so. The literature offers little information regarding parameters of recoverability from mastitis such as its recoding difficulty, economic importance, presence of genetic variance, and genetic correlations with other traits of interest (e.g., with susceptibility). A few studies (Fogsgaard *et al.*, 2015b; Franzén *et al.*, 2012) have been performed to evaluate the genetic and non-genetic merits of recoverability

from mastitis in dairy cows. Franzén *et al.* (2012) developed a model using transition probabilities for genetic evaluation of susceptibility to- and recoverability from mastitis using simulation studies. Recently, Fogsgaard (2015) studied behavioural changes with regard to feed intake and production performances during and weeks after recovery from CM. These authors have proposed the introduction of the recovery aspect as a new trait in the analyses to enhance the genetic evaluation of udder health (Franzén *et al.*, 2012) and to promote normal levels of production, feed intake and behaviour (Fogsgaard, 2015) in dairy cows. The studies compiled in this thesis (papers I-III), have incorporated this novel idea of modelling recoverability to improve the genetic evaluation of udder health. The information from the genetics of recoverability will result in higher accuracy of estimates. Moreover, since mastitis is a common and unavoidable problem, evaluation of ability of animals to recover should be of interest. In papers I-III, we have evaluated and presented important genetic parameters such as the between traits genetic correlation, size of the additive genetic variance, and associated SNP variants in the bovine genome not only for susceptibility but also for recoverability from mastitis.

Despite considerable research, response to mastitis has not yet been a success story for the animal geneticists and breeders. One of the main factors for this is the low accuracy of selection mainly resulting from less accurate diagnosis of health status and failure of methods and models to capture available information in disease data. For example, often genetic evaluation of mastitis is based on two phenotypes: healthy or diseased, which is prone to loss of information and failure to take the dynamic nature of susceptibility into account, leading to low accuracy. Lipschutz-Powell *et al.* (2012) showed failure of conventional models (such as sire models) to capture the genetic variation even if it is present in a disease data. The methods and models developed and used in this thesis (papers I-III) have been tailored to address this issue by adopting the concept of transition model to define traits of interest. Furthermore, Several studies (Koeck *et al.*, 2010; Negussie *et al.*, 2008) suggest that CM and SCC in an interval may have a more similar genetic basis than CM and SCC in a different interval, and therefore, the application of test-day (longitudinal) models for SCC would be more appropriate. In addition to addressing dynamics of SCC using transition methods, papers II and III, adopted longitudinal models considering time-dependent variables and lactation effect to model time-dependent observed transitions. The observed transitions were modeled as a linear combination of systematic effects and of a function of time to reflect the dynamic nature of susceptibility and recoverability within and across lactations.

Moreover, the use of OCC from AMS (papers II-IV) makes the methods and models developed in this thesis more relevant for the dairy industry. The global dairy industry is in a continuous structural change. Particularly, in most developed countries, the average size of dairy herds is continuously increasing and in parallel, the use of AMS is becoming more popular (Barkema *et al.*, 2015). The adoption of AMS by dairy farms presents opportunities to cheaply and easily collect data on a daily basis, or even at individual milkings. This thesis offers methods and models to exploit the information collected via AMS. Parameter estimates obtained from such information rich data are more accurate and relevant to enhance the overall genetic evaluation of mastitis as information collected via AMS is more objective and timely than information collected with human observation (reviewed by Barkema *et al.* (2015)).

5.2 Analysis of major findings

5.2.1 Performance of methods and models

In paper I extensive simulation analyses of susceptibility to- and recovery from mastitis were performed to develop a bivariate model and test its performance in reproducing different genetic parameters. We put emphasis on the between traits genetic correlations to evaluate the ability of the model in reproducing the simulated true parameter values. The mean (of 20 replicates) estimate of genetic correlation deviated by −0.055, 0.005, and 0.008 for the alternative with simulated true values of 0.0, 0.2, and −0.2, respectively, over all daughter group sizes and mastitis incidences. Figures 6 and 7 show the 95% highest posterior density (HPD) interval of genetic correlations for the 20 replicates for the daughter group size of 240 and 60, respectively. In most cases (93% of the replicates), the true parameter value was within the HPD interval, indicating that the methods and models can provide accurate estimates for the between trait genetic correlation. However, in situations where there is little information about recovery because of the fact that only diseased cows can recover, estimates were characterised by a wider HPD interval, indicating relatively high degree of uncertainty (Figure 7, left).

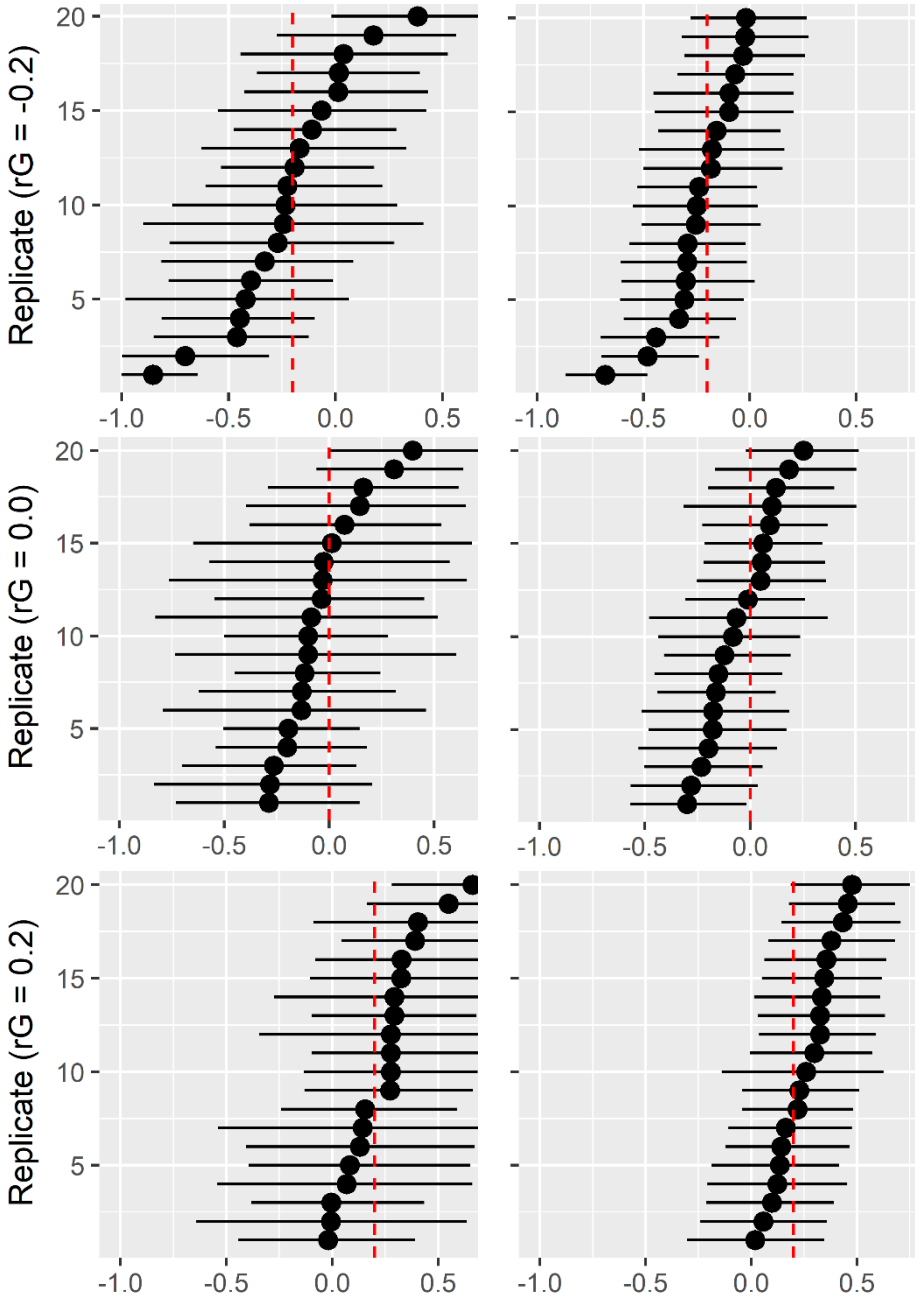

Figure 6. 95% highest posterior density intervals for genetic correlations between susceptibility to – and recoverability for mastitis cases of 0.28 (left) and 0.95 (right) per lactation for the daughter group size of 240 in each replicate (20 replications). Dashed red lines indicate true (simulated) values of genetic correlations of r_G = -0.2, 0.0 and 0.2. Dark dots represent the posterior mean in each replicate derived from posterior distribution of sample size of 2000.

