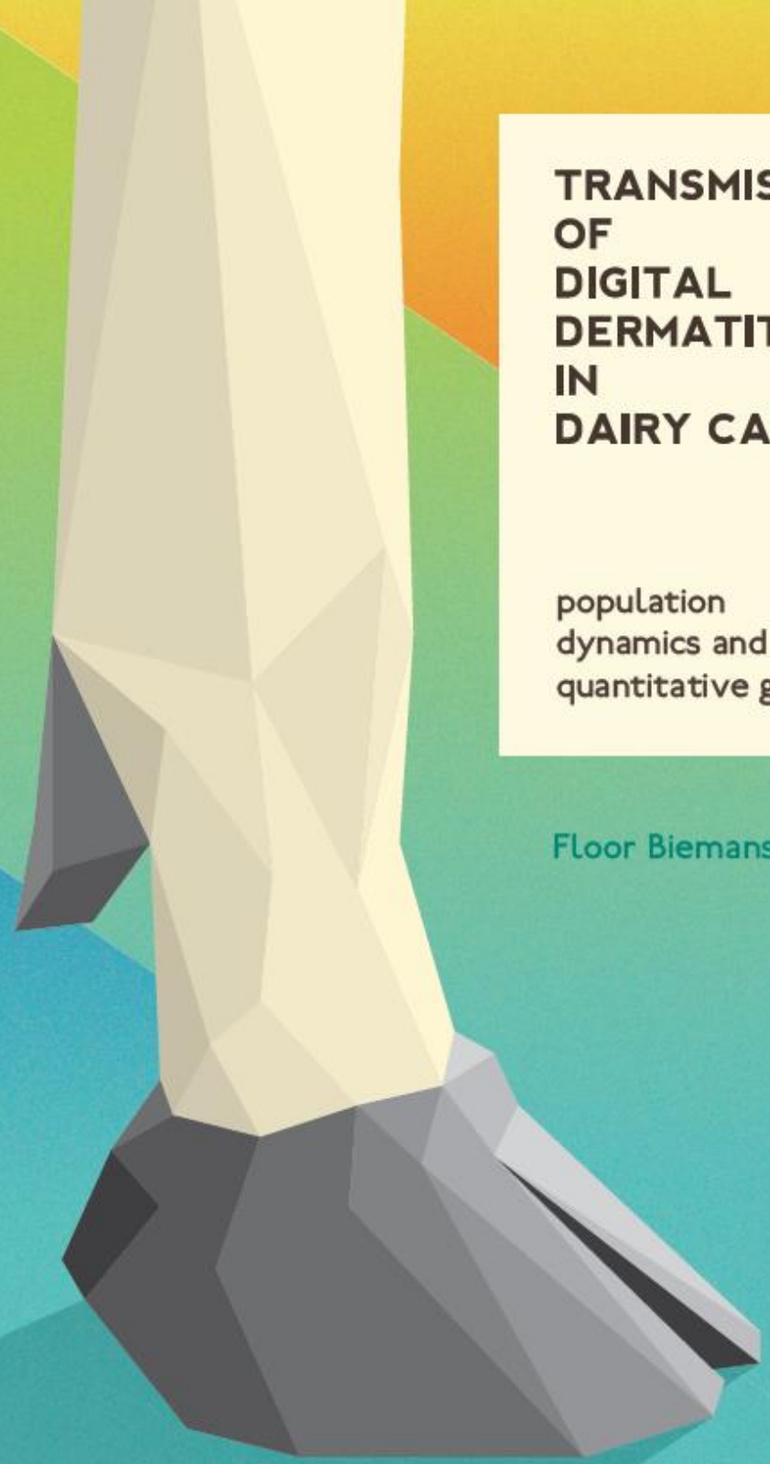
TRANSMISSION OF DIGITAL DERMATITIS IN DAIRY CATTLE

population dynamics and host quantitative genetics

Floor Biemans



TRANSMISSION OF DIGITAL DERMATITIS IN DAIRY CATTLE population dynamics and host quantative genetics Floor Biemans

Propositions

1. Eradication of digital dermatitis is possible with selective breeding in Holstein Friesian dairy cattle.
   (this thesis)

2. To lower the R_0 of digital dermatitis to a value below one, management interventions should focus on prevention and treatment of class M4 lesions.
   (this thesis)

3. Only when we are able to brew beer in space, we are ready to live on planets other than earth.

4. To protect humankind from the detrimental consequences of constant economic growth, governance must be organized on a global scale.

5. To reduce environmental impact, scientist should set an example by having conferences in virtual reality rather than traveling the world to meet in real life.

6. In a civilised country, the entire parliament should vote in favour of gay marriage.

Propositions belonging to the thesis entitled:
"Transmission of digital dermatitis in dairy cattle: Population dynamics and host quantitative genetics"

Floor Biemans
Wageningen, 9 May 2018

Transmission of digital dermatitis in dairy cattle:
Population dynamics and host quantitative genetics

Floor Biemans

Thesis committee

Promotor
Prof. Dr M.C.M. de Jong
Professor of Quantitative Veterinary Epidemiology
Wageningen University & Research

Co-promotor
Dr P. Bijma
Assistant professor Animal Breeding and Genomics
Wageningen University & Research

Other members
Prof. Dr F.A. van Eeuwijk, Wageningen University & Research
Prof. Dr T. Meuwissen, Norwegian University of Life Sciences, Ås, Norway
Dr A. Wilson, University of Edinburgh, Edinburgh, Scotland
Prof. Dr G. van Schaik, Utrecht University, The Netherlands

This research was conducted under the auspices of the Graduate School Wageningen Institute of Animal Sciences (WIAS).

Transmission of digital dermatitis in dairy cattle: Population dynamics and host quantitative genetics

Floor Biemans

Thesis

Submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 9 May 2018
at 1.30 p.m. in the Aula.

Floor Biemans
Transmission of digital dermatitis in dairy cattle: Population dynamics and host quantitative genetics,
204 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summaries in English and Dutch

ISBN 978-94-6343-763-9
DOI: https://doi.org/10.18174/443357

Abstract

Biemans, F. (2018). Transmission of digital dermatitis in dairy cattle: Population dynamics and host quantitative genetics. PhD thesis, Wageningen University and Research, Wageningen, The Netherlands

Susceptibility, infectivity, the contact rate, and the duration of the infectious period together determine the basic reproduction ratio (R_0). The R_0 is the average number of secondary cases caused by a typical infectious individual in a fully susceptible population. It determines the ability of an infection to establish itself in a population. The threshold value is one; if $R_0 < 1$ a typical infectious individual will infect on average less than one susceptible individual and the disease will die out with certainty. If $R_0 > 1$ a major outbreak is possible, and sometimes such a disease may persist in a population. For endemic diseases in homogeneous populations, the prevalence in the equilibrium follows from R_0 as $1 - \frac{1}{R_0}$. Breeding strategies that aim to reduce the prevalence of endemic diseases should thus aim to reduce R_0. Because R_0 depends on both susceptibility and infectivity of the host population, genetic variation in both those traits should be taken into account. This thesis focusses on Digital Dermatitis (DD) in dairy cattle. DD is an endemic infectious claw disease associated with lameness. We collected time-series data on individual disease that might facilitate genetic selection against DD. In this thesis, we investigated transmission dynamics for DD and estimated genetic effects for both host susceptibility and host infectivity. We proposed a generalized linear mixed model to estimate SNP effects on both host susceptibility and host infectivity from time-series data on individual disease status. The model accounted for variation in exposure of susceptible individuals to infectious group mates, and for the infectivity genotypes of those group mates. The power to detect SNP effects was high for susceptibility but lower for infectivity. We applied the model to field data on DD to investigate the contribution of different disease classes to R_0. The estimated R_0 was 2.36, to which the class with irregular skin contributed 88.5%. Genomic estimated breeding values for R_0 ranged from 0.62 to 6.68 with an accuracy of ~0.6. There were 135 SNPs with a suggestive association for several host susceptibility traits, and heritability estimates for these traits ranged from 0.09 to 0.37. These results show that genetic selection against DD is very promising; there is substantial heritable variation and a meaningful accuracy can be obtained from a limited amount of data.

Voor mijn ouders en bro

Contents

1

General introduction

1.1 Introduction

Infectious diseases in livestock are a worldwide concern. First, they pose a threat to human health when they are zoonotic. Second, excessive or incorrect use of vaccines may lead to vaccine escape strains in viruses (Wilson et al., 1998; Brueggemann et al., 2007), and incorrect use of antibiotics can lead to resistant bacteria (Neu, 1992). Third, infectious diseases affect animal health and welfare (Broom, 2006), and last, they cause economic losses due to a decrease in production, costs of treatment and prevention measures, time spent treating the disease, and in some cases culling of the animal (Read and Walker, 1998; Bennett and IJpelaar, 2005; van der Linde et al., 2010). So, it is essential to keep searching for (additional) methods to fight infectious diseases. Selection and breeding for host individuals (*i.e.* livestock) with desirable traits with respect to infectious diseases can be such an additional method (Binder and Levitt, 1998; Stear et al., 2001).

1.2 Selection and breeding

Host susceptibility and host infectivity are two (sets of) host traits that affect disease transmission. Susceptibility is the relative risk to get infected when exposed to a typical infectious individual or to the infectious material of this typical infected individual in the environment. Susceptibility is a characteristic of the focal individual, and has a direct genetic effect (DGE). A DGE is a heritable effect of an individual that affects the phenotype (disease status) of the individual itself. Infectivity, on the other hand, is the relative propensity of an individual to infect a typical susceptible individual. Infectivity affects the disease status of other individuals rather than the disease status (fitness) of the focal individual itself, and has therefore an indirect genetic effect (IGE). An IGE is a heritable effect of an individual on the phenotype of another individual (Wolf et al., 1998; Bijma and Wade, 2008). For infectious diseases it is noteworthy that susceptibility has an indirect genetic effect, individuals with a low susceptibility have a lower chance of being infected and, therefore, have a reduced chance of infecting others (Anche et al., 2014).

Indirect genetic effects can have a considerable and sometimes unexpected effect on the rate and direction of evolution by natural selection, and on response to selective breeding (Griffing, 1967; Moore et al., 1997; Bijma and Wade, 2008). This means that IGEs can, in principle, be used for genetic improvement of populations, and they need to be studied to understand the direction of selection response. However, studies on infectious diseases tend to focus on individual differences in susceptibility (sometimes measured as resistance) only (Woolhouse et al., 1998; Springbett et al., 2003). To make optimal use of all heritable variation that exist with

respect to transmission of infectious diseases, both host susceptibility and host infectivity should be taken into account. A key question is, however, how much genetic variation there is in (especially) infectivity.

There is evidence of phenotypic variation in infectivity, *e.g.*, super-shedders that shed many more infectious units compared to other individuals of the same species (Chase-Topping et al., 2008). Super-shedders could well be more infectious, *i.e.*, be super-spreaders. Super-spreaders are assumed to infect disproportionally more host individuals compared to other individuals of the same species, not only because they shed more infectious units, but also for example because they make more infective contacts (Stein, 2011). If this observed variation in infectivity is (partly) genetically determined, infectivity can be selected against in order to reduce the transmission of infectious diseases.