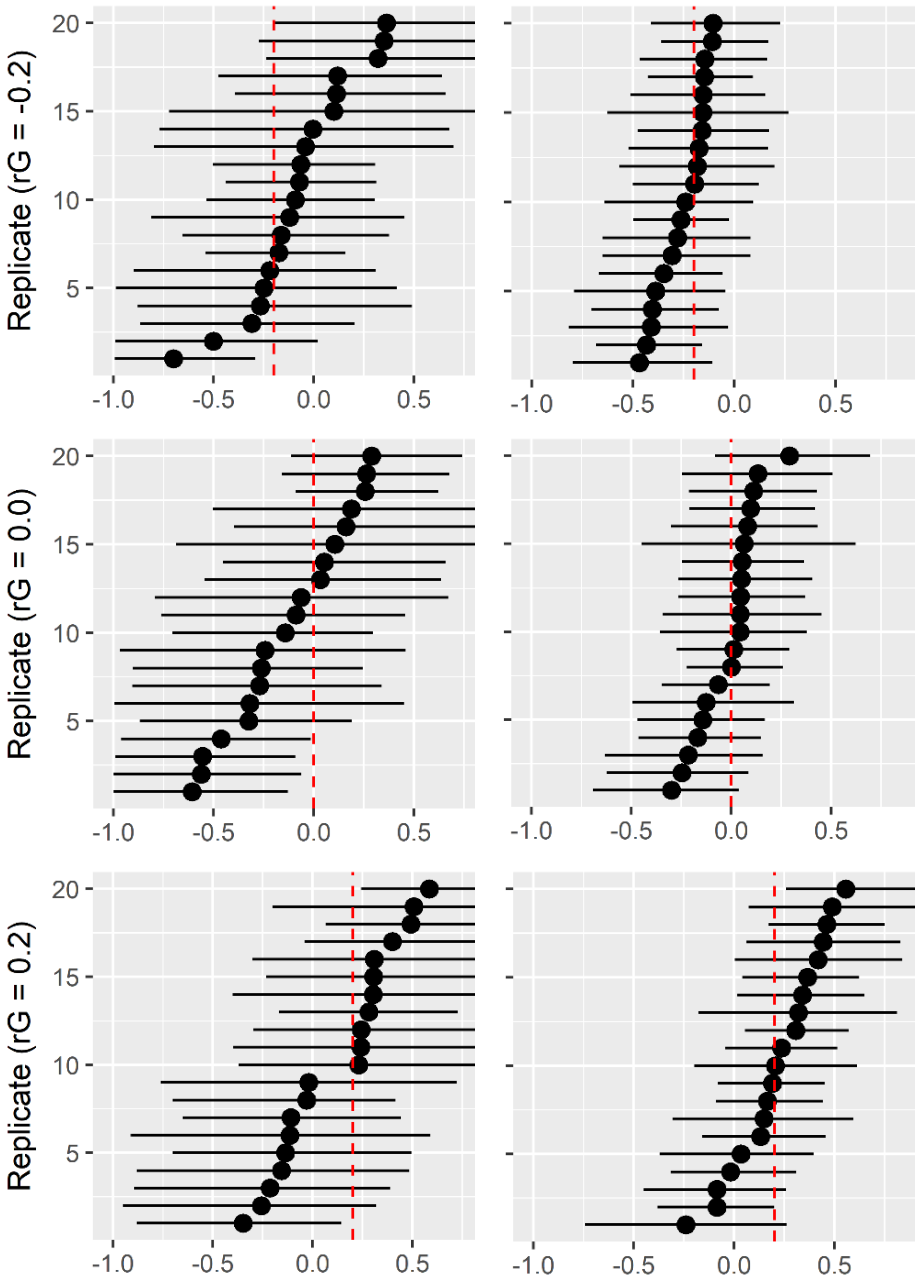

Figure 7. 95% highest posterior density intervals for genetic correlations between susceptibility to – and recoverability for mastitis cases of 0.28 (left) and 0.95 (right) per lactation for the daughter group size of 60 in each replicate (20 replications). Dashed red lines indicate true (simulated) values of genetic correlations of r_G = -0.2, 0.0 and 0.2. Dark dots represent the posterior mean in each replicate derived from posterior distribution of sample size of 2000.

Heritability estimates were approximately in agreement with the true value (0.039). However, due to the smaller number of observations, the heritability estimates for DH were slightly upwardly biased. In a study with the same simulation design, different models (linear, threshold, and survival) were compared and a range of heritability estimates were reported (Carlén *et al.*, 2006). For survival analysis of time to first mastitis they reported a value of 0.037, which is very close to the true value. This supports our estimates, because their survival analysis model is conceptually equivalent to analysing state transitions with a probit model. One difference, and advantage, is that our model allows repeated mastitis cases, whereas their survival analysis did not.

The model was also evaluated for its performance in estimating accuracies, measured via the Pearson product-moment correlation between EBV and TBV. As expected, accuracies were higher for the larger population size and higher mastitis incidence. In earlier analysis of single trait threshold sire model using the same simulation design, the accuracy of breeding values ranged from 0.53 to 0.60 (Carlén *et al.*, 2006) for the daughter group of 60 in scenario 1. The corresponding accuracy in paper I (ranging from 0.56 to 0.57) was consistent with that study. The accuracies of breeding values for DH (ranging from 0.26 to 0.48) compared to HD were, however, low in all cases. Similar accuracies of breeding values for DH that ranged from 0.24-0.48 were reported by Franzén et al. (2012). The reason for this is the much reduced information in the DH data; because of little information in recoverability.

5.2.2 Application of the methods and models

The bivariate model developed through simulation in paper I with added time function as well as several systematic effects to allow for difference of susceptibility or recoverability over the course of lactation was applied to real data. Since this was the first time that the method developed in paper I is applied to real data, we had little options to compare and contrast the genetic parameter estimates particularly for the DH trait. A few studies (e.g., Urioste *et al.* (2012)) have evaluated different traits but conceptually equivalent to the DH trait in this thesis. Urioste *et al.* (2012) analysed average days diseased (per lactation) in Swedish Holstein cows as a trait and reported increased days diseased for later parities which is comparable to the results in paper II for DH where recovery rate was highest in parity 1 (fewer days diseased) and lowest in parity 3 (more days diseased). They also reported parity specific heritabilities that ranged from 0.06 - 0.17, with lower heritability for late parity and higher heritability for earlier parity. Both the results by Urioste *et al.* (2012) and our heritability estimates from the sire ($h^2 = 0.08$) and animal ($h^2 = 0.06$) models in paper II indicate the presence of heritable variation in the DH trait. The

posterior distributions of heritabilities (Figure 8) also indicate that the trait DH has similar size of heritability as the trait HD and hence, could be considered a new trait for selection.

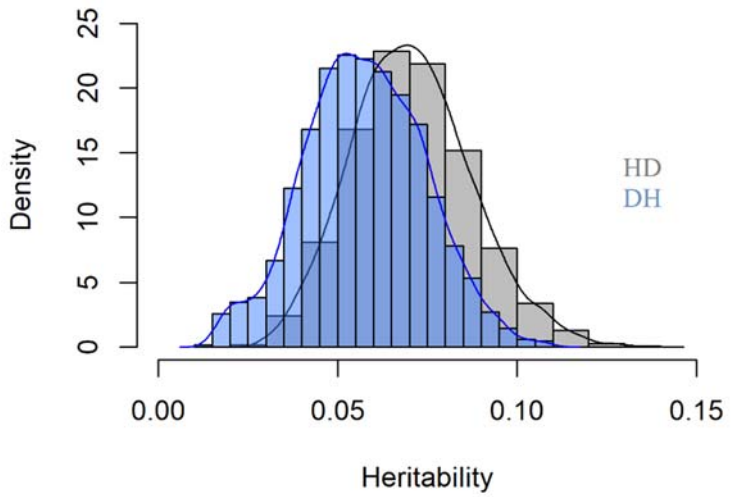

Figure 8. Posterior distributions of heritability estimates from a Markov chain Monte Carlo sample size of 5000 for susceptibility (HD, gray and overlapped area) to- and recoverability (DH, blue and overlapped area) from mastitis obtained from the animal model.

The heritability estimate for HD was consistent with literature values, indicating the successful application of the methods and models to real data. A review by Heringstad *et al.* (2000) indicated that estimates of heritability for susceptibility to CM from threshold models ranged from 0.06 to 0.12. Other studies (Koeck *et al.*, 2010; de Haas *et al.*, 2008) showed heritability of SCC-based traits ranging from 0.09 to 0.13, which is consistent with the heritability for HD in paper II. Despite differences in analyses and definitions of SCC-related traits, results from several previous reports (Koeck *et al.*, 2010; de Haas *et al.*, 2008; Lassen *et al.*, 2003) and our estimates in paper II fell within the range of estimates presented by Heringstad *et al.* (2000).

The genetic correlation between HD and DH in paper II is strongly negative. Urioste *et al.* (2012) reported strong positive genetic correlations (0.97) between average days diseased and average SCC in early lactation (5–150 days) implying that in genetic terms either of the traits could be explained by the other one. A strong negative genetic correlations could point at a single underlying genetic mechanism affecting both HD and DH. However, the observed genetic correlation between HD and DH (-0.83 and -0.90 from the sire and animal models, respectively) in paper II is not complete (i.e. 1), meaning there may be some merit in selecting also for the DH trait.

Contrary to the observed high negative genetic correlation (paper II), association signals in paper III were mapped in different locations, showing the traits to be rather polygenic. Much larger GWAS would be required to get a better estimate of which SNP affect both traits. We also observed no overlap of signals across parities implying not only the traits but also lactations could be polygenic. Effect of genes has been shown to vary depending on parity or age of cows (Wojdak-Maksymiec *et al.*, 2013). Wojdak-Maksymiec *et al.* (2013) reported that allele T of the gene *TNF-α* was associated with a lower number of mastitis cases in lower parities and a higher number of mastitis cases in higher parities. Similarly the casual variants or genes could be different for HD and DH confirming the findings that quantitative trait loci (QTLs) even for highly correlated traits (e.g., SCC and CM) present on the same chromosome do not necessarily overlap (Sender *et al.*, 2013). Complexity of the traits was manifested with the absence of strong association signals suggesting that numerous genes with small effects could be involved in both directions of the disease, making it difficult for GWAS to identify causal genes as described in Hayes *et al.* (2010). Further exploration of the genetic basis of susceptibility and recoverability would require a much larger GWAS study in terms of animals with records. With the increased use of milking robots with automated recording and the routine genotyping of cows a much larger GWAS using the same approach should be feasible in the foreseeable future.

Despite the absence of strong association signals, several SNP variants were identified within or nearby genes known for their role in immunity and wound healing. Most of the suspected candidate genes for HD and DH are found in chromosomes frequently reported (e.g., on the cattle QTL database; https://www.animalgenome.org/cgi-bin/QTLdb/BT/index) for their association with mastitis traits defined from SCCs. From the several candidate genes identified by nearby or overlapping SNP variants, here we discuss just a few of them that have been reported for their association with immune systems. From literature and gene annotation databases (e.g., the human gene database, http://www.genecards.org/), it is plausible to suggest that MARCH3 (involved in regulation of the endosomal transport pathway (Fukuda *et al.*, 2006)), MAST3 (highly expressed in antigen-presenting cells and in lymphocytes (Labbe et al., 2008) and STAB2 (involved in lymphocyte homing, cell adhesion, and receptor scavenging (Howard *et al.*, 2015; Schledzewski *et al.*, 2011)) as casual candidate genes for susceptibility to mastitis. For recovery from mastitis we suggest EPS15L1 (essential for lymphocyte development (Seiler *et al.*, 2015)), PDGFD (involved in macrophage recruitment and wound healing, (Uutela *et al.*, 2004)), and PTX3 (involved in regulating inflammation, http://www.genecards.org/) as candidate for the casual genes.

5.2.3 More detailed and dynamic health classification

Another point this thesis tried to consider is the unavoidable misclassifications when SCC is used as an indicator for mastitis. In the boundary-based classification (papers I-III), if the boundary between healthy and diseased is too low, high random fluctuations around "normal" SCC levels will lead to falsely classified cases of mastitis (Franzén *et al.*, 2012). Bishop and Woolliams (2010) demonstrated that non-genetic factors such as imperfect sensitivity and specificity of diagnosis are likely to impact genetic parameters for disease traits. The misclassification and probability of inaccurate diagnosis may be improved by going from a strict limit between two states to a more flexible and realistic classification. Recently, a dynamic and realistic assignment to udder health classes that reflect the continuous nature of susceptibility to a disease has been recommended by Løvendahl and Sørensen (2016). In paper IV, we introduced a more dynamic health classification which takes, severity of possible infection and trend of OCC increase into account. This will potentially improve the genetic evaluation of mastitis even further as variations observed at every milking can be accounted in either of the health classes (*very healthy, normal healthy, short-term less healthy, short-term sick, acute ,* and *persistent* or *long-term sick*).

Results presented in paper IV show possibilities to capture additive genetic variance not only in the two extremes of health classes (healthy and diseased) but also from intermediate health classes. The health classes: *very healthy*, *normal healthy*, *short-term less healthy*, *short-term sick*, *acute,* and *long-term sick* had heritability of 0.11, 0.08, 0.06, 0.09, 0.00, and 0.07, respectively. Results in paper IV indicate the presence of considerable genetic variance for cows' presence in health classes defined for longer periods, whereas the variations in health classes defined for short-term and sudden changes (e.g., acute) were mostly attributed to environmental factors. Methods and models presented in paper IV can help to capture the changes in SCC that each cow exhibits in every milking and to make selection more efficient by taking into account the dynamic nature of disease susceptibility.

5.3 Limitations of the studies

The methods and results presented in this thesis have shown possibilities of joint modelling and analyses of susceptibility to- and recoverability from mastitis. The application of a transition method (paper I) as well as the application of test-day (papers II and III) and the more detailed health classes (paper IV) are tailored to better capture changes in SCC and account the dynamics of SCC a cow may exhibit during lactation. However, the major

limitation of the studies compiled in this thesis is absence of recorded cases of CM. In the genetic evaluation of udder health the use of SCC compared to use of recorded cases of CM has been justified, because: 1) SCC is easy and inexpensive to collect; 2) SCC is genetically highly correlated with CM; 3) SCC is 2 to 3 times as heritable as CM; 4) SCC is an indicator of not only SCM but also CM (reviewed by Dekkers *et al.* (1998)). Despite these justifications there are studies (Schukken *et al.*, 2011; Suriyasathaporn *et al.*, 2000) questioning the concept of selection against higher SCC because of its potential to jeopardise the immune system. In addition, though SCC is genetically highly correlated with CM and SCM, the correlation has never been complete (i.e. 1), meaning health classes are inaccurate and misclassifications are unavoidable when SCC is used as an indicator for mastitis. Moreover, there are reports (Koeck *et al.*, 2010) showing that CM and SCC may describe different aspects of udder health, making it difficult to get accurate health classes. Accounting both SCC and recorded cases of CM could help to get accurate health classes which determine the rate of genetic gain. Recently Canada has developed a selection index of mastitis resistance derived from both CM and SCC (Van Doormaal & Beavers, 2014).

Another limitation of this thesis (specifically papers I-III) is the observed incomparable data structure between the HD and DH traits. The trait recoverability can only be observed in cows with mastitis history. In other words, only diseased animals can recover, limiting the number of observations for the DH dataset. For example, the observations for HD were above one million per replicate, but for DH observations were in ten thousands (paper I). Results in paper I, indicate that the predictive ability of the model was sensitive to the number of observations. The more daughters per sire combined with more mastitis incidence, the more accurate EBVs were observed. Additionally, the size of the real data used in papers II - IV was much smaller than the simulated data in paper I. This is shown by the observed large SE on the genetic parameters in papers II and III. For example, the SE of heritability estimates from the sire model for HD and DH are 0.002 and 0.006 in paper I whereas the average corresponding estimate in paper II is at least 10 times larger than these estimates. We also think that the whole GWAS in the current thesis has probably been influenced by the small data size. We say this because several of the QTLs for mastitis reported in literature have not been replicated in our GWAS (paper III). Therefore, much larger data and probably whole genome sequencing will be required to efficiently detect marker effects for both traits. The statistical power of GWAS can be increased by increasing the sample size of genotyped cows and using high density SNP arrays than the current marker density from

the Illumina 50K SNP array data (Mao, 2016). Nevertheless, the current GWAS had some power to pick up large effects, if they segregated.

Studies compiled in this thesis did not answer how recovery was achieved. We do not know if recovery was achieved with antibiotics or without antibiotics and hence, treatment was not accounted in the methods and models. By ignoring treatment we assume all treatment effects were equal for every animal. This is a very simplified assumption and should be considered as limitation in this study. In reality, recovery from mastitis could be achieved with antibiotics treatments or naturally without treatments by the ability of cows to overcome infection. From genetics point of view both ways of recovery are relevant because not only the natural recoverability but also the response to treatment is somewhat genetically controlled. For efficient genetic gain, however, it is important to distinguish how recovery is achieved and to use appropriate data in the genetic evaluations.

5.4 Future research and directions

Better modelling for genetic evaluations and accurate classification of health status are necessary to improve udder health through selection and breeding. This thesis has introduced better models and methods to accommodate the dynamic nature of susceptibility and recoverability over time. However, the SCC-based classification of cows as healthy or mastitic presented in this thesis is critical, and therefore, it is important that the methods and models are further investigated using recorded cases of CM and recovery, including records of treatments. To improve the indictor traits based health classification, the EMR could be supplemented by other traits that are routinely recorded. For example, electrical conductivity of milk (EC) is automatically recorded at every milking in most herds using AMS. The genetic correlation between EC and mastitis is strong, ranging from 0.65 to 0.8 (reviewed by, Norberg (2005), making it a suitable trait to supplement the EMR-based health classification.

The introduction of the recoverability in the genetic evaluation of mastitis opens a new avenue of future research. Selection and genetic improvement of cows to overcome mastitic infection requires balance against other traits. Do cows that recover from mastitis have the ability to come back to full capacity of production and other performances? A study (Fogsgaard *et al.*, 2015a) has shown that fast recovered cows can quickly return to normal levels of production, feed intake and behaviour. However, the study by Fogsgaard *et al.* (2015a) was limited in its genetic scope. Therefore it is important that genetic correlation between recoverability and other traits of dairy cattle is investigated in the future.

Another interesting point that deserves future investigation is the contradicting results shown in papers II and III. Understanding the biological background for the observed high negative genetic correlation (paper II) and absence of overlapping association signals (paper III) is not easy without further investigations. SCC are usually extremely polygenic and under high environmental influence, however, as the HD and DH traits are genetically highly correlated, we expect overlapping signals for both HD and DH on same chromosome, but results indicate rather absence of overlapping association signals. This may indicate the need for (much) more data than the current genotype data. The findings in paper III could, therefore, considered as a starting point for larger GWAS, especially for recoverability as this is the first GWAS to report associated SNP variants and candidate genes.

Mastitis is not only polygenic but also multifactorial and different pathogens may invoke a different immune response (Oviedo-Boyso *et al.*, 2007). Studies (Holmberg *et al.*, 2012; Sørensen *et al.*, 2009; de Haas *et al.*, 2002b) show presence of genetic variation for pathogen-specific mastitis, implying that selection without the knowledge of the pathogens involved may be effective in reducing the incidence of mastitis due to specific pathogens but less effective in reducing the incidence of mastitis caused by other pathogens. Applying the methods and models to pathogen specific mastitis data could therefore be considered for future research and directions. We also hope the methods and models developed in this thesis will be further evaluated and applied to a wider context of disease data modelling. Adding the genetics of recoverability in the genetic evaluation of disease data could be of specific benefit in situations of high disease prevalence. Other endemic diseases with high prevalence that are negatively affecting the industry such as the porcine reproductive and respiratory syndrome virus in pigs or nematode infections in sheep could benefit from the methods and models developed in this thesis. PRRSV inflicts high negative economic impact in pig production by interfering in the reproductive performance of sows (Holtkamp *et al.*, 2013) and nematode infections negatively impacts weight gain, wool production and milk yield in sheep (reviewed by, Mavrot *et al.* (2015).