Knowledge of the amount of genetic variation in host infectivity is limited. It is expected that genetic variation in infectivity is not exhausted by natural selection, because infectivity affects the disease status of other individuals instead of the disease status (fitness) of the focal individual. Infectivity will only be affected by natural selection when feedback mechanisms like kin and group selection are present (Bijma and Wade, 2008). Even when such feedback mechanisms are present, selection on IGEs is generally weaker than on DGEs (Bijma, 2010). Therefore, there might be more genetic variation in infectivity compared to susceptibility.

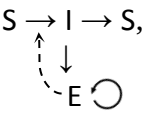
To fully understand and quantify the impact of genetic variation in host susceptibility and host infectivity on disease transmission, we should combine knowledge from the field of quantitative genetics with knowledge from the field of quantitative epidemiology.

1.3 Epidemiology

The field of quantitative epidemiology uses mathematical models and statistical methods to study disease spread through a population. In the often used compartmental model, individuals are assigned to a compartment that represents a specific disease status. Transitions between these stages can be modelled deterministically or stochastically. The rate at which transitions occur is determined by the model parameters. Both epidemics and endemics can be modelled with compartmental models. In this thesis, the focus is on transmission of endemic infectious diseases only.

An endemic infectious disease can be modelled, for example, with a Susceptible-Infectious-Susceptible-model (SIS-model). In this model, the total population (with size N) is divided into a susceptible compartment (S) and an infectious compartment

(I), with S + I = N. The symbols S and I denote the disease status of the individuals in the compartment as well as the number of individuals with that disease status. Since an infection is often transmitted via the environment (with the exception of *e.g.* sexual transmitted diseases), an extra environmental route (E) should be added to the model (see for example de Rueda et al., 2015):



here, E denotes the amount of infection coming from the environment.

In stochastic SIS-models, each individual moves between the two compartments with certain rate (probability per unit of time) determined by the model parameters. For S → I, the total transmission rate is $\beta S \frac{I}{N}$, where β is the transmission rate parameter which contains information on the contact rate and transmission probability given contact (Kermack and McKendrick, 1927; Roberts and Heesterbeek, 1993). With genetic heterogeneity among host individuals, the β depends on the susceptibility genotype of the susceptible individuals and the infectivity genotype of the infectious individuals (Bishop and MacKenzie, 2003; Nath et al., 2004). For I → S the recovery rate is αI, where α is the recovery rate parameter that gives information on the average duration of the infectious period ($1/\alpha$) and part of the infectiousness of the infected individuals.

Together, the transmission rate parameter and the duration of the infectious period (the recovery rate parameter) determine the basic reproduction ratio R_0. The R_0 is the average number of secondary cases caused by a typical infectious individual in a fully susceptible population (May and Anderson, 1987). It is a population parameter, rather than a parameter of a single individual. The R_0 contains information on the ability of an infection to establish itself in the population (May and Anderson, 1987). The threshold value is one, if $R_0 < 1$ an infectious individual will infect on average less than one susceptible individual and the disease will die out. If $R_0 > 1$ a disease can persist at an endemic level in a population, but in the stochastic model is may die out by chance.

An endemic equilibrium exist when the transmission rate is equal to the recovery rate. At an endemic equilibrium level, S and I are constant, so each infectious individual infects on average exactly one susceptible individual. Thus, the effective reproduction ratio (R_E), the average number of secondary cases caused by a typical infectious individual, is one. The relation between the susceptible fraction and the

basic reproduction ratio is: $R_E = \frac{S}{N} R_0 = 1$. The endemic prevalence level follows from this as, $\frac{I}{N} = 1 - \frac{S}{N} = 1 - \frac{1}{R_0}$. Figure 1.1 shows the relation between R_0 and the infected fraction of the population in the endemic equilibrium.

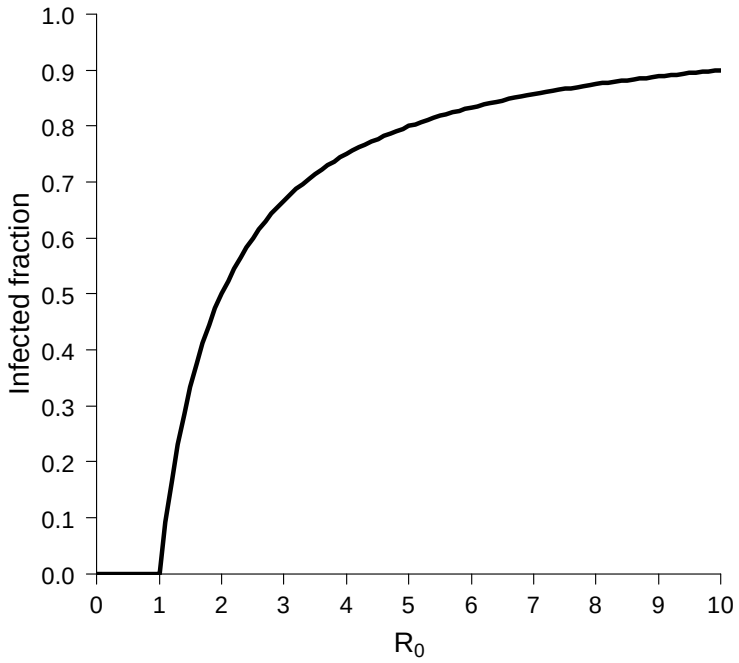


Figure 1.1. Relation between the basic reproduction ratio R_0 and the infected fraction of the population (I/N, prevalence) in the endemic equilibrium, $\frac{I}{N} = 1 - \frac{1}{R_0}$.

1.4 Selection and breeding for reduced disease transmission

Classic quantitative genetic approaches connect the disease of an individual to their own breeding value. Thereby they capture variation in susceptibility only, variation in exposure of susceptible individuals to infectious herd mates and variation in infectivity of the herd mates are ignored. However, the endemic prevalence level of a disease is determined by R_0. Breeding strategies that aim to reduce the prevalence level should thus focus on reducing R_0, preferably to a value below one. Because R_0 depends on both susceptibility and infectivity, the genetic variation in both of these traits should be taken into account. Estimating genetic variation in infectivity is difficult because infectivity is expressed by infected individuals only, and because genetic variation must be estimated indirectly from the number of susceptible group

mates that become infected. However, it is likely that the quality of the estimates can be improved when data on disease status are recorded multiple times (Pooley et al., 2014; Anacleto et al., 2015). So, breeding strategies should take variation in susceptibility, infectivity, and variation in exposure into account.

1.5 Digital Dermatitis

This thesis focusses on the endemic disease Digital Dermatitis (DD). DD is an infectious claw disease that mainly affects dairy and beef cattle, but is also observed in sheep (Sullivan et al., 2014), goats (Sullivan et al., 2015) and North American elk (Clegg et al., 2015). In dairy cattle, lesions occur usually on the hind feet above the interdigital space next to the heel bulbs (Figure 1.2) (Walker et al., 1995). The lesions can develop filiform papillae and be surrounded by hyperkeratotic skin with hairs 2-3 times longer than normal (Read and Walker, 1998). DD is associated with lameness; a severely affected cow bears its weight on the toe of the affected foot, shakes the foot as if in pain, and shows reluctance to move (Bassett et al., 1990; Collighan and Woodward, 1997; Read and Walker, 1998).

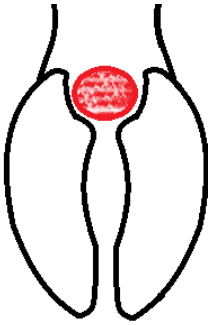


Figure 1.2. Digital Dermatitis lesions (red circle) usually occur above the interdigital space next to the heel bulbs of the hind feet of dairy cattle.

Digital Dermatitis is transmitted via the environment, where the infectious "agent" is an amalgamation of microbes with a predominant bacterial component (Demirkan et al., 1999; Read and Walker, 1998; Rodríguez-Lainz et al., 1996; Sogstad et al., 2005; Vink et al., 2009). Spirochetes of the genus *Treponema spp.* are the most common bacteria in DD lesions (Clegg et al., 2015). *Treponema spp.* with a preference for keratinised cells probably penetrate the epidermal layers, while producing a keratolytic toxin. This degeneration of keratin layers stimulates epidermal proliferation and hyperplasia (Blowey et al., 1994).

Digital Dermatitis is prevalent in many countries, like Mexico (Argáez-Rodríguez et al., 1997), the Republic of Chile (Rodriguez-Lainz et al., 1998), the United Kingdom (Laven and Logue, 2006), the Netherlands (Holzhauer, 2006), the United States (Wells et al., 1999), Sweden (Manske et al., 2002), and Denmark (Nielsen et al., 2009). In the Netherlands, about 90% of the herds was affected by DD in 2003 (Holzhauer et al., 2006). In a cross-sectional study, 21.2% of the cows were affected, with within herd prevalences ranging from 0% to 83.0% (Holzhauer et al., 2006). Factors such as breed, herd size, lactation stadium, flooring system, and climate affect the prevalence in a herd (Holzhauer et al., 2006).

Lesions can be classified with the standardized system that was developed by Döpfer *et al.* (1997) and extended by Berry *et al.* (2012). This system has six distinct classes (M0, M1, M2, M3, M4, and M4.1). Class M0 is skin without a macroscopically visible lesion. Class M1 is a small lesion of 0-2 cm, class M2 is a lesion of >2 cm, class M3 is a lesion covered by a scab, class M4 is irregular skin with dyskeratotis or surface proliferation, and class M4.1 is a small lesion (M1) in addition to irregular skin (M4) (Döpfer et al., 1997; Döpfer, 2009; Berry et al., 2012). Figure 1.3 to 1.8 show the distinct DD classes. Cows scored as M0 are considered susceptible to DD, while cows scored M1, M2, M3, M4, or M4.1 are infected.

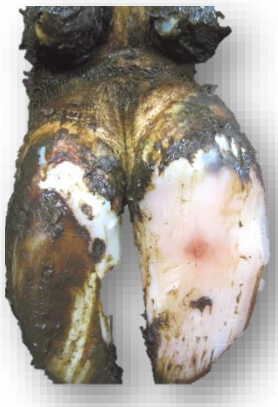


Figure 1.3. Claw without Digital Dermatitis; class M0, skin without a macroscopically visible lesion.

Figure 1.4. Claw with Digital Dermatitis class M1, a small lesion of 0-2 cm.

Figure 1.5. Claw with Digital Dermatitis class M2, a lesion of >2 cm.

Figure 1.6. Claw with Digital Dermatitis class M3, a lesion covered by a scab.

Figure 1.7. Claws with Digital Dermatitis class M4, irregular skin with dyskeratotis (a) and irregular skin with surface proliferation (b).

Figure 1.8. Claw with Digital Dermatitis class M4.1, a small lesion (M1) in addition to irregular skin (M4).