The thesis also leaves an open research topic on the models and methods for future investigation. Ideally, the transition-based traits definitions and methods (papers I-III) could be expanded to the more detailed and newly defined dynamic health classes presented in paper IV. The transition methods in combination with the newly defined health classes have the potential to capture more genetic variation and hence to further improve the genetic evaluation of udder health. Estimation of genetic parameters for- and among the multiple-transitions from/to *very healthy*, *normal healthy*, *short-term less healthy*, *short-*

term sick, *acute,* and *persistent* are important for efficient decision support and selection. Expanding the methods to multiple-transitions will most likely need (much) more data than the current data in papers II-IV. Getting large data is hopefully only a matter of time as more and more farms are using AMS (Barkema *et al.*, 2015) that are capable of cheaply and easily collecting data on a daily basis, or even at individual milkings. Hence there will be more data in the future and the true utility of these novel models is likely to increase.

In addition to the large phenotype data collected from AMS in the foreseeable future more cows are routinely genotyped as genotyping costs are rapidly coming down. The dairy cattle breeding industry is shifting to genomic selection (Boichard *et al.*, 2012; Hayes *et al.*, 2009). In the era of genomic selection, methods and models developed in this thesis could be instrumental to enhance the genetic and genomic evolution of health traits. Genomic selectin has particular advantageous features to the dairy cattle breeding industry and compared to the traditional (progeny testing proven bull-based) breeding programme, it could double the rate of genetic gain (Schefers & Weigel, 2012). Summarised by Schefers and Weigel (2012), the accelerated genetic gain from genomic selection is attributed to: 1) a greater accuracy of predicted genetic merit for young animals, 2) a shorter generation interval from the use of young animals, and 3) a higher selection intensity from testing large group of animals. Though sufficient and accurate phenotype is a requirement for genomic selection, once the reference population is well established, genomic selection is a novel tool to improve the genetics of traits that are more difficult/expensive to measure (e.g., mastitis).

6 Conclusions

Unlike the higher milk yield production resulted from genetic improvement as well as improved nutrition and management, response to selection against mastitis has not yet been a success story for the animal geneticists and breeders. One of the main factors for this is the genetic evaluation of mastitis is prone to loss of information, leading to low accuracy. This thesis has presented methods and models that can make use of more information by analysing the changes in SCC that each cow exhibits during lactation in both directions of the disease. From the methods and results presented in this thesis, we can make the following conclusions:

- The thesis has introduced methods and models for joint genetic evaluation of susceptibility to- and recoverability from mastitis.
- The methods and models are capable of correctly reproducing genetic parameters and they offer an option to estimate breeding value for both directions of the disease.
- The methods and models are applicable to real data and they produce and estimate genetic parameter consistent with literature values.
- Although susceptibility to- and recoverability from mastitis are strongly negatively correlated, recoverability is as heritable as susceptibility and could be considered a new trait for selection.
- The absence of strong association signal in the GWAS proves that mastitis is a complex trait and different functional genes could be involved over lactations and directions of the disease.
- This thesis is the first in reporting SNP-trait association particularly for recoverability from mastitis in Danish Holstein cows.
- This novel methods and modelling, if adopted, should enhance the genetic evaluation of mastitis by its ability to capture time-dependent and additional information from recoverability in the modelling and analyses of disease data, specifically mastitis.

7 Summary of the thesis

Mastitis in dairy cows is a common disease and has large negative economic and animal welfare consequences for the dairy industry. For this reason, mastitis has been the focus of many dairy cattle research projects over the last decades. However, in the genetic analyses, only susceptibility to contract mastitis has been considered, leaving aside the other aspect of the disease - the recoverability. Since mastitis is a frequent and unavoidable problem, evaluation of the ability of cows to overcome an infection should, therefore, be of interest. The aim of this thesis was to improve the genetic evaluation of udder health by introducing methods and models that can make use of the information contained in both directions of the disease (susceptibility and recoverability). To achieve this we have conducted four studies presented in the following four scientific papers:

In paper I, extensive simulation analyses were performed to develop a bivariate model for joint genetic evaluation of susceptibility to- and recoverability from mastitis. Mastitis incidences measured via somatic cell count (OCC) were simulated and a transition model was applied to define traits of interest. The mastitic or healthy states were converted to two series of transitions indicators: one for healthy to diseased (HD, to define mastitis susceptibility) and the other for diseased to healthy (DH, to define recoverability). The traits were analysed with a bivariate threshold sire model using the Bayesian statistics approach in DMU. Heritabilities for HD were consistent with the simulated (true) value (0.039). On the other hand, heritabilities for DH were a bit upward biased and this could be due to the fact that recoverability can only be observed if there has been mastitis, limiting the number of observations for the DH dataset. The average genetic correlations between HD and DH deviated by −0.055, 0.005, and 0.008 from the alternative simulated values of 0.0, 0.2, −0.2, respectively. The study demonstrated that

both traits can be modeled jointly and simulated correlations could be correctly reproduced.

In paper II, the model developed in paper I with an added time function as well as several systematic effects was applied to real data from Danish Holstein cows using automatic milking systems equipped with online somatic cell count (OCC) measuring units. The OCC were converted to elevated mastitis risk, a continuous variable (on a [0-1] scale) indicating the risk of mastitis. Risk values above 0.6 were assumed to indicate that a cow had mastitis. As in paper I, the mastitic or healthy states were converted to HD and DH. In addition to the bivariate threshold sire model, the traits were analysed with a threshold animal model for comparison. The heritabilities (ranging from 0.06 to 0.08) were of similar size across models and traits. The between trait genetic correlation was -0.83. Results indicated the presence of a genetic component in both traits and a strong genetic correlation.

In paper III, we performed genome-wide association studies to find positions on the genome that affect HD and/or DH. Phenotypic statistical analyses were performed to correct the phenotypes for different fixed and random effects. After quality control a total of 39378 autosomal single-nucleotide polymorphisms (SNPs) were used for the association analysis. Single SNP regression analysis was performed and substitution effect of each SNP was tested with a t-test. Contrary to the observed high negative genetic correlation (in paper II), association signals were mapped in different locations, suggesting that the traits could be regulated by different genes. Many of the SNP variants associated either with HD or DH were located in or very near to genes that have been reported for their role in the immune system. Genes involved in lymphocyte developments (e.g., *MAST*3 and *STAB*2) and genes involved in macrophage recruitment and regulation of inflammations (e.g., *PDGFD* and *PTX*3) were suggested as possible casual genes for HD and DH, respectively. However, complexity of the traits was manifested with the absence of strong association signals suggesting that numerous genes with small effects could be involved in both directions of the disease.

In paper IV, a more dynamic health classification which takes, severity of possible infection into account, was introduced. Considerable genetic variance was detected for cows' presence in health classes defined for longer periods, whereas the variations in health classes defined for short-term and sudden changes (e.g., acute) were mostly attributed to environmental factors.

Although susceptibility to- and recoverability from mastitis are strongly negatively correlated, recoverability has a similar size of genetic component as susceptibility and could be considered a new trait for selection. This thesis has introduced a method for joint estimation of breeding values for susceptibility

to- and recoverability from mastitis. Modelling and analyses of the genetics of recoverability could be of specific benefit in situations of high disease incidence. This novel modelling approach, if adopted, should enhance the genetic evaluation of mastitis by its ability of capturing time-dependent and additional information from recovery in the modelling and analyses.

8 Sammanfattning

Mastit (juverinflammation) hos mjölkkor är en vanligt förekommande sjukdom som har stora negativa ekonomiska och välfärdsmässiga konsekvenser för mejeriindustrin. Mastit har därför varit fokus för många forskningsprojekt de senaste årtiondena. Genetiska analyser av juverhälsa har hittills fokuserat på kons mottaglighet för mastits. Den andra aspekten av sjukdomen, kons tillfrisknande, har tidigare varit outforskad. Eftersom mastit är en vanligt förekommande och oundviklig sjukdom, borde kons återhämtningsförmåga efter en infektion vara av intresse. Målet med denna avhandling var att förbättra den genetiska evalueringen av juverhälsa genom att introducera metoder och modeller som kan använda information från sjukdomens hela utveckling, både mottaglighet och tillfrisknande. Vi har utfört fyra studier som är presenterade i fyra vetenskapliga artiklar.

I artikel I utfördes omfattande simuleringsanalyser för att utveckla en bivariat modell för simultan genetisk evaluering av mottaglighet för- och tillfrisknande från mastit. Fall av mastit, klassificerade genom celltal, simulerades och en övergångsmodell (transition model) användes för att definiera de önskade egenskaperna. Sjuka och friska tillstånd omvandlades till två typer av övergångsindikatorer: en från frisk till sjuk (HD, definition av mottaglighet för mastit) och en annan från sjuk till frisk (DH, definition av tillfrisknande från mastis). Egenskaperna blev analyserade med en bivariat, tröskel- och tjurmodell med hjälp av Bayesianska statistiska funktioner i DMU. Arvbarhet för HD sammanföll med det simulerade sanna värdet (0.039). Däremot var arvbarheten för DH lite för hög. Detta kan möjligtvis förklaras med att tillfrisknande bara kan observeras om en ko fått mastit, vilket begränsar antalet observationer för DH. Den genomsnittliga genetiska korrelationen mellan HD och DH avvek med -0.055, 0.005 och 0.008 från de simulerade värdena på 0.0, 0.2, och -0.2. Studien visade att båda egenskaperna

kan modelleras simultant och att simulerade korrelationer kan bli korrekt återgivna.

I artikel II vidareutvecklades modellen från artikel I med en tidsfunktion och flera systematiska effekter och användes sedan på data från danska Holsteinkor mjölkade i automatiska mjölkningssystem utrustade med enheter för mätning av celltal (OCC) online. Mätningarna blev omvandlade till ”ökad mastitrisk”, en kontinuerlig variabel på en [0-1] skala som indikerar risken för mastit. Värden över 0.6 antogs indikera att en ko hade mastit. Precis som i artikel I blev sjuka och friska tillstånd konverterade till HD och DH. I tillägg till den bivariata, tröskel- och tjurmodellen blev egenskaperna också analyserade med en tröskel- och djurmodell i jämförande syfte. Arvbarheterna (0.06 till 0.08) var i samma storleksordning för alla modeller och egenskaper. Den genetiska korrelationen mellan egenskaperna var -0.83. Resultaten indikerar förekomsten av en genetisk komponent för båda egenskaperna samt en stark genetisk korrelation mellan dem.

I artikel III utfördes en GWAS studie (genome-wide association study) för att hitta positioner på genomet som påverkar HD och/eller DH. Fenotypiska statistiska analyser utfördes för att justera fenotyperna för olika systematiska och slumpmässiga effekter. Efter kvalitetskontroll återstod 39378 autosoma SNP (single nucleotide polymorphism) som användes för associationsanalyser. Enkel SNP regressionsanalys utfördes och substitutionseffekter för varje enskild SNP testades för signifikans med hjälp av t-test. I motsats till de observerade höga negativa korrelationerna (i artikel II) blev associationssignalerna lokaliserade på olika platser. Detta indikerar att egenskaperna kan vara reglerade av olika gener. Många av SNP varianterna associerade med HD eller DH er lokaliserade på eller mycket nära gener som är kända för deras påverkan på immunsystemet. Gener involverade i lymfocytutveckling (t.ex. *MAST3* och *STAB2*) och gener involverade i makrofagrekrytering och reglering av inflammation (t.ex. *PDGFD* and *PTX3*) föreslogs som möjliga kausala gener för HD respektive DH. Emellertid ger egenskapernas komplexitet utslag i bristen på starka associationssignaler vilket indikerar att åtskilliga gener med liten effekt kan vara involverade i sjukdomens båda riktningar (HD och DH).

I artikel IV blev en mer dynamisk hälsoklassificering introducerad, vilken tar hänsyn till graden av infektion. Betydande genetisk varians påvisades för en kos närvaro i hälsoklasser definierade för långa perioder, medan variationen i hälsoklasser definierade för korta perioder och plötsliga ändringar (t.ex. akut mastit) mest kunde förklaras av miljömässiga faktorer.

Även om mottaglighet för- och tillfrisknande från mastit är starkt negativt korrelerade så är tillfrisknandet påverkat av gener i samma grad som mottagligheten och kan introduceras som en ny egenskap för genetisk

selektion. Den här avhandlingen har introducerat en metod för simultan estimering av avelsvärden för mottaglighet för- och tillfrisknande från mastit. Modellering och analyser av gentiken bakom tillfrisknandet kan vara särskilt fördelaktigt i situationer med hög incidens av mastit. Om denna nya modelleringsansats införs, kan den förbättra den genetiska evalueringen av mastit genom sin förmåga att tillvarata tidsberoende samt utökad information från tillfrisknandet i modeller och analyser.

9 Sammendrag

Mastitis hos malkekøer er almindeligt forekommende og har store konsekvenser for malkekvægssektoren både økonomisk og dyrevelfærdsmæssigt. Mastitis har derfor været omdrejningspunktet for adskillige forskningsprojekter i de seneste årtier. Genetiske analyser af yversundhedsegenskaber har indtil videre udelukkende fokuseret på koens modtagelighed for mastitis. Derved overses den anden side af sygdommen, nemlig det at blive rask. Da mastitis er et hyppigt og uundgåelig problem burde koens evne til at komme sig efter en infektion også være af interesse. Formålet med denne afhandling var at forbedre den avlsmæssige evaluering af yversundhed ved at introducere metoder og modeller, som kan gøre brug af information på begge sider af et sygdomstilfælde, modtagelighed og raskmelding. I den forbindelse har vi udført fire studier, som er præsenteret i fire videnskabelige artikler:

I artikel I blev der udført omfattende simuleringsanalyser som grundlag for udviklingen af en bivariat model til samtidig genetisk evaluering af modtagelig og raskmelding for mastitis. Mastitis incidens målt på baggrund af celletal blev simuleret, og en overgangsmodel (transition model) blev anvendt til at definere de ønskede egenskaber. Henholdsvis syge og raske tilstande blev omdannet til to typer overgangsindikatorer: én for rask til syg (HD, til definition af mastitismodtagelighed) og en anden for syg til rask (DH, til definition af raskmelding). Egenskaberne blev analyseret med en bivariat tærskel- og tyremodel ved hjælp af den bayesianske statistiske funktionalitet i DMU. Arvbarheder for HD var sammenfaldende med de simulerede (sande) værdier (0.039). På den anden side var arvbarhederne for DH en smule for høje. Dette kan muligvis forklares ved at raskmelding kun kan observeres, hvis der har været et forudgående mastitistilfælde. Derved begrænses antallet af observationer i DH datasættet. Den gennemsnitlige genetiske korrelation mellem HD og DH afveg med hhv. -0.055, 0.005 og 0.008 for de simulerede

værdier på hhv. 0.0, 0.2 og -0.2. Undersøgelsen demonstrerede at begge egenskaber kan modelleres samtidig og at simulerede korrelationer i høj grad kan genfindes.

I artikel II blev modellen fra artikel I udvidet med en tidsfunktion og flere systematiske effekter for derefter at blive anvendt på data fra danske Holstein køer malket i automatiske malkesystemer påsat udstyr til online måling af celletal (OCC). Disse målinger blev omdannet til "øget mastitis risiko" – en kontinuert variabel, på en [0-1] skala, som indikerer risikoen for mastitis. Det blev antaget, at en ko havde mastitis, når risikoen var over 0.6. Ligesom i artikel I blev syge og raske tilstande konverteret til hhv. HD og DH. Udover en bivariat tærskel- og tyremodel blev egenskaberne også analyseret vha. tærskel- og dyremodel. Arvbarhederne (0.06 til 0.08) var ens på tværs af modeller og egenskaber. Den genetiske korrelation mellem de to egenskaber var på -0.83. Resultaterne antyder tilstedeværelse af en genetisk komponent for begge egenskaber og en stærk genetisk korrelation.

I artikel III udførte vi en GWAS undersøgelse (genome-wide association study) for at fastslå områder på genomet som påvirker HD og/eller DH. Fænotypiske, statistiske analyser blev udført for at justere fænotyperne for forskellige systematiske og tilfældige effekter. I alt 39,378 autosomale SNP (single nucleotide polymorphisms) var til rådighed for associationsanalyserne efter kvalitetskontrol. Enkel SNP regressionsanalyse blev udført og substitutionseffekter for hver enkel SNP blev testet for signifikans med t-test. I modsætning til de den stærke negative korrelation (i artikel II) blev associationssignalerne lokaliseret forskellige steder. Dette antyder, at egenskaberne kan være reguleret af forskellige gener. Mange af SNP varianterne associeret med hhv. HD eller DH er lokaliseret meget tæt på gener, som er kendt for deres påvirkning af immunsystemet. Gener involveret i lymfocytudvikling (f.eks. *MAST3* og *STAB2*) og gener involveret i makrofagrekruttering og regulering af betændelsestilstande (f.eks. *PDGFD* og *PTX3*) blev foreslået som mulige kausale gener. Imidlertid giver egenskabernes kompleksitet sig udslag i manglen på stærke associationssignaler. Dette antyder at adskillige gener kan være involveret på både modtagelighedens og raskmeldingens side af et mastitistilfælde.

I artikel IV blev en mere dynamisk yversundhedsklassification introduceret. Der blev fundet betydelig genetisk varians for en kos tilstedeværelse i en klasse af længere varighed sammenlignet med klasser af kortere varighed og pludselige ændringer (f.eks. akut mastitis). De sidstnævnte kunne mest forklares vha. miljømæssige faktorer.

Selvom modtagelighed for og raskmelding fra mastitis er stærkt negativt korrelerede, er raskmelding påvirket af gener i samme størrelsesorden som

modtagelighed og kan introduceres som en ny egenskab for genetisk selektion. Denne afhandling har introduceret en metode til samtidig estimering af avlsværdier for modtagelighed for og raskmelding fra mastitis. Modellering og analyser af genetikken bag raskmelding kan være særlig fordelagtig i situationer med høj mastitisincidens. Dette nye modelleringsaspekt, hvis indført, kunne forbedre den genetiske evaluering for yversundehed pga. evnen til at opfange ekstra og tidsafhængig information fra raskmelding i modellering og analyser.

References

Abdel-Shafy, H., Bortfeldt, R.H., Reissmann, M. & Brockmann, G.A. (2014). Short communication: Validation of somatic cell score-associated loci identified in a genome-wide association study in German Holstein cattle. *Journal of Dairy Science,* 97(4), ss. 2481-2486.

Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Banet, J.F., Billis, K., Giron, C.G., Hourlier, T., Howe, K., Kahari, A., Kokocinski, F., Martin, F.J., Murphy, D.N., Nag, R., Ruffier, M., Schuster, M., Tang, Y.A., Vogel, J.H., White, S., Zadissa, A., Flicek, P. & Searle, S.M.J. (2016). The Ensembl gene annotation system. *Database-the Journal of Biological Databases and Curation*.

Barkema, H.W., von Keyserlingk, M.A., Kastelic, J.P., Lam, T.J., Luby, C., Roy, J.P., LeBlanc, S.J., Keefe, G.P. & Kelton, D.F. (2015). Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. *Journal of Dairy Science,* 98(11), ss. 7426-45.