Once infected, a cow can present several classes before it recovers (Döpfer et al., 2012). A cow might not present all the classes, and the order it presents the classes when infected is not necessarily from M1 to M4.1. Based on the classes an infected cow presents, the cow can be categorized as one of three types: a type I cow never show class M2 but might have class M1 or M4; a type II cow show class M2 once, thereafter this class is absent for a long period; a type III cow show class M2 lesions repeatedly over a short period of time (Döpfer, 2009). This indicates that there is variation in the classes an animal presents, and in the duration of these classes. Furthermore, it is possible that the infectivity of the lesions differs between classes and animals. Some animals might therefore be more infectious than others are, and these differences in infectivity might be genetic.

1.6 The gap

Genetic variation in host susceptibility and host infectivity provides an opportunity to reduce disease transmission with selection and breeding, and may also help to unravel the mechanisms underlying disease transmission. However, there is a gap between animal breeding and quantitative epidemiology that must be bridged. Classic quantitative genetic approaches only capture variation in susceptibility, and fail to take into account variation in infectivity and exposure.

1.7 Aim

In this thesis, we aim to estimate genetic effects on both host susceptibility and host infectivity for an endemic disease with a generalized linear model. The model is applied to data on DD. We extend the generalized linear model developed by Anche et al. (2015). They proposed this model to estimate relative effects of genes on susceptibility and infectivity from binary data on disease status collected at the end of an epidemic. Anacleto et al. (2015) showed that the quality of the infectivity estimates can be improved when data are recorded multiple times during the infection chain. Because, when the number (and possible phenotypes and genotypes) of the infectious individuals to which a susceptible individual was exposed are known when it did or did not become infected, this provides information on who infected whom (Pooley et al., 2014). An endemic is more suitable for multiple recordings compared to an epidemic as the occurrence of new cases is more predictable. In this study we, therefore, focus on the endemic disease DD.

In chapter 2, we test the extended model in a simulation study. We simulate an endemic disease and observe the disease status of individuals repeatedly. We estimate genetic variation in susceptibility, and use the variation in the exposure of susceptible individuals to infectious herd mates to estimate genetic variation in infectivity. Bias and precision of the estimates are quantified for different effect sizes and the optimal recording interval is identified

In chapter 3, we use phenotype data on DD to determine how the R_0 for DD is composed. We investigate the distribution of the classes that are first observed after infection, the average duration of each class, and the infectivity of each class. With this information we determine the contribution of each class to R_0.

In chapter 4, the phenotype data on DD are combined with genotype data of the same animals. Here we estimate genetic variance components for host susceptibility, infectivity, and R_0 for DD. Furthermore, we investigate the effect of including both susceptibility and infectivity in the model. Finally, different models are compared for the ability to predict whether or not a susceptible animal gets infected.

In chapter 5, we perform several genome-wide association studies (GWAS) to detect single nucleotide polymorphisms (SNPs) associated with DD. We perform the GWAS with two different models. First a linear model is used to detect SNPs that are associated with host susceptibility to the different M-classes and the presence of active lesions. Next a generalized linear model is used to detect SNPs that are associated with host susceptibility and host infectivity to DD.

Finally, I discuss the broader perspective of the study in the general discussion (chapter 6). I will focus on disease traits that were not considered in this thesis. Some

affect the basic reproduction ratio, like the duration of the infectious period and the indirect infectivity of a cow via the environment. Next, I will address breeding against infectious diseases in practice, *i.e.*, the correlation between the estimated breeding value for R_0 and the estimated breeding values for milk production and DD that are currently used. Furthermore, I will address the use of sensor systems for phenotype collection. Finally, I will propose an additional explanation for the lack of power to estimate differences in infectivity.

1.8 References

Anacleto, O., Garcia-Cortés, L.A., Lipschutz-Powell, D., Woolliams, J.A., Doeschl-Wilson, A.B., 2015. A novel statistical model to estimate host genetic effects affecting disease transmission. Genetics 201, 871-884.

Anche, M., de Jong, M., Bijma, P., 2014. On the definition and utilization of heritable variation among hosts in reproduction ratio R0 for infectious diseases. Heredity 113, 364-374.

Anche, M.T., Bijma, P., De Jong, M.C., 2015. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47, 1-15.

Argáez-Rodríguez, F.d.J., Hird, D.W., de Anda, J.H., Read, D.H., Rodríguez-Lainz, A., 1997. Papillomatous digital dermatitis on a commercial dairy farm in Mexicali, Mexico: Incidence and effect on reproduction and milk production. Preventive Veterinary Medicine 32, 275-286.

Bassett, H., Monaghan, M., Lenhan, P., Doherty, M., Carter, M., 1990. Bovine digital dermatitis. Veterinary Record 126, 164-165.

Bennett, R., IJpelaar, J., 2005. Updated estimates of the costs associated with thirty four endemic livestock diseases in Great Britain: a note. Journal of Agricultural Economics 56, 135-144.

Berry, S.L., Read, D.H., Famula, T.R., Mongini, A., Döpfer, D., 2012. Long-term observations on the dynamics of bovine digital dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. The Veterinary Journal 193, 654-658.

Bijma, P., 2010. Fisher's fundamental theorem of inclusive fitness and the change in fitness due to natural selection when conspecifics interact. Journal of evolutionary biology 23, 194-206.

Bijma, P., Wade, M., 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. Journal of evolutionary biology 21, 1175-1188.

Binder, S., Levitt, A.M., 1998. Preventing emerging infectious diseases: a strategy for the 21st century. US Department of Health and Human Services, Centers for Disease Control and Prevention.

Bishop, S.C., MacKenzie, K.M., 2003. Genetic management strategies for controlling infectious diseases in livestock populations. Genetics Selection Evolution 35, S3.

Blowey, R., Done, S., Cooley, W., 1994. Observations on the pathogenesis of digital dermatitis in cattle. Veterinary Record 135, 117-115.

Broom, D.M., 2006. Behaviour and welfare in relation to pathology. Applied Animal Behaviour Science 97, 73-83.

Brueggemann, A.B., Pai, R., Crook, D.W., Beall, B., 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States.

Chase-Topping, M., Gally, D., Low, C., Matthews, L., Woolhouse, M., 2008. Super-shedding and the link between human infection and livestock carriage of Escherichia coli O157. Nature Reviews Microbiology 6, 904-912.

Clegg, S., Mansfield, K., Newbrook, K., Sullivan, L., Blowey, R., Carter, S., Evans, N., 2015. Isolation of digital dermatitis treponemes from hoof lesions in wild North American elk (Cervus elaphus) in Washington State, USA. Journal of clinical microbiology 53, 88-94.

Collighan, R.J., Woodward, M.J., 1997. Spirochaetes and other bacterial species associated with bovine digital dermatitis. FEMS Microbiology Letters 156, 37-41.

de Rueda, C.B., de Jong, M.C., Eblé, P.L., Dekker, A., 2015. Quantification of transmission of foot-and-mouth disease virus caused by an environment contaminated with secretions and excretions from infected calves. Veterinary research 46, 1.

Döpfer, D., 2009. Digital Dermatitis: The dynamics of digital dermatitis in dairy cattle and the manageable state of disease, CanWest Veterinary Conference, Banff AB, Canada.

Döpfer, D., Holzhauer, M., Boven, M.v., 2012. The dynamics of digital dermatitis in populations of dairy cattle: Model-based estimates of transition rates and implications for control. The Veterinary Journal 193, 648-653.

Döpfer, D., ter Huurne, A., Cornelisse, J., Van Asten, A., Koopmans, A., Meijer, F., Schukken, Y., Szakall, I., Klee, W., Bosma, R., 1997. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and Campylobacter faecalis. Veterinary Record 140, 620-623.

Griffing, B., 1967. Selection in reference to biological groups I. Individual and group selection applied to populations of unordered groups. Australian Journal of Biological Sciences 20, 127-140.

Holzhauer, M., 2006. Claw Health in Dairy Cows in The Netherlands: Epidemiological aspects of different claw disorders in dairy cattle in The Netherlands. Utrecht University, Utrecht.

Holzhauer, M., Hardenberg, C., Bartels, C., Frankena, K., 2006. Herd-and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. Journal of dairy science 89, 580-588.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. The Royal Society, pp. 700-721.

Laven, R.A., Logue, D.N., 2006. Treatment strategies for digital dermatitis for the UK. The Veterinary Journal 171, 79-88.

Manske, T., Hultgren, J., Bergsten, C., 2002. Topical treatment of digital dermatitis associated with severe heel-horn erosion in a Swedish dairy herd. Preventive Veterinary Medicine 53, 215-231.

May, R.M., Anderson, R.M., 1987. Transmission dynamics of HIV infection. Nature 326, 137.

Moore, A.J., Brodie III, E.D., Wolf, J.B., 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution, 1352-1362.

Nath, M., Woolliams, J., Bishop, S., 2004. Identifying critical parameters in the dynamics and control of microparasite infection using a stochastic epidemiological model. Journal of animal science 82, 384-396.

Neu, H.C., 1992. The crisis in antibiotic resistance. Science 257, 1064-1073.

Nielsen, B.H., Thomsen, P.T., Sørensen, J.T., 2009. A study of duration of digital dermatitis lesions after treatment in a Danish dairy herd. Acta Veterinaria Scandinavica 51, 27.

Pooley, C., Bishop, S., Marion, G., 2014. Estimation of single locus effects on susceptibility, infectivity and recovery rates in an epidemic using temporal data, Proceedings of the 10th world congress of genetics applied to livestock production, pp. 17-22.

Read, D.H., Walker, R.L., 1998. Papillomatous digital dermatitis (footwarts) in California dairy cattle: clinical and gross pathologic findings. Journal of Veterinary Diagnostic Investigation 10, 67-76.

Roberts, M., Heesterbeek, H., 1993. Bluff your way in epidemic models. Trends in microbiology 1, 343-348.

Rodriguez-Lainz, A., Melendez-Retamal, P., Hird, D.W., Read, D.H., 1998. Papillomatous digital dermatitis in Chilean dairies and evaluation of a screening method. Preventive Veterinary Medicine 37, 197-207.

Springbett, A., MacKenzie, K., Woolliams, J., Bishop, S., 2003. The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165, 1465-1474.

Stear, M., Bishop, S., Mallard, B., Raadsma, H., 2001. The sustainability, feasibility and desirability of breeding livestock for disease resistance. Research in veterinary science 71, 1-7.

Stein, R.A., 2011. Super-spreaders in infectious diseases. International Journal of Infectious Diseases 15, e510-e513.

Sullivan, L., Blowey, R., Carter, S., Duncan, J., Grove-White, D., Page, P., Iveson, T., Angell, J., Evans, N., 2014. Presence of digital dermatitis treponemes on cattle and sheep hoof trimming equipment. Veterinary Record 175, 201.

Sullivan, L., Evans, N., Clegg, S., Carter, S., Horsfield, J., Grove-White, D., Duncan, J., 2015. Digital dermatitis treponemes associated with a severe foot disease in dairy goats. The Veterinary record 176, 283-283.

van der Linde, C., de Jong, G., Koenen, E.P.C., Eding, H., 2010. Claw health index for Dutch dairy cattle based on claw trimming and conformation data. Journal of Dairy Science 93, 4883-4891.

Walker, R.L., Read, D.H., Loretz, K.J., Nordhausen, R.W., 1995. Spirochetes isolated from dairy cattle with papillomatous digital dermatitis and interdigital dermatitis. Veterinary Microbiology 47, 343-355.

Wells, S., Garber, L., Wagner, B., 1999. Papillomatous digital dermatitis and associated risk factors in US dairy herds. Preventive Veterinary Medicine 38, 11-24.

Wilson, J., Nokes, D., Carman, W., 1998. Current status of HBV vaccine escape variants–a mathematical model of their epidemiology. Journal of viral hepatitis 5, 25-30.

Wolf, J.B., Brodie III, E.D., Cheverud, J.M., Moore, A.J., Wade, M.J., 1998. Evolutionary consequences of indirect genetic effects. Trends in Ecology & Evolution 13, 64-69.

Woolhouse, M., Stringer, S., Matthews, L., Hunter, N., Anderson, R., 1998. Epidemiology and control of scrapie within a sheep flock. Proceedings of the Royal Society of London. Series B: Biological Sciences 265, 1205-1210.

2

A model to estimate effects of SNPs on host susceptibility and infectivity for an endemic infectious disease

Floor Biemans[1,2], Mart C.M. de Jong[1], Piter Bijma[2]

[1] Quantitative Veterinary Epidemiology Group, Wageningen University & Research, 6700 AH Wageningen, the Netherlands

[2] Animal Breeding and Genomics, Wageningen University & Research, 6700 AH Wageningen, the Netherlands

Genetics Selection Evolution (2017) 49:53

Abstract

Infectious diseases in farm animals affect animal health, decrease animal welfare and can affect human health. Selection and breeding of host individuals with desirable traits regarding infectious diseases can help to fight disease transmission, which is affected by two types of (genetic) traits: host susceptibility and host infectivity. Quantitative genetic studies on infectious diseases generally connect an individual's disease status to its own genotype, and therefore capture genetic effects on susceptibility only. However, they usually ignore variation in exposure to infectious herd mates, which may limit the accuracy of estimates of genetic effects on susceptibility. Moreover, genetic effects on infectivity will exist as well. Thus, to design optimal breeding strategies, it is essential that genetic effects on infectivity are quantified. Given the potential importance of genetic effects on infectivity, we set out to develop a model to estimate the effect of single nucleotide polymorphisms (SNPs) on both host susceptibility and host infectivity. To evaluate the quality of the resulting SNP effect estimates, we simulated an endemic disease in 10 groups of 100 individuals, and recorded time-series data on individual disease status. We quantified bias and precision of the estimates for different sizes of SNP effects, and identified the optimum recording interval when the number of records is limited.

We present a generalized linear mixed model to estimate the effect of SNPs on both host susceptibility and host infectivity. SNP effects were on average slightly underestimated, *i.e.* estimates were conservative. Estimates were less precise for infectivity than for susceptibility. Given our sample size, the power to estimate SNP effects for susceptibility was 100% for differences between genotypes of a factor 1.56 or more, and was higher than 60% for infectivity for differences between genotypes of a factor 4 or more. When disease status was recorded 11 times on each animal, the optimal recording interval was 25 to 50% of the average infectious period.

Our model was able to estimate genetic effects on susceptibility and infectivity. In future genome-wide association studies, it may serve as a starting point to identify genes that affect disease transmission and disease prevalence.

2.1 Introduction

Infectious diseases in farm animals affect animal health, decrease animal welfare and can affect human health (Broom, 2006). Infectious diseases also cause economic losses due to disease-related costs, treatment costs, costs for prevention measures, and reduced production (Bennett and IJpelaar, 2005). Bacterial infections are often treated with antibiotics, which can lead to antibiotic-resistant bacteria (Neu, 1992). Viral infections can be prevented with vaccination, which can lead to vaccine escape strains (Wilson et al., 1998; Brueggemann et al., 2007). Thus, it is highly desirable to search for additional ways to fight transmission of infectious diseases. One such approach consists of selecting and breeding host populations for desirable traits regarding infectious diseases (Binder and Levitt, 1998).

Two main sets of host traits affect transmission of infectious diseases: host susceptibility and host infectivity. Susceptibility is the relative risk of an individual to become infected when exposed to a typical (average) infectious individual or (for infectious diseases transmitted via the environment) the infectious material excreted by a typical infectious individual. Infectivity is the relative propensity of an infected individual to infect a typical (average) susceptible individual.

Studies that investigate host genetic effects related to infectious diseases generally focus on host disease status, and link this to the genotype of the host (Woolhouse et al., 1998; Springbett et al., 2003). By linking own disease status to own genotype, only genetic effects on susceptibility are captured and variation in exposure of susceptible individuals to infectious herd mates is ignored, which may limit the accuracy of estimates of genetic effects on susceptibility. Moreover, there is evidence that genetic variability in infectivity exists as well. Variability in infectivity is found in, for example, super-shedders, *i.e.*, individuals that shed many more infectious units than the average individual in the population (Chase-Topping et al., 2008). This variability in shedding was found among individuals infected with the "same" pathogen and, thus could be due to host genetic differences.

A host genetic effect on infectivity is an example of an indirect genetic effect (IGE) (Moore et al., 1997; Anche et al., 2014), which is a heritable effect of one individual on the phenotype of another individual (Bijma and Wade, 2008). IGE can have profound effects on the rate and direction of evolution by natural selection and on response to selective breeding (Wolf et al., 1998; Bijma et al., 2007b; Bijma and Wade, 2008; Anche et al., 2014). Thus, genetic effects on infectivity can be used for genetic improvement of populations that suffer from infectious diseases (Lipschutz-Powell et al., 2012b; Anche et al., 2014) but its use requires different breeding strategies (Bijma and Wade, 2008; Lipschutz-Powell et al., 2012a; Anche et al., 2014).

To design optimal breeding strategies, it is essential to first quantify the genetic effects on infectivity.

Genome-wide association studies (GWAS), in which effects of single nucleotide polymorphisms (SNPs) on a phenotype are estimated, are a common way to quantify genetic effects. To estimate effects of SNPs on susceptibility and infectivity, a generalized linear model with a complementary log-log link function can be used. This model has been applied to data on the final disease status of individuals after an epidemic disease (Lipschutz-Powell et al., 2014; Anche et al., 2015), but many diseases are endemic. Furthermore, it is likely that the quality of the estimates improves when data on individual disease status are recorded over multiple (short) time intervals during the infection chain in a population. Each interval can then be seen as an incomplete epidemic, in which only a fraction of the susceptible individuals become infected. For each interval, the infectious individuals to which a susceptible focal individual is exposed are known. Thus, more information on who infected whom and on the rate of infection is available, compared to information on the final disease status only, which is expected to improve the quality of the estimates for host genetic effects on susceptibility and infectivity (Anacleto et al., 2015).

Given the potential importance of genetic effects on infectivity, we set out to develop a model to estimate the effects of SNPs on both host susceptibility and host infectivity for an endemic disease. The model accounts for variation among susceptible individuals in exposure to infectious herd mates and for the genotypes of those herd mates. To evaluate the quality of the SNP effects estimated by the model, we simulated an endemic disease and recorded data on individual disease status multiple times during the endemic. We quantified bias and precision of the estimates for different sizes of SNP effects, and identified the optimal recording interval.

2.2 Methods

2.2.1 Transmission model

Our objective was to develop a model to estimate the effect of single SNPs on disease transmission. Thus, we considered a genetically heterogeneous population of diploid individuals, with one locus for the susceptibility effect γ, and one locus for the infectivity effect φ. The susceptibility locus had two alleles, allele G with value γ_G and allele g with value γ_g. The infectivity locus also had two alleles, allele F with value φ_F and allele f with value φ_f. We assumed additive allele effects on the log-scale, by simulating effects as multiplicative on the original scale such that model

terms could be formulated as allele counts within individuals (Anche et al., 2015). Thus, susceptibility values were $\gamma_{GG} = \gamma_G\gamma_G = \gamma_G^2$ for genotype GG, $\gamma_{Gg} = \gamma_{gG} = \gamma_G\gamma_g$ for genotype Gg/gG, and $\gamma_{gg} = \gamma_g^2$ for genotype gg. Likewise, infectivity values were $\varphi_{FF} = \varphi_F^2$ for genotype FF, $\varphi_{Ff} = \varphi_{fF} = \varphi_F\varphi_f$ for genotype Ff/fF, and $\varphi_{ff} = \varphi_f^2$ for genotype ff. Note that multiplicative allele effects on the original scale introduce dominance on the original scale. Because the value for the heterozygote is lower than the average value of both homozygotes, *i.e.*, $\gamma_{Gg} < 0.5(\gamma_{GG} + \gamma_{gg})$, the dominance is negative (see Discussion).

An endemic disease was modelled with a stochastic compartmental susceptible-infected-susceptible-model (SIS-model). In a SIS-model, two events can occur: infection of a susceptible individual and recovery of an infected individual. Infected individuals were immediately infectious and recovered individuals were immediately susceptible again. Thus, no lasting immunity to disease was assumed. Events (infection and recovery) occurred randomly with a probability per unit of time, depending on model parameters and disease status of individuals in the population.

In a genetically homogeneous population, the expected rate with which susceptible individuals become infected equals $\frac{dS}{dt} = \beta I \frac{S}{N}$, where I is the number of infectious individuals, S the number of susceptible individuals, and $S + I = N$, *i.e.*, the size of the closed population in which the endemic takes place (Kermack and McKendrick, 1927). The transmission rate parameter β is a population specific constant that contains information on the contact rate and transmission probability between hosts (Roberts and Heesterbeek, 1993).

In a genetically heterogeneous population, β varies between pairs of individuals, depending on the susceptibility genotype of the susceptible individual and the infectivity genotype of the infectious individual. We assumed that, between individuals, the susceptibility genotype and the infectivity genotype have independent effects, which is known as separable mixing in epidemiology (Diekmann et al., 1990), *i.e.*, the susceptibility effect of individuals that are susceptible is independent of the infectivity effect of individuals that are infectious. Thus, the transmission rate parameter β_{ij} from an infectious individual with infectivity genotype j ($j = FF$, Ff or ff) to a recipient susceptible individual with susceptibility genotype i ($i = GG$, Gg or gg) was defined as:

$$\beta_{ij} = c\gamma_i\varphi_j,$$

where γ_i is the susceptibility value for genotype i and φ_j the infectivity value for genotype j. Without loss of generality, we chose $\gamma_g = \varphi_f = 1$ as reference allele

values. Therefore, $\gamma_{gg} = \varphi_{ff} = 1$, so that $\beta_{ggff} = c$. Thus, c represents the transmission rate parameter from an infectious individual with infectivity genotype ff to a susceptible individual with susceptibility genotype gg. Since, $\gamma_g = \varphi_f = 1$, γ_G represents the ratio of the value of allele G over the value of allele g, and φ_F represents the ratio of the value of allele F over the value of allele f. For example, $\gamma_{GG}/\gamma_{Gg} = \gamma_G$, and $\gamma_{Gg}/\gamma_{gg} = \gamma_G$.