Bennedsgaard, T.W., Enevoldsen, C., Thamsborg, S.M. & Vaarst, M. (2003). Effect of mastitis treatment and somatic cell counts on milk yield in Danish organic dairy cows. *Journal of Dairy Science,* 86(10), ss. 3174-83.

Bishop, S.C. & Woolliams, J.A. (2010). On the Genetic Interpretation of Disease Data. *PLoS One,* 5(1).

Boichard, D., Ducrocq, V., Croiseau, P. & Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies,* 339(7-8), ss. 274-7.

Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M.N., Boscher, M.Y., Druet, T., Genestout, L., Colleau, J.J., Journaux, L., Ducrocq, V. & Fritz, S. (2012). Genomic selection in French dairy cattle. *Animal Production Science,* 52(2-3), ss. 115-120.

Brotherstone, S. & Goddard, M. (2005). Artificial selection and maintenance of genetic variance in the global dairy cow population. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences,* 360(1459), ss. 1479-88.

Campbell, J.R. & Marshall, R.T. (2016). *Dairy Production and Processing: The Science of Milk and Milk Products*: Waveland Press.

Carlén, E. (2008). Genetic Evaluation of Clinical Mastitis in Dairy Cattle. *PhD Thesis*. Swedish University of Agricultural Sciences, Uppsala.

Carlén, E., Emanuelson, U. & Strandberg, E. (2006). Genetic Evaluation of Mastitis in Dairy Cattle Using Linear Models, Threshold Models, and Survival Analysis: A Simulation Study. *Journal of Dairy Science,* 89(10), ss. 4049-4057.

Carlén, E., Schneider, M.D. & Strandberg, E. (2005). Comparison between linear models and survival analysis for genetic evaluation of clinical mastitis in dairy cattle. *Journal of Dairy Science,* 88(2), ss. 797-803.

de Haas, Y. (2003). Somatic cell count patterns: Improvement of udder health by genetics and management. *PhD Thesis*. Wageningen University.

de Haas, Y., Barkema, H.W. & Veerkamp, R.F. (2002a). The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. *Journal of Dairy Science,* 85(5), ss. 1314-1323.

de Haas, Y., Barkema, H.W. & Veerkamp, R.F. (2002b). Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. *Animal Science,* 74, ss. 233-242.

de Haas, Y., Ouweltjes, W., ten Napel, J., Windig, J.J. & de Jong, G. (2008). Alternative somatic cell count traits as mastitis indicators for genetic selection. *Journal of Dairy Science,* 91(6), ss. 2501-2511.

Dekkers, J.C.M., Boettcher, P.J. & Mallard, B.A. (1998). Genetic improvement of udder health. *6th World Congress on Genetics Applied to Livestock Production. Armidale, Australia, 25:3-10.* .

Djabri, B., Bareille, N., Beaudeau, F. & Seegers, H. (2002). Quarter milk somatic cell count in infected dairy cows: a meta-analysis. *Veterinary Research,* 33(4), ss. 335-57.

Dohoo, I.R., Smith, J., Andersen, S., Kelton, D.F., Godden, S. & Mastitis Research Workers, C. (2011). Diagnosing intramammary infections: evaluation of definitions based on a single milk sample. *Journal of Dairy Science,* 94(1), ss. 250-61.

Fogsgaard, K.K. (2015). Recovery from Mastitis in Dairy Cows – Development of Behaviour, Milk Production and Inflammatory Markers in the Weeks during and after Naturally Occurring Clinical Mastitis. *PhD Thesis*.: Aarhus University, Department of Animal Science - Behaviour and stressbiology.

Fogsgaard, K.K., Bennedsgaard, T.W. & Herskin, M.S. (2015a). Behavioral changes in freestall-housed dairy cows with naturally occurring clinical mastitis. *Journal of Dairy Science,* 98(3), ss. 1730-1738.

Fogsgaard, K.K., Løvendahl, P., Bennedsgaard, T.W. & Østergaard, S. (2015b). Changes in milk yield, lactate dehydrogenase, milking frequency, and interquarter yield ratio persist for up to 8 weeks after antibiotic treatment of mastitis. *Journal of Dairy Science,* 98(11), ss. 7686-98.

Franzén, J., Thorburn, D., Urioste, J.I. & Strandberg, E. (2012). Genetic evaluation of mastitis liability and recovery through longitudinal analysis of transition probabilities. *Genetics Selection Evolution,* 44(1).

Fukuda, H., Nakamura, N. & Hirose, S. (2006). MARCH-III is a novel component of endosomes with properties similar to those of MARCH-II. *Journal of Biochemistry,* 139(1), ss. 137-145.

Gernand, E. & Konig, S. (2014). Random regression test-day model for clinical mastitis: Genetic parameters, model comparison, and correlations with indicator traits. *Journal of Dairy Science,* 97(6), ss. 3953-63.

Gernand, E., Rehbein, P., von Borstel, U.U. & Konig, S. (2012). Incidences of and genetic parameters for mastitis, claw disorders, and common health traits recorded in dairy cattle contract herds. *Journal of Dairy Science,* 95(4), ss. 2144-56.

Gianola, D. & Rosa, G.J. (2015). One hundred years of statistical developments in animal breeding. *Annual Review of Animal Biosciences,* 3, ss. 19-56.

Goddard, M.E. (2009). View to the future: could genomic evaluation become the standard? *Interbull Bulletin (39), pp. 83-88*.

Goddard, M.E. & Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics,* 10(6), ss. 381-391.

Govignon-Gion, A., Dassonneville, R., Baloche, G. & Ducrocq, V. (2012). Genetic evaluation of mastitis in dairy cattle in France. *Interbull bulletin*(46), ss. 121-126.

Govignon-Gion, A., Dassonneville, R., Baloche, G. & Ducrocq, V. (2016). Multiple trait genetic evaluation of clinical mastitis in three dairy cattle breeds. *Animal,* 10(4), ss. 558-65.

Green, M.J., Bradley, A.J., Medley, G.F. & Browne, W.J. (2007). Cow, farm, and management factors during the dry period that determine the rate of clinical mastitis after calving. *Journal of Dairy Science,* 90(8), ss. 3764-3776.

Grohn, Y.T., Eicker, S.W., Ducrocq, V. & Hertl, J.A. (1998). Effect of diseases on the culling of Holstein dairy cows in New York State. *Journal of Dairy Science,* 81(4), ss. 966-78.

Hadfield, J.D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software,* 33(2), ss. 1-22.

Hagnestam-Nielsen, C. & Ostergaard, S. (2009). Economic impact of clinical mastitis in a dairy herd assessed by stochastic simulation using different methods to model yield losses. *Animal,* 3(2), ss. 315-328.

Halasa, T., Huijps, K., Osteras, O. & Hogeveen, H. (2007). Economic effects of bovine mastitis and mastitis management: a review. *The Veterinary Quarterly,* 29(1), ss. 18-31.

Halasa, T., Nielen, M., De Roos, A.P.W., Van Hoorne, R., de Jong, G., Lam, T.J.G.M., van Werven, T. & Hogeveen, H. (2009). Production loss due to new subclinical mastitis in Dutch dairy cows estimated with a test-day model (vol 92, pg 599, 2009). *Journal of Dairy Science,* 92(3), ss. 1315-1315.

Hansson, H. & Lagerkvist, C.J. (2014). Decision making for animal health and welfare: integrating risk-benefit analysis with prospect theory. *Risk Analysis,* 34(6), ss. 1149-59.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges (vol 92, pg 433, 2009). *Journal of Dairy Science,* 92(3), ss. 1313-1313.

Hayes, B.J., Pryce, J., Chamberlain, A.J., Bowman, P.J. & Goddard, M.E. (2010). Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. *Plos Genetics,* 6(9).

Heringstad, B., Klemetsdal, G. & Ruane, J. (2000). Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. *Livestock Production Science,* 64(2-3), ss. 95-106.

Hiitio, H., Vakkamaki, J., Simojoki, H., Autio, T., Junnila, J., Pelkonen, S. & Pyorala, S. (2017). Prevalence of subclinical mastitis in Finnish dairy cows: changes during recent decades and impact of cow and herd factors. *Acta Veterinaria Scandinavica,* 59.

Hirschhorn, J.N. & Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Review Genetics,* 6(2), ss. 95-108.

Hogan, J.S. & Smith, K.L. (1987). A practical look at environmental mastitis. *Compendium on Continuing Education for the Practicing Veterinarian,* 9(10), ss. F341-F344, F346.

Hogeveen, H., Huijps, K. & Lam, T.J. (2011). Economic aspects of mastitis: new developments. *New Zealand Veterinary Journal,* 59(1), ss. 16-23.

Holmberg, M., Fikse, W.F., Andersson-Eklund, L., Artursson, K. & Lunden, A. (2012). Genetic analyses of pathogen-specific mastitis. *Journal of Animal Breeding and Genetics,* 129(2), ss. 129-137.

Holtkamp, D.J., Kliebenstein, J.B., Neumann, E.J., Zimmerman, J.J., Rotto, H.F., Yoder, T.K., Wang, C., Yeske, P.E., Mowrer, C.L. & Haley, C.A. (2013). Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *Journal of Swine Health and Production,* 21(2), ss. 72-84.

Howard, J.T., Jiao, S.H., Tiezzi, F., Huang, Y.J., Gray, K.A. & Maltecca, C. (2015). Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *Bmc Genetics,* 16.

Hutchison, J., Cole, J. & Bickhart, D. (2014). Use of young bulls in the United States. *Journal of Dairy Science,* 97(5), ss. 3213-3220.