The *total* infectivity to which susceptible individuals are exposed at time t, depends on the total number of infectious individuals of each genotype at that time $I_j(t)$ and is measured by $\sum_j(\varphi_j\, I_j(t))$. Thus, the infection rate at time t for susceptible individuals with genotype i $(Infection\, rate_i(t))$, depends on the susceptibility of genotype i and on the total infectivity of infectious group mates:

$$Infection\, rate_i(t) = c\, \gamma_i \frac{S_i(t)}{N} \sum_j \big(\varphi_j I_j(t)\big), \qquad \text{(Equation 2.1)}$$

where $S_i(t)$ is the number of susceptible individuals with genotype i at time t.

The probability per unit of time for an individual to recover and become susceptible again was given by the recovery rate parameter α and was assumed to be the same for all genotypes. Note that a single α does not imply the same infectious period for all individuals; because α is a stochastic rate, the length of the infectious period follows an exponential distribution and thus shows random phenotypic, albeit not genetic, variation among individuals.

2.2.2 Generalized linear model

To estimate the effect of single SNPs on both host susceptibility and host infectivity, we developed a generalized linear model (GLM). The GLM was based on the infection rate given by Equation 2.1. We assumed that the recording interval, the disease status of individuals at recording, and the genotypes of individuals were known.
For the sake of readability, the index t is dropped in the following and, hence, S, S_i, I, and I_j refer to the number of individuals at the beginning of the interval. Then, the probability P_i for a single susceptible individual with genotype i to get infected when exposed to all infectious individuals during an interval Δt, follows from assuming a Poisson process within Δt. It is the probability of a non-zero outcome from a Poisson distribution, and follows from Equation 2.1 with $S_i = 1$,

$$P_i = 1 - e^{-c\gamma_i(\sum_j \varphi_j I_j)\Delta t/N}. \qquad \text{(Equation 2.2)}$$

The second term on the right-hand side is the zero-term of the Poisson distribution, which gives the probability of no infection. Thus, the number individuals with genotype i that become infected during Δt, *i.e.*, cases C_i, follows a binomial distribution with binomial total S_i, *i.e.*, depends on the number of susceptible individuals of genotype i at the start of the interval and the probability to become infected given by Equation 2.2 (Velthuis et al., 2003). Equation 2.2 assumes that infections are only caused by individuals that were infectious at the beginning of the interval (I_j). In other words, the effect on the P_i of individuals that became infected or recovered during the interval is ignored in Equation 2.2. This assumption is increasingly violated at longer recording intervals. Thus, we investigated the effect of the recording interval on the quality of the estimates and whether an optimum recording interval exists.

Because the probability to become infected follows from the complement of the zero-term of the Poisson distribution (Equation 2.2), the complementary log-log is the appropriate link function to connect the explanatory variables to the expected value of the observed variable (Velthuis et al., 2003; Anche et al., 2015). Thus, a GLM with a complementary log-log link function was used to estimate effects of SNPs:

$$\mathrm{cloglog}(P_i) = \log(-\log(1 - P_i)) =$$

$$\log(c) + \log(\gamma_i) + \log\left(\sum_j \frac{I_j}{I}\varphi_j\right) + \log\left(\frac{I}{N}\Delta t\right),$$

where I is the total number of infected individuals at the beginning of the interval, such that $\frac{I_j}{I}$ represents the fraction of infectious individuals with infectivity genotype j at the beginning of the interval. As noted by Anche et al. (2015), this model is linear in $\log(\gamma_i)$ but not in $\log(\varphi_j)$. To linearize the model, the arithmetic mean of φ, $\sum_j \frac{I_j}{I}\varphi_j$, was approximated by the corresponding geometric mean, $\prod_j \varphi_j^{\frac{I_j}{I}}$ (Anche et al., 2015), such that $\log\left(\sum_j \frac{I_j}{I}\varphi_j\right) \approx \log\left(\prod_j \varphi_j^{\frac{I_j}{I}}\right) = \sum_j \frac{I_j}{I}\log(\varphi_j)$. Now, the GLM is linear in both $\log(\gamma_i)$ and $\log(\varphi_j)$:

$$\mathrm{cloglog}(P_i) \approx \log(c) + \log(\gamma_i) + \sum_j \frac{I_j}{I}\log(\varphi_j) + \log\left(\frac{I}{N}\Delta t\right).$$

Details on the error caused by this approximation are in the appendix of (Anche et al., 2015), and are less than 5% for infectivity effects up to a factor of 3 (*i.e.*, φ_F between 0.33 and 3.0).

By assuming multiplicative allele effects on the original scale, allele effects were additive on the log-scale. For susceptibility, for example, $\log(\gamma_{gg}) = 0$, $\log(\gamma_{Gg}) = \log(\gamma_G)$, and $\log(\gamma_{GG}) = 2\log(\gamma_G)$. Thus, under this assumption, the model can be expressed in terms of allele counts (Anche et al., 2015). Furthermore, we added a random group effect to account for possible additional (stochastic) differences in transmission between groups. When a random group effect is added to the model, the standard deviations of the estimated parameters are higher than those from a model without group included as a random effect. Although we did not simulate group effects in this study, they must be estimated in real data. Thus, we included a random group effect to better reflect the standard errors on the allele effect estimates that may be found in real data. A generalized linear mixed model (GLMM) allows for the inclusion of random effects resulting in the following final GLMM:

$$\text{cloglog}(P_i) = c_0 + c_1 CountG + c_2 CountF + \log\left(\frac{I}{N}\Delta t\right). \qquad \text{(Equation 2.3)}$$

Where $c_0 = \log(c)$ is the intercept. To achieve that $\gamma_g = \varphi_f = 1$, such that $\log(\gamma_g) = \log(\varphi_f) = 0$, we counted alleles G and F within individuals, rather than alleles g and f, such that the regression coefficients represent the value of a single copy of allele G or F. For example, the ratio of γ_G versus γ_g is $\gamma_G = e^{c_1}$, which is estimated by $\hat{\gamma}_G = e^{\widehat{c_1}}$. Thus, $CountG$ represents the number of G-alleles at the susceptibility locus of the susceptible individual, takes values 0, 1 or 2, and has coefficient $c_1 = \log(\gamma_G)$. $CountF$ represents the average number of F-alleles at the infectivity locus in the infected individuals, takes real values between 0 and 2, and has coefficient $c_2 = \log(\varphi_F)$. $CountF$ is calculated as $\frac{2I_{FF}+I_{Ff}}{I}$, where I_{FF} is the number of infected individuals with genotype FF at the beginning of the interval and I_{Ff} is the corresponding number of infected individuals with genotype Ff. The denominator of $CountF$ is I rather than $2I$ because $CountF$ is the average number of F alleles rather than its proportion. Table 2.1 summarizes the relationship between the regression coefficients of the GLMM and the transmission rate parameters for each genotype. The final model term, $\log\left(\frac{I}{N}\Delta t\right)$, is a known offset, *i.e.*, an "explanatory variable" with coefficient equal to 1. The time period Δt determines the interpretation of the transmission rate parameter. For example, rates are per day when the time period Δt is expressed in days.

Table 2.1. Relationship between the transmission rate parameters and the regression coefficients of the generalized linear mixed model for each genotype.

Transmission rate parameter[1]	Expression in terms of regression coefficients
β_{ggff}	e^{c_0}
β_{Ggff}	$e^{c_0+c_1}$
β_{GGff}	$e^{c_0+2c_1}$
β_{ggFf}	$e^{c_0+c_2}$
β_{GgFf}	$e^{c_0+c_1+c_2}$
β_{GGFf}	$e^{c_0+2c_1+c_2}$
β_{ggFF}	$e^{c_0+2c_2}$
β_{GgFF}	$e^{c_0+c_1+2c_2}$
β_{GGFF}	$e^{c_0+2c_1+2c_2}$

[1] The first two subscripts of β indicate the susceptible genotype of susceptible individuals, the second two subscripts indicate the infectivity genotype of infectious individuals. It follows that $\gamma_G = e^{c_1}$ and $\varphi_F = e^{c_2}$.

2.2.3 Simulations

To evaluate the quality of the estimates from the above model, we simulated an endemic disease and quantified bias and precision of SNP effects estimated based on Model 3. Bias was defined as the difference between the estimated and true effects of each SNP and relative bias was defined as the bias relative to the true size of the effect. Absolute bias and relative bias were calculated on the original scale. Precision was measured by the root mean squared error (RMSE) of the estimated SNP effects on the original scale. Simulations were conducted in R version 3.2.3. and data were analysed with the R-package lme4 (Bates et al., 2014; R Core Team, 2017), using the glmer() function to solve the GLMM with Gauss-Hermite quadrature methods.

A group (defined as closed and random mixing) consisted of 100 individuals, which resembles for example, a dairy herd in the Netherlands. In dairy herds, a limited number of sires is used, so that cows in the same herd are (slightly) more related to each other than to cows in other herds. We simulated such genetic heterogeneity by sampling allele frequencies for susceptibility (p_G and $p_g = 1 - p_G$), and infectivity (p_F and $p_f = 1 - p_F$) for each group from a beta distribution with a mean of 0.5 and standard deviation of 0.05. We chose a beta distribution for p to ensure that allele frequencies are between 0 and 1. For the mean allele frequency, we used 0.5, which is simply the centre of the 0 to 1 interval. We assumed that the susceptibility effect of an individual and that same individual's infectivity effect were

not correlated. Within groups, genotypes were sampled assuming Hardy-Weinberg equilibrium. The loci for susceptibility and infectivity were simulated in linkage equilibrium.

Next, an initial disease status was modelled for each individual. Because interest was in obtaining data from the endemic phase of the disease, the endemic phase was started at the equilibrium in terms of number of susceptible and infectious individuals (details are in the Appendix). The next event, infection or recovery, was sampled using the direct method of the Gillespie's algorithm (Gillespie, 1977), where the probability that a specific event occurred was proportional to the rate with which that event occurred (see (Anche et al., 2015) for an example). Thus, time-period between events was sampled from an exponential distribution with the sum of the rates of infection and recovery as parameter. If the endemic phase died out (no infectious individuals in the population), a random individual was infected immediately. This case was excluded from the analysed data, but included as explanatory variable in the model for subsequent cases.

One replicate consisted of 10 groups of 100 individuals each. In each replicate, individual disease status was recorded 11 times, and individual genotypes were known.

2.2.4 Scenarios

Table 2.2 shows the input values for scenarios 1 and 2.

In scenario 1, we varied γ_G and φ_F simultaneously between 0.3 and 1, while keeping $\gamma_g = \varphi_f = 1$, to investigate statistical power to identify SNP effects on susceptibility and infectivity. A value for $\gamma_g = 0.3$, for example, means that the Gg genotype is $1/0.3 = 3\frac{1}{3}$ times less susceptible than the gg genotype, while the GG genotype is $1/0.3^2 = 11.1$ times less susceptible than the gg genotype.

In scenario 2, we varied the recording interval while keeping the total number of recordings constant, in order to find the optimal recording interval. The recording interval ranged from 4.8 to 133.3% of the average infectious period ($1/\alpha$). For all recording intervals in scenario 2, $\gamma_G = \varphi_F = 0.4$. To check whether the optimal recording interval depends on the effect size, we also investigated a scenario with $\gamma_G = \varphi_F = 0.6$.