International Dairy Federation (1987). Bovine mastitis : definition and guidelines for diagnosis. *Bulletin of the International Dairy Federation 211, 24 pp.*

International Dairy Federation (2011). Suggested interpretation of mastitis terminology. *Bulletin of the International Dairy Federation. Brussels, Belgium; 2011 (448/2011).*

Jamrozik, J., Koeck, A., Miglior, F., Kistemaker, G.J., Schenkel, F.S., Kelton, D.F., Doormaal, B.J.v. & van Doormaal, B.J. (2013). Genetic and genomic evaluation of mastitis resistance in Canada. *Interbull bulletin*(47), ss. 43-51.

Jamrozik, J. & Schaeffer, L.R. (2012). Test-day somatic cell score, fat-to-protein ratio and milk yield as indicator traits for sub-clinical mastitis in dairy cattle. *Journal of Animal Breeding and Genetics,* 129(1), ss. 11-19.

Kehrli, M.E. & Shuster, D.E. (1994). Factors Affecting Milk Somatic-Cells and Their Role in Health of the Bovine Mammary-Gland. *Journal of Dairy Science,* 77(2), ss. 619-627.

Koeck, A., Heringstad, B., Egger-Danner, C., Fuerst, C., Winter, P. & Fuerst-Waltl, B. (2010). Genetic analysis of clinical mastitis and somatic cell count traits in Austrian Fleckvieh cows. *Journal of Dairy Science,* 93(12), ss. 5987-5995.

Kuang, Y., Tani, K., Synnott, A.J., Ohshima, K., Higuchi, H., Nagahata, H. & Tanji, Y. (2009). Characterization of bacterial population of raw milk from bovine mastitis by culture-independent PCR-DGGE method. *Biochemical Engineering Journal,* 45(1), ss. 76-81.

Lassen, J., Hansen, M., Sorensen, M.K., Aamand, G.P., Christensen, L.G. & Madsen, P. (2003). Genetic relationship between body condition score, dairy character, mastitis, and diseases other than mastitis in first-parity Danish Holstein cows. *Journal of Dairy Science,* 86(11), ss. 3730-3735.

Lipschutz-Powell, D., Woolliams, J.A., Bijma, P. & Doeschl-Wilson, A.B. (2012). Indirect Genetic Effects and the Spread of Infectious Disease: Are We Capturing the Full Heritable Variation Underlying Disease Prevalence? *PLoS One,* 7(6).

Lund, T., Miglior, F., Dekkers, J.C.M. & Burnside, E.B. (1994). Genetic-Relationships between Clinical Mastitis, Somatic-Cell Count, and Udder Conformation in Danish Holsteins. *Livestock Production Science,* 39(3), ss. 243-251.

Løvendahl, P. & Sørensen, L.P. (2016). Frequently recorded sensor data may correctly provide health status of cows if data are handled carefully and errors are filtered away. *Biotechnologie Agronomie Societe Et Environnement,* 20(1), ss. 3-12.

Madsen, P. & Jensen, J. (2013). A user's Guide to DMU. A Packagefor Analysing Multivariate Mixed Models. Version 6, release 5.2. University of Aarhus, Faculty Agricultural Sciences (DJF), Department of Genetics and Biotechnology, Research Centre Foulum, Tjele, Denmark.

Mao, X. (2016). Population level genome-wide association studies in dairy cattle. *Ph.D. Thesis.* Science and technology. Aarhus University.

Mavrot, F., Hertzberg, H. & Torgerson, P. (2015). Effect of gastro-intestinal nematode infection on sheep performance: a systematic review and meta-analysis. *Parasit Vectors,* 8, s. 557.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology,* 17(1), s. 122.

Meredith, B.K., Kearney, F.J., Finlay, E.K., Bradley, D.G., Fahey, A.G., Berry, D.P. & Lynn, D.J. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *Bmc Genetics,* 13.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics,* 157(4), ss. 1819-1829.

Miglior, F., Muir, B.L. & Van Doormaal, B.J. (2005). Selection indices in Holstein cattle of various countries (vol 88, pg 1255, 2005). *Journal of Dairy Science,* 88(4), ss. 1613-1613.

Negussie, E., Stranden, I. & Maentysaari, E.A. (2008). Genetic association of clinical mastitis with test-day somatic cell score and milk yield during first lactation of Finnish Ayrshire cows. *Journal of Dairy Science,* 91(3), ss. 1189-1197.

Nielsen, C. (2009). Economic impact of mastitis in dairy cows. *PhD Thesis.* Swedish University of Agricultural Sciences, Uppsala, Sweden.

Nielsen, C. & Emanuelson, U. (2013). Mastitis control in Swedish dairy herds. *Journal of Dairy Science,* 96(11), ss. 6883-93.

Norberg, E. (2005). Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review. *Livestock Production Science,* 96(2-3), ss. 129-139.

Nyman, A.K., Ekman, T., Emanuelson, U., Gustafsson, A.H., Holtenius, K., Waller, K.P. & Sandgren, C.H. (2007). Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. *Preventive Veterinary Medicine,* 78(2), ss. 142-160.

Nyman, A.K., Persson Waller, K., Bennedsgaard, T.W., Larsen, T. & Emanuelson, U. (2014). Associations of udder-health indicators with cow factors and with intramammary infection in dairy cows. *Journal of Dairy Science,* 97(9), ss. 5459-73.

Oltenacu, P.A. & Broom, D.M. (2010). The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Animal Welfare,* 19, ss. 39-49.

Osteras, O., Solbu, H., Refsdal, A.O., Roalkvam, T., Filseth, O. & Minsaas, A. (2007). Results and evaluation of thirty years of health recordings in the Norwegian dairy cattle population. *Journal of Dairy Science,* 90(9), ss. 4483-97.

Oviedo-Boyso, J., Valdez-Alarcon, J.J., Cajero-Juarez, M., Ochoa-Zarzosa, A., Lopez-Meza, J.E., Bravo-Patino, A. & Baizabal-Aguirre, V.M. (2007). Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. *Journal of Infection,* 54(4), ss. 399-409.

Pantoja, J.C.F., Hulland, C. & Ruegg, P.L. (2009). Somatic cell count status across the dry period as a risk factor for the development of clinical mastitis in the subsequent lactation. *Journal of Dairy Science,* 92(1), ss. 139-148.

Parker Gaddis, K.L., Cole, J.B., Clay, J.S. & Maltecca, C. (2014). Genomic selection for producer-recorded health event data in US dairy cattle. *Journal of Dairy Science,* 97(5), ss. 3190-9.

Philipsson, J. & Lindhe, B. (2003). Experiences of including reproduction and health traits in Scandinavian dairy cattle breeding programmes. *Livestock Production Science,* 83(2-3), ss. 99-112.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* 81(3), ss. 559-575.

R Core Team (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* http://www.R-project.org/.

Ramirez, N.F., Keefe, G., Dohoo, I., Sanchez, J., Arroyave, O., Ceron, J., Jaramillo, M. & Palacio, L.G. (2014). Herd- and cow-level risk factors associated with subclinical mastitis in dairy farms from the High Plains of the northern Antioquia, Colombia. *Journal of Dairy Science,* 97(7), ss. 4141-50.

Rollin, E., Dhuyvetter, K.C. & Overton, M.W. (2015). The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Preventive Veterinary Medicine,* 122(3), ss. 257-64.

Rupp, R. & Boichard, D. (2003). Genetics of resistance to mastitis in dairy cattle. *Veterinary Research,* 34(5), ss. 671-688.

Sahana, G., Guldbrandtsen, B., Thomsen, B. & Lund, M.S. (2013). Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. *Animal Genetics,* 44(6), ss. 620-626.

Sargeant, J.M., Scott, H.M., Leslie, K.M., Ireland, M.J. & Bashiri, A. (1998). Clinical mastitis in dairy cattle in Ontario: Frequency of occurrence and bacteriological isolates (vol 39, pg 33, 1998). *Canadian Veterinary Journal-Revue Veterinaire Canadienne,* 39(4), ss. 240-240.

Schaeffer, L.R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics,* 123(4), ss. 218-23.

Schefers, J.M. & Weigel, K.A. (2012). Genomic selection in dairy cattle: integration of DNA testing into breeding programs. *Animal Frontiers,* 2(1), ss. 4-9.

Schledzewski, K., Geraud, C., Arnold, B., Wang, S.J., Grone, H.J., Kempf, T., Wollert, K.C., Straub, B.K., Schirmacher, P., Demory, A., Schonhaber, H., Gratchev, A., Dietz, L., Thierse, H.J., Kzhyshkowska, J. & Goerdt, S. (2011). Deficiency of liver sinusoidal scavenger receptors stabilin-1 and-2 in mice causes glomerulofibrotic nephropathy via impaired hepatic clearance of noxious blood factors. *Journal of Clinical Investigation,* 121(2), ss. 703-714.

Schroeder, J. (2012). Bovine Mastitis and Milking Management. North Dakota State University. https://www.ag.ndsu.edu/pubs/ansci/dairy/as1129.pdf

Schukken, Y.H., Gunther, J., Fitzpatrick, J., Fontaine, M.C., Goetze, L., Holst, O., Leigh, J., Petzl, W., Schuberth, H.J., Sipka, A., Smith, D.G., Quesnell, R., Watts, J., Yancey, R., Zerbe, H., Gurjar, A., Zadoks, R.N., Seyfert, H.M. & members of the Pfizer mastitis research, c. (2011). Host-response patterns of intramammary infections in dairy cows. *Veterinary Immunology and Immunopathology,* 144(3-4), ss. 270-89.

Schukken, Y.H., Wilson, D.J., Welcome, F., Garrison-Tikofsky, L. & Gonzalez, R.N. (2003). Monitoring udder health and milk quality using somatic cell counts. *Veterinary Research,* 34(5), ss. 579-596.

Seegers, H., Fourichon, C. & Beaudeau, F. (2003). Production effects related to mastitis and mastitis economics in dairy cattle herds. *Veterinary Research,* 34(5), ss. 475-91.

Seiler, C., Gebhart, N., Zhang, Y., Shinton, S.A., Li, Y.S., Ross, N.L., Liu, X.J., Li, Q., Bilbee, A.N., Varshney, G.K., LaFave, M.C., Burgess, S.M., Balciuniene, J., Balciunas, D., Hardy, R.R., Kappes, D.J., Wiest, D.L. & Rhodes, J. (2015). Mutagenesis Screen Identifies agtpbp1 and eps15L1 as Essential for T lymphocyte Development in Zebrafish. *PLoS One,* 10(7).

Sender, G., Korwin-Kossakowska, A., Pawlik, A., Hameed, K.G.A. & Oprzadek, J. (2013). Genetic basis of mastitis resistance in dairy cattle - a review. *Annals of Animal Science,* 13(4), ss. 663-673.

Shook, G.E. & Schutz, M.M. (1994). Selection on somatic cell score to improve resistance to mastitis in the United States. *Journal of Dairy Science,* 77(2), ss. 648-58.

Strillacci, M.G., Frigo, E., Schiavini, F., Samore, A.B., Canavesi, F., Vevey, M., Cozzi, M.C., Soller, M., Lipkin, E. & Bagnato, A. (2014). Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. *Bmc Genetics,* 15.

Suriyasathaporn, W., Schukken, Y.H., Nielen, M. & Brand, A. (2000). Low somatic cell count: a risk factor for subsequent clinical mastitis in a dairy herd. *Journal of Dairy Science,* 83(6), ss. 1248-55.

Sørensen, L.P., Bjerring, M. & Løvendahl, P. (2016). Monitoring individual cow udder health in automated milking systems using online somatic cell counts. *Journal of Dairy Science,* 99(1), ss. 608-620.

Sørensen, L.P., Mark, T., Madsen, P. & Lund, M.S. (2009). Genetic correlations between pathogen-specific mastitis and somatic cell count in Danish Holsteins. *Journal of Dairy Science,* 92(7), ss. 3457-3471.

Thompson-Crispi, K., Atalla, H., Miglior, F. & Mallard, B.A. (2014). Bovine mastitis: frontiers in immunogenetics. *Frontiers in Immunology,* 5, s. 493.

Urioste, J.I., Franzén, J., Windig, J.J. & Strandberg, E. (2012). Genetic relationships among mastitis and alternative somatic cell count traits in the first 3 lactations of Swedish Holsteins. *Journal of Dairy Science,* 95(6), ss. 3428-3434.

Uutela, M., Wirzenius, M., Paavonen, K., Rajantie, I., He, Y.L., Karpanen, T., Lohela, M., Wiig, H., Salven, P., Pajusola, K., Eriksson, U. & Alitalo, K. (2004). PDGF-D induces macrophage recruitment, increased interstitial pressure, and blood vessel maturation during angiogenesis. *Blood,* 104(10), ss. 3198-3204.

Van Doormaal, B. & Beavers, L. (2014). Mastitis Resistance Selection: Now a Reality! *Available: https://www.cdn.ca/articles.php*.

Watts, J.L. (1988). Etiological agents of bovine mastitis. *Veterinary Microbiology,* 16(1), ss. 41-66.

Vazquez, A.I., Gianola, D., Bates, D., Weigel, K.A. & Heringstad, B. (2009). Assessment of Poisson, logit, and linear models for genetic analysis of

clinical mastitis in Norwegian Red cows. *Journal of Dairy Science,* 92(2), ss. 739-48.

Vazquez, A.I., Perez-Cabal, M.A., Heringstad, B., Rodrigues-Motta, M., Rosa, G.J.M., Gianola, D. & Weigel, K.A. (2012). Predictive ability of alternative models for genetic analysis of clinical mastitis. *Journal of Animal Breeding and Genetics,* 129(2), ss. 120-128.

Viguier, C., Arora, S., Gilmartin, N., Welbeck, K. & O'Kennedy, R. (2009). Mastitis detection: current trends and future perspectives. *Trends in Biotechnology,* 27(8), ss. 486-93.

Wilmink, J. (1987). Adjustment of test-day milk, fat and protein yields for age, season and stage of lactation. *Livestock Production Science,* 16, ss. 335-348.

Wojdak-Maksymiec, K., Szyda, J. & Strabel, T. (2013). Parity-dependent association between TNF-alpha and LTF gene polymorphisms and clinical mastitis in dairy cattle. *Bmc Veterinary Research,* 9.

Wu, J., Hu, S. & Cao, L. (2007). Therapeutic effect of nisin Z on subclinical mastitis in lactating cows. *Antimicrobial Agents and Chemotherapy,* 51(9), ss. 3131-5.

Acknowledgements

After a long and (unexpectedly) challenging journey, it is a real excitement to find myself writing this doctoral thesis. I would like to thank all the people who enabled me to make this thesis a reality.

DJ de Koning, comes first, really an outstanding and very compassionate supervisor! Your immense contributions and encouraging guidance throughout the course of my doctoral studies was extraordinary. You were there always to bring a solution even at times when I was not a self-starter in solving my own problems. I was always impressed with your swift and helpful response both in academic and administrative matters. Thank you very much!

Luc Janss, I would like to extend my sincere appreciation for your endless will to give a hand! I have really appreciated your support and patience. Thank you for all the positive contribution and thank you for the comments and suggestions that helped to significantly improve the manuscripts.

Freddy Fikse, I would like to extend my sincere appreciation for your contributions in data analyses and interpretation of results. Without your academic help in analyses, it would have been difficult to publish.

Jessica Franzén, many thanks to you! Your guidance in the very initial stage and in the very last stage of the PhD programme have helped me a lot. Your contribution and suggestions in writing this thesis was invaluable and most importantly timely. I am very grateful for the time you spent in reading and editing my thesis during your summer time and it was mainly this kind of compassionate help that enabled me to complete this thesis. A big thank you!

Peter Løvendahl, I would like to express my sincere appreciation. Your comments and suggestions were invaluable and (always) very encouraging. It was your words ‘as a first draft it is very good, this seems to work!’ that really pushed me for more analyses and to put more effort in all the projects. Your extraordinary contribution during meetings and in interpretation of results and paper writings was rather a lot and invaluable! Thank you very much Peter!

Lars Peter, I’m sincerely grateful for your kind help in preparing and editing the data for me. Your help in the modelling and analyses was also helpful. I am very grateful for the opportunity you gave me to work with you especially at the last stage of my doctoral studies.

Next to my supervisors, I would like to thank other people who have positively contributed to this thesis from its inception to completion.

Erling Strandberg and Anne Lundén, I am sincerely grateful for both of you for selecting me for the prestigious Erasmus Mundus scholarship and for giving me the opportunity to join the EGS-ABG family! I am grateful for the time you spent with me during the interview which was the only gateway to this programme.

Susanne Eriksson, thank you a lot for all the support! As a postgraduate studies coordinator your guidance and help especially with regard to my study plan and support letters was exceptionally helpful and effective.

Elise Norberg, thank you very much for supporting and enabling me to complete the EGS-ABG joint PhD programme. I am so honoured to know you!

Per Madsen, I'm sincerely grateful for all the help in modelling and DMU handling. The time you took in explaining and replying my questions through email have contributed a lot to complete this thesis. Thank you very much!

I would like to thank all the staff members in both of my host institutions. Specifically, thank you Helena Pettersson, Monica Jansson and Cano Merkan at SLU in Uppsala and Louise Pedersen and the late Karin Smedegaard in Foulum for facilitating and helping me with regard to working environments.

Thank you all my amazing fellow EGS-ABG students: Ahmed Ismael, Amabel Tenghe, André Hidalgo, Bingjie Li, Chrisy Rochus, Doreen Schwochow, Edin Hamzic, Gabriel Rovere, Gareth Difford, Grum Gebreyesus, Hadeer Moursy,

Hadi Esfandyari, Irene van den Berg, Juan Cordero, Jovana Marjanovic, Merina Shrestha, Nancy Amuzu, Sandrine Duchemin, Shizi Wang, Tao Xiang, Thu Hong Le and Xiaowei Mao for all the shared time in courses seminars, conferences and other meetings. You all but specifically Chrisy, Grum, and Gabriel, have positively contributed to the completion of this thesis.

Thanks also goes to the PhD students at the department: Agnese Viluma, Anne Loberg. Emelie Z. König, Gabriela Bottani, Hadrien Gourlé, Helena Eken Asp, Kim J. Fegraeus, Maximillian Manzi, Per Arvelius, Sofia Nyman, Suvi Mäkeläinen, and Tomas Klingström for all the shared time during fikas, lunch breaks, seminars and other meetings.

Towards the end of my doctoral studies there were times of difficulties and I felt like I had nowhere to turn to. But I found that our faculties and departments at both universities are full of considerate people with open doors to turn to. I don't want to mention the names of these people who helped me during those challenging times. **But I cannot thank them enough!!**

I had a horrible accident in a bad day but in a good country, Sweden! Thanks so much for the incredible medical team especially surgeons: Mattias Wahlborg, Maria Mannberg, Daniel Hunde and the nurses at the Orthopedic unit of the Akademiska Sjukhuset and the nurses at Gottsunda and Kungsgärdets vårdcentral in Uppsala. Definitely I am very grateful to the Health Care System in Sweden!

I cannot end the thank-you note without acknowledging the generous financial support from the Erasmus-Mundus fellowship programme, twice☺. It is a great honour to be an alumni of Erasmus Mundus both at Doctoral (**EGS-ABG**) and Master (**EMABG**) programmes!

Berihu Gebremedhin Welderufael
September 29, 2017
Uppsala.