Table 2.2. Input values for the simulations.

Variable	Scenario 1 SNP effect	Scenario 2 Recording interval
Group size	100	100
Trans. rate par. ref. type (c)[1]	0.8 – 0.145	0.6
Recovery rate (α)	0.0476	0.0476
Average infectious period ($1/\alpha$)	21 days	21 days
Value susceptibility allele g (γ_g)	1	1
Value susceptibility allele G (γ_G)	0.3 – 1	0.4
Value infectivity allele f (φ_f)	1	1
Value infectivity allele F (φ_F)	0.3 – 1	0.4
Frequency allele g (p_g)	Beta(0.5,0.05)	Beta(0.5,0.05)
Frequency allele f (p_f)	Beta(0.5,0.05)	Beta(0.5,0.05)
Basic reproduction ratio (R_0)	3.0	3.0
Endemic reproduction ratio (R)[2]	2.1 to 3.0	2.4
Recording interval (% of $1/\alpha$)	66.6%	4.8 to 133.3%
Recording frequency	11 times (10 intervals)	11 times (10 intervals)

[1] Transmission rate parameter for the reference genotype $ggff$

[2] Details on the calculation of the endemic reproduction ratio are in the appendix

2.3 Results

Estimates in this section are averages of 200 replicates, except for Figure 2.1, which shows the result for one replicate. Infectivity estimates were not corrected for the geometric mean approximation because the error caused by this approximation was found to be small, as quantified (Tables 2.3 and 2.5).

Figure 2.1 shows an example of the percentage of infected individuals for each susceptibility and infectivity genotype during 100 days of an endemic. The distribution of the susceptibility genotypes within the infected individuals differed from the genotype frequency for susceptibility in the whole population (Figure 2.1a). The gg genotype was overrepresented in the infected individuals because this genotype had above-average susceptibility, while the GG genotype was underrepresented in the infected individuals. Most of infected individuals, however, had the Gg genotype, simply because there were more individuals with this genotype. An overview of genotype specific prevalences for different allele values is in Table A2.1 of the Appendix. The distribution of the infectivity genotypes within the infected individuals was similar to the genotype frequencies in the whole

population, because the susceptibility and infectivity loci were unlinked and in linkage equilibrium (Figure 2.1b).

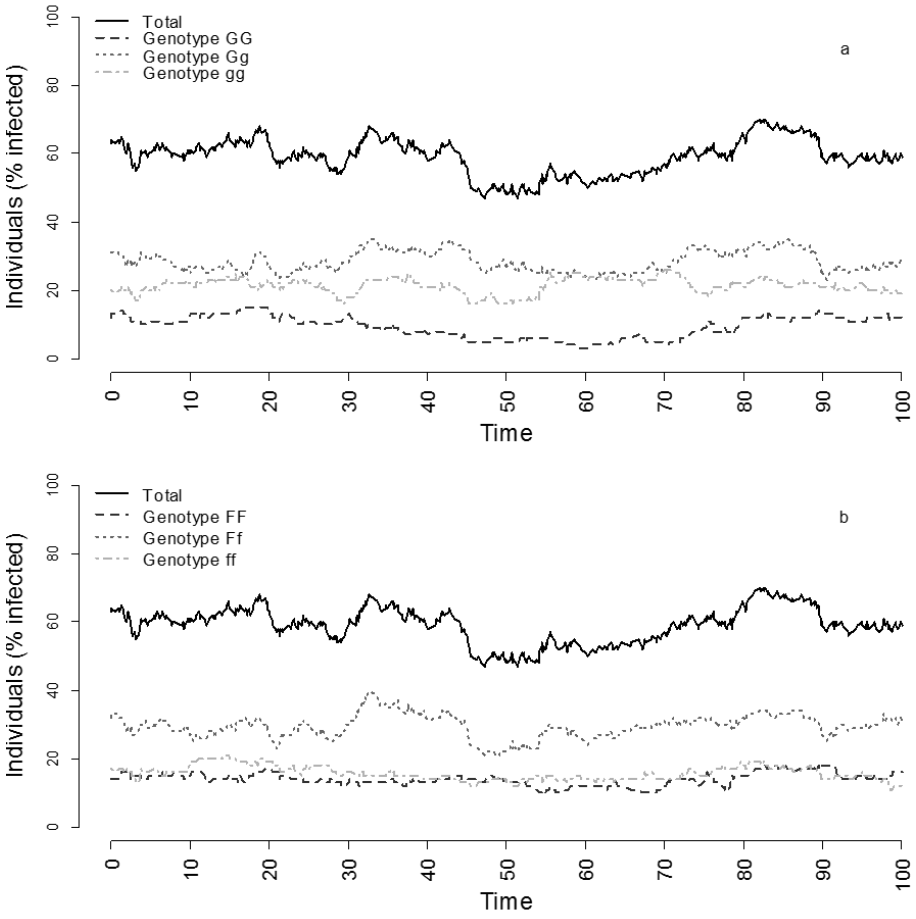


Figure 2.1. Percentage of infected individuals within a given susceptibility (panel a) and infectivity (panel b) genotype during 100 days of an endemic disease. Results are from one representative replicate with $p_g = p_f = 0.5$ and $\gamma_G = \varphi_F = 0.4$.

In scenario 1, we varied γ_G from 1 to 0.3, so that the susceptibility effect, $\gamma_g - \gamma_G$, varied from 0 to 0.7 (Tables 2.3 and 2.4). Since $\gamma_g = 1$, the susceptibility value of the G allele ranged from 100 to 30% of that of the g allele. Tables 2.3 and 2.4 show

the estimates of susceptibility and infectivity effects, bias, precision, and power for different allele effect sizes. All SNP effects were underestimated. As expected, absolute bias increased with absolute size of the effect, for both susceptibility and infectivity. However, relative bias decreased with absolute size of the effect. Precision was measured by the RMSE, with higher values indicating less precision.

Table 2.3. Estimates of the effect of susceptibility, bias, precision, and power for different allele effect sizes.

Input	Estimate	Bias		RMSE[2]	Power
$(\gamma_g - \gamma_G)$[1]	$(\gamma_g - \hat{\gamma}_G)$[1]	Absolute	Relative		
0.0	-0.001	-0.001	-0.1%	0.033	2%
0.1	0.087	-0.013	-13.4%	0.033	78%
0.2	0.173	-0.027	-13.3%	0.039	100%
0.3	0.265	-0.035	-11.7%	0.043	100%
0.4	0.358	-0.042	-10.5%	0.046	100%
0.5	0.457	-0.043	-8.6%	0.047	100%
0.6	0.558	-0.042	-7.1%	0.046	100%
0.7	0.663	-0.037	-5.3%	0.039	100%

[1] $\gamma_g = 1$

[2] Precision was measured by RMSE and results are averages of 200 replicates.

Table 2.4. Estimates of the effect of infectivity, bias, precision, power, and error caused by the geometric mean approximation (GMA).

Input	Estimate	Bias		RMSE[2]	Power	GMA error[3]
$(\varphi_f - \varphi_F)$[1]	$(\varphi_f - \hat{\varphi}_F)$[1]	Absolute	Relative			
0.0	-0.011	-0.011	-1.1%	0.215	2.0%	-0.0002
0.1	0.029	-0.071	-71.4%	0.212	5.0%	0.0001
0.2	0.125	-0.075	-37.5%	0.191	10.5%	0.0005
0.3	0.197	-0.103	-34.3%	0.185	23.0%	0.0008
0.4	0.279	-0.121	-30.2%	0.203	44.0%	0.0017
0.5	0.350	-0.150	-30.0%	0.222	60.0%	0.0029
0.6	0.449	-0.151	-25.2%	0.200	80.0%	0.0052
0.7	0.529	-0.171	-24.5%	0.203	90.5%	0.0082

[1] $\varphi_f = 1$

[2] Precision was measured by RMSE and results are averages of 200 replicates.

[3] $\hat{\varphi}_F - \hat{\varphi}_{F_{corrected\ for\ GMA}}$

For infectivity, the RMSE was 4.3 to 6.6 times higher than for susceptibility. There was no clear relationship between RMSE and true size of the SNP effect. For susceptibility, power to detect a SNP effect, defined as the probability to find a significant effect given that it exists, *i.e.*, the percentage of replicates with a significant SNP effect (P < 0.05), was 100% for all values of γ_G, except for $\gamma_g - \gamma_G = 0.1$, for which power was 78%. For infectivity, power increased from 5% for $\varphi_f - \varphi_F = 0.1$ to 90.5% for $\varphi_f - \varphi_F = 0.7$.

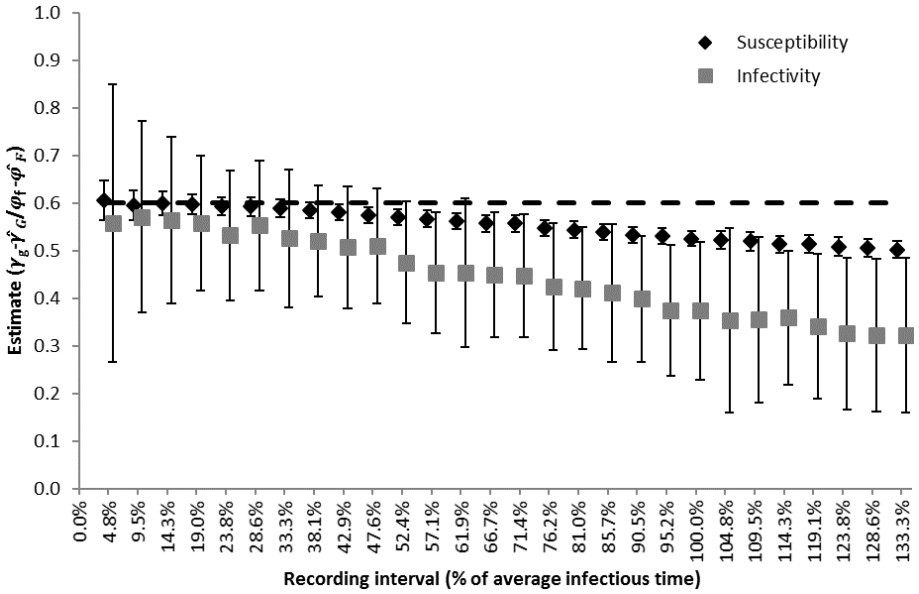


Figure 2.2. Susceptibility and infectivity estimates for different recording intervals. Markers show the estimates, which were averaged over 200 replicates. Input was $\gamma_g - \gamma_G = \varphi_f - \varphi_F = 0.6$ (dashed line). Error bars show the standard deviation among replicates on the original scale. Further inputs are in Table 2.2, Scenario 2.

In scenario 2, we varied the recording interval while keeping the total number of recordings constant. Figure 2.2 shows estimates of susceptibility and infectivity for different recording intervals. Table 2.5 shows the corresponding precision, power, and error caused by the geometric mean approximation (GMA).

Table 2.5. Precision, power, and error caused by the geometric mean approximation (GMA) for different recording intervals.

Recording interval as % of infectious time	RMSE[1]		Power (%)		GMA error[2]
	Susceptibility	Infectivity	Susceptibility	Infectivity	
4.8	0.042	0.294	100.0	47.0	0.0154
9.5	0.031	0.203	100.0	65.5	0.0133
14.3	0.025	0.178	100.0	75.5	0.0115
19.0	0.022	0.147	100.0	80.5	0.0105
23.8	0.020	0.152	100.0	80.5	0.0090
28.6	0.021	0.144	100.0	83.0	0.0088
33.3	0.022	0.163	100.0	84.5	0.0088
38.1	0.023	0.141	100.0	86.0	0.0080
42.9	0.026	0.158	100.0	84.0	0.0076
47.6	0.030	0.151	100.0	88.5	0.0074
52.4	0.034	0.179	100.0	85.5	0.0061
57.1	0.037	0.193	100.0	78,5	0.0053
61.9	0.042	0.213	100.0	81.5	0.0055
66.7	0.046	0.200	100.0	80.5	0.0052
71.4	0.046	0.200	100.0	82.5	0.0052
76.2	0.055	0.220	100.0	77.0	0.0045
81.0	0.059	0.219	100.0	79.5	0.0042
85.7	0.063	0.238	100.0	77.0	0.0041
90.5	0.069	0.241	100.0	75.5	0.0037
95.2	0.071	0.263	100.0	72.0	0.0032
100.0	0.076	0.268	100.0	65.0	0.0033
104.8	0.079	0.313	100.0	73.0	0.0028
109.5	0.083	0.301	100.0	68.5	0.0030
114.3	0.088	0.278	100.0	69.0	0.0028
119.1	0.089	0.301	100.0	66.0	0.0026
123.8	0.094	0.317	100.0	60.5	0.0024
128.6	0.096	0.320	100.0	66.0	0.0022
133.3	0.099	0.322	100.0	60.0	0.0023

[1] Precision was measured by RMSE and results are averages of 200 replicates. Further inputs are in Table 2.2, Scenario 2.

[2] $\hat{\varphi}_F - \hat{\varphi}_{F_{corrected\ for\ GMA}}$

For all intervals, SNP effects were underestimated, except for the 4.8%-interval, for which the susceptibility effect was slightly overestimated, $\hat{\gamma}_{G_{4.8\%}} = 0.605$. Underestimation increased with length of the recording interval, which was more pronounced for infectivity. For susceptibility, bias was smallest (-0.04%) for the 14.3% interval, while for infectivity, bias was smallest (-4.8%) for the 9.5% interval. For susceptibility, power was 100% for all intervals, while precision was highest from the 25% interval to the 50% interval, and decreased for longer and shorter intervals. For infectivity, both power and precision were highest from the 25% interval to the 50% interval, and decreased for longer and shorter intervals. We found the same optimal recording interval for susceptibility and infectivity with $\gamma_G = \varphi_F = 0.6$ (results not shown).

2.4 Discussion

Given the potential importance of genetic effects on infectivity, we developed a model to estimate effects of host SNPs on both susceptibility and infectivity. The model accounts for variation among susceptible individuals in the exposure to infectious herd mates, and for the genotypes of those herd mates. To test our model, we simulated an endemic disease in 10 groups of 100 individuals and recorded time-series data on individual disease status. For different SNP effects and recording intervals, we quantified bias and precision of model estimates. SNP effects were on average underestimated, thus estimates were conservative. Underestimation of SNP effects on infectivity increased with length of the recording interval. In spite of the limited sample size simulated, power to detect SNP effects for susceptibility was high. Power to detect effects for infectivity was lower but became higher than 60% when the allele effect size was greater than a factor of 0.5. The optimal recording interval was similar for susceptibility and infectivity, around 25 to 50% of the length of the average infectious period.

In the development of our model, we followed Anche et al. (2015), who considered epidemic diseases modelled by a SIR model. Given the importance of endemic diseases for livestock populations, we extended their approach to endemic diseases following a SIS model. Moreover, we considered time-series data on individual disease status, whereas Anche et al. (2015) considered the final disease status of individuals after an epidemic had gone through the population. With time-series data, more information is available on who infected whom and on the variation among susceptible individuals in exposure to infectious herd mates. This increases the accuracy of SNP-estimates, particularly for infectivity (Anacleto et al., 2015). We expect that our model can be easily extended to time-series data on

epidemic diseases that follow a SIR model, because the underlying principle is the same. Each time-period can be treated as an incomplete epidemic, where the number of susceptible and infectious individuals at the beginning of the period and the number of cases during the period must be recorded.

Both susceptibility effects and infectivity effects were underestimated, which was more pronounced for longer recording intervals, likely because of unobserved infections and recoveries in-between the recording time points. Regarding underestimation of the susceptibility effect, a case is missed when a susceptible individual becomes infected and also recovers within the same time interval. Since recovery rate was the same for all genotypes, the probability to miss a case was higher for genotypes that are more susceptible. Hence, genotypes with higher susceptibility have a larger proportion of missed cases, which reduces the estimate of the susceptibility effect. Regarding underestimation of the infectivity effect, we used the number of infectious individuals of each genotype at the start of the time-interval, $I_j(t)$, as explanatory variable in our model. However, there is loss and gain of infectious individuals during the interval because on the one hand, some of the initially infectious individuals may recover during the interval and thus no longer contribute to transmission, while on the other hand, some of the initially susceptible individuals may become infected during the interval and contribute to transmission from that time onwards. This loss and gain of infectious individuals is not accounted for by the model, which is more pronounced for longer intervals. In a (dynamic) equilibrium, the number of infectious individuals will, on average, tend to move towards its median value. Hence, the number of infectious individuals of a certain genotype at the beginning of the interval is systematically more extreme than the actual number of infectious individuals of that genotype averaged over the interval. Thus, in the model, the variance of the $CountF$-term is systematically too large, especially for longer intervals. This explains underestimation of the infectivity effect (*i.e.*, c_2) and the increase of this underestimation when the recording interval is longer. However, when the recording interval is short, there are no or only a few infections within an interval and, thus, the number of cases is too limited for precise estimations. Thus, given a fixed total number of recordings, short recording intervals lead to reduced precision of estimates, whereas long intervals lead to bias (Figure 2.2). When the number of recordings is unlimited, the optimal recording interval will be short because the large number of records compensates for the limited precision of individual records but not for their bias.

An assumption of our model is that cases within an interval are caused by the infected individuals at the beginning of that interval. Thus, there is a gain and loss of

infectious individuals that is not accounted for by the model. The impact of this error depends on the number of cases and the number of recoveries relative to the number of infected individuals at the beginning of the interval. In an endemic equilibrium, the number of cases within an interval equals, on average, the number of recoveries within an interval, $C = \alpha I$. Hence, when expressed relative to the number of infected individuals at the beginning of the interval, the number of cases and the number of recoveries are both defined by the recovery rate α. Thus, the impact of the error caused by the assumption is determined by α, which suggests that the recovery rate (which equals the incidence in the endemic equilibrium) determines the optimum recording interval, rather than prevalence.

Estimates of genetic effects on infectivity were less accurate than those on susceptibility. This is partly because infectivity is expressed only by the infected individuals. Furthermore, there is a trade-off between the quality of the susceptibility and infectivity estimates in relation to group size (Anacleto et al., 2015). In large groups, more information is available on the order in which individuals become infected, which leads to better susceptibility estimates, while in small groups it is easier to establish who infected whom, which leads to better infectivity estimates. Because large groups have multiple infected individuals at any given point in time, genetic differences in infectivity have to be estimated indirectly from the number of susceptible group mates that become infected and from the genotype fractions among the infected individuals at different points in time. Thus, especially in populations that consist of large groups, more records and groups are needed to estimate genetic effects on infectivity than on susceptibility.

We assumed that allele effects on susceptibility and infectivity were additive on the log-scale, such that the model could be formulated in terms of allele counts within individuals and the model could be tested without introducing estimation errors that might be present with additive allele effects on the original scale. Allele effects were, therefore, multiplicative on the original scale. With multiplicative allele effects, negative dominance is introduced on the original scale. The magnitude of the dominance relative to the additive effect, denoted as d/a following Falconer et al. (1996) is:

$$\frac{d}{a} = \frac{\gamma_{Gg} - 0.5(\gamma_{gg} + \gamma_{GG})}{0.5(\gamma_{gg} - \gamma_{GG})},$$

with $\gamma_G < \gamma_g$. So, for example, for a twofold effect with $\gamma_G = 0.5$ and $\gamma_g = 1.0$, the dominance deviation is one third of the additive effect. For a tenfold effect, $d/a = -0.81$. Hence, in our model, alleles that cause a large increase in susceptibility or

infectivity are almost completely recessive. Recessive alleles for susceptibility may be plausible because selection against recessive alleles with detrimental effects on fitness is inefficient, particularly when the frequency of the recessive allele is low. Hence, alleles that cause a large increase in susceptibility but are still segregating are probably recessive. Whether completely recessive alleles for infectivity are also plausible, is unknown at present.

An alternative perspective is that our model estimates the average effects of alleles on the log-scale, regardless of presence or absence of dominance on the log-scale. This is analogous to using ordinary additive models for estimating SNP effects, where the model captures the full average effect (α) of an allele, including the relevant dominance component ($\alpha = a + (q - p)d$; (Falconer et al., 1996)).

We determined $\hat{\alpha}$ for additive and multiplicative allele effects, to determine the impact on estimates when allele effects are additive on the original scale instead of multiplicative. Input values for the additive simulation were $\gamma_{GG} = \varphi_{FF} = 0.16$, $\gamma_{Gg} = \varphi_{Ff} = 0.58$, and $\gamma_{gg} = \varphi_{ff} = 1$. So, with $p = q = 0.5$, the average effect $\alpha = 0.42$ (Falconer et al., 1996). Estimates were $\hat{\gamma}_G = 0.45$ and $\hat{\gamma}_F = 0.63$, such that $\hat{\alpha} = \frac{\hat{\gamma}_{gg} - \hat{\gamma}_{GG}}{2} = \frac{1 - 0.45^2}{2} = 0.40$ for susceptibility and $\hat{\alpha} = 0.30$ for infectivity. For the multiplicative simulation, input values were $\gamma_{GG} = \varphi_{FF} = 0.16$, $\gamma_{Gg} = \varphi_{Ff} = 0.4$, and $\gamma_{gg} = \varphi_{ff} = 1$, such that, with $p = q = 0.5$, the average effect $\alpha = 0.42$. Estimates were $\hat{\gamma}_G = 0.44$ and $\hat{\gamma}_F = 0.55$, so $\hat{\alpha} = 0.40$ for susceptibility, and $\hat{\alpha} = 0.35$ for infectivity. This suggests that our model performs worse if allele effects are additive on the original scale instead of multiplicative.

We estimated the effect of two SNPs, one for infectivity and one for susceptibility, without fitting the effect of other genes that may affect these traits. This approach is similar to genome-wide association studies (GWAS) or candidate gene studies, where SNP effects are often fitted one at a time. Hence, the model presented here can be used as a starting point to explore and identify which loci affect the trait of interest. One approach could be to estimate both the susceptibility effect and the infectivity effect of the same SNP, one SNP at a time. This would imply full linkage disequilibrium (LD) between the susceptibility SNP and the infectivity SNP, because they are one and the same SNP. However, in contrast to GWAS for ordinary ("direct") traits, this would not imply full confounding of the two effects, because they are expressed in phenotypes of distinct individuals. Nevertheless, both effects may be partially confounded because herd mates are usually related. Hence, for GWAS, further research is required to investigate the effect of LD between SNPs for susceptibility and infectivity. Note that, while we considered absence of LD between loci in the simulated data, the statistical model that we developed

(Equation 2.2) does not make this assumption, because SNP effects are simply fitted as fixed effects in our model. Thus, estimates of SNP effects represent *partial* regression coefficients and, therefore, account for LD. As in any single-SNP GWAS, there may be genes elsewhere in the genome that affect the same trait and show LD with the SNP of interest. Such genes would bias estimates of the SNP of interest. Hence, after an initial single-SNP GWAS, the significant SNPs should ideally be fitted simultaneously, in order to account for LD. Moreover, in GWAS studies, significance thresholds need to account for multiple testing to avoid many false positives, and GWAS studies need to take population stratification into account. For traits affected by direct effects only, stratification can be accounted for by including a random polygenic effect in the model, with a covariance-structure given by a genomic relationship matrix. For infectious disease data, that model would need to be extended with polygenic effects for infectivity of infected contact individuals (Anacleto et al., 2015). The latter model may also be suitable for genomic prediction, where the purpose is to estimate breeding values of individuals, rather than single gene effects.

Anche et al. showed that relatedness within groups resulted in better estimates of susceptibility and infectivity (Anche et al., 2015). When relatedness within groups is high, individuals with above/below average susceptibility will also have group mates with above/below average susceptibility, and individuals with above/below average infectivity will also have group mates with above/below average infectivity. Relatedness within groups, therefore, increases variation between groups, which improves the estimates (Anche et al., 2015). However, results from the field of indirect genetic effects indicate that relatedness may lead to confounding of direct and indirect effects. For example, when groups consist of a single family, direct and indirect effects are fully confounded (Bijma et al., 2007a). This result may extend to infectious disease data when loci for susceptibility and infectivity are in LD. Further research is needed to identify the optimal group structure with respect to relatedness for estimating genetic effects on susceptibility and infectivity.

Knowledge of the amount of genetic variation in infectivity is very limited at present. In general, natural selection has a tendency to exhaust heritable variation in traits related to individual fitness. Infectivity, however, is an indirect genetic effect, that affects disease status of other individuals rather than that of the individual itself. Natural selection targets such indirect genetic effects only in the presence of feed-back mechanisms, such as with kin and group selection (Bijma and Wade, 2008). Even in the presence of such feed-back mechanisms, selection on indirect genetic effects is weaker than on direct genetic effects (Bijma, 2010). Thus, infectivity may have been less exposed to natural selection and may exhibit more genetic variation.

Presence of genetic variation is also suggested by the existence of super spreaders (Stein, 2011). The model presented here can be used as a starting point to determine the amount of genetic variation that is present for infectivity in populations. This may also help to better estimate effects on susceptibility because the model accounts for variation among susceptible individuals in their exposure to infectious herd mates and for the genotypes of those herd mates.

When our model is extended with the relevant polygenic effects (as discussed previously), it can be used to estimate SNP effects on susceptibility and infectivity, in particular when more data on disease status and genotype become available. Opportunities to measure disease status on a regular basis lie in the increasing number of sensor systems that are used and will be used in the future (Steeneveld and Hogeveen, 2015). Current sensor systems are able to record animal activity, temperature, cells in milk, etc. In the future, these types of sensor data may provide regular information about the disease status of an animal. In addition, the number of animals that are genotyped increases rapidly. Combining the model developed here with genotype and sensor data may considerably enhance breeding against infectious diseases in livestock.

2.5 Conclusions

We developed a generalized linear mixed model to estimate SNP effects on both host susceptibility and host infectivity from time-series data on individual disease status for an endemic disease. In contrast to common models used in animal breeding, our model accounts for variation among susceptible individuals in their exposure to infectious herd mates and for the genotypes of those herd mates. With the use of simulated data, we quantified bias and precision of SNP effects estimated by the model and showed that the optimal recording interval is between 25 and 50% of the average infectious period when disease status is observed 11 times. When the recording interval was close to optimal, SNP effects were on average slightly underestimated. Infectivity estimates were less precise than susceptibility estimates. In future genome-wide association studies, the model presented here may be useful to estimate SNP effects that affect disease transmission and disease prevalence.

2.6 Acknowledgements

This work is part of the research programme “Meer met minder” with project number 847.13.004, which is financed by the Netherlands Organisation for Scientific Research (NWO).

2.7 References

Anacleto, O., Garcia-Cortés, L.A., Lipschutz-Powell, D., Woolliams, J.A., Doeschl-Wilson, A.B., 2015. A novel statistical model to estimate host genetic effects affecting disease transmission. Genetics 201, 871-884.

Anche, M., de Jong, M., Bijma, P., 2014. On the definition and utilization of heritable variation among hosts in reproduction ratio R0 for infectious diseases. Heredity 113, 364-374.

Anche, M.T., Bijma, P., De Jong, M.C., 2015. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47, 1-15.

Anderson, R.M., May, R.M., Anderson, B., 1992. Infectious diseases of humans: dynamics and control. Wiley Online Library.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Bennett, R., IJpelaar, J., 2005. Updated estimates of the costs associated with thirty four endemic livestock diseases in Great Britain: a note. Journal of Agricultural Economics 56, 135-144.

Bijma, P., 2010. Fisher's fundamental theorem of inclusive fitness and the change in fitness due to natural selection when conspecifics interact. Journal of evolutionary biology 23, 194-206.

Bijma, P., Muir, W.M., Ellen, E.D., Wolf, J.B., Van Arendonk, J.A., 2007a. Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. Genetics 175, 289-299.

Bijma, P., Muir, W.M., Van Arendonk, J.A., 2007b. Multilevel selection 1: quantitative genetics of inheritance and response to selection. Genetics 175, 277-288.

Bijma, P., Wade, M., 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. Journal of evolutionary biology 21, 1175-1188.

Binder, S., Levitt, A.M., 1998. Preventing emerging infectious diseases: a strategy for the 21st century. US Department of Health and Human Services, Centers for Disease Control and Prevention.

Broom, D.M., 2006. Behaviour and welfare in relation to pathology. Applied Animal Behaviour Science 97, 73-83.

Brueggemann, A.B., Pai, R., Crook, D.W., Beall, B., 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States.

Chase-Topping, M., Gally, D., Low, C., Matthews, L., Woolhouse, M., 2008. Super-shedding and the link between human infection and livestock carriage of Escherichia coli O157. Nature Reviews Microbiology 6, 904-912.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J., 1990. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28, 365-382.

Falconer, D.S., Mackay, T.F., Frankham, R., 1996. Introduction to quantitative genetics (4th edition). Trends in Genetics 12, 280.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry 81, 2340-2361.

Heffernan, J., Smith, R., Wahl, L., 2005. Perspectives on the basic reproductive ratio. Journal of the Royal Society Interface 2, 281-293.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. The Royal Society, pp. 700-721.

Lipschutz-Powell, D., Woolliams, J.A., Bijma, P., Doeschl-Wilson, A.B., 2012a. Indirect genetic effects and the spread of infectious disease: are we capturing the full heritable variation underlying disease prevalence?

Lipschutz-Powell, D., Woolliams, J.A., Bijma, P., Pong-Wong, R., Bermingham, M.L., Doeschl-Wilson, A.B., 2012b. Bias, accuracy, and impact of indirect genetic effects in infectious diseases. Frontiers in genetics 3.

Lipschutz-Powell, D., Woolliams, J.A., Doeschl-Wilson, A.B., 2014. A unifying theory for genetic epidemiological analysis of binary disease data. Genetics Selection Evolution 46, 1-12.

Moore, A.J., Brodie III, E.D., Wolf, J.B., 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution, 1352-1362.

Neu, H.C., 1992. The crisis in antibiotic resistance. Science 257, 1064-1073.

R Core Team, 2017. R: A Language and Environment for Statistical Computing, 3.4.0 ed. R Foundation for Statistical Computing, Vienna, Austria.

Roberts, M., Heesterbeek, H., 1993. Bluff your way in epidemic models. Trends in microbiology 1, 343-348.

Springbett, A., MacKenzie, K., Woolliams, J., Bishop, S., 2003. The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165, 1465-1474.

Steeneveld, W., Hogeveen, H., 2015. Characterization of Dutch dairy farms using sensor systems for cow management. Journal of dairy science 98, 709-717.

Stein, R.A., 2011. Super-spreaders in infectious diseases. International Journal of Infectious Diseases 15, e510-e513.

Velthuis, A., De Jong, M., Kamp, E., Stockhofe, N., Verheijden, J., 2003. Design and analysis of an Actinobacillus pleuropneumoniae transmission experiment. Preventive veterinary medicine 60, 53-68.

Wilson, J., Nokes, D., Carman, W., 1998. Current status of HBV vaccine escape variants–a mathematical model of their epidemiology. Journal of viral hepatitis 5, 25-30.

Wolf, J.B., Brodie III, E.D., Cheverud, J.M., Moore, A.J., Wade, M.J., 1998. Evolutionary consequences of indirect genetic effects. Trends in Ecology & Evolution 13, 64-69.

Woolhouse, M., Stringer, S., Matthews, L., Hunter, N., Anderson, R., 1998. Epidemiology and control of scrapie within a sheep flock. Proceedings of the Royal Society of London. Series B: Biological Sciences 265, 1205-1210.

2.8 Appendix

2.8.1 Distribution of susceptible genotypes in the endemic equilibrium

In the absence of heterogeneity, the expected equilibrium prevalence follows from the basic reproduction ratio R_0, and can be calculated as $1 - \frac{1}{R_0}$ (Anderson et al., 1992; Heffernan et al., 2005). R_0 can be expressed as a function of average susceptibility $\bar{\gamma}$, average infectivity $\bar{\varphi}$, the transmission rate parameter of the reference genotypes c, and the recovery rate parameter α, $R_0 = \bar{\gamma}\bar{\varphi}\frac{c}{\alpha}$ (Anche et al., 2014). Average susceptibility was calculated as $\bar{\gamma} = \sum_i p_i \gamma_i$, and average infectivity as $\bar{\varphi} = \sum_j p_j \varphi_j$, where i indexes susceptibility genotypes, j indexes infectivity genotypes, and p_i is the frequency of genotype i in the population.

However, with heterogeneity the equilibrium prevalence differs from the above result. At the start of an endemic phase, only few individuals are infected. Therefore, the *susceptible* fraction with susceptibility genotype i $\left(fracS_i = \frac{S_i}{S}\right)$ is similar to the genotype frequency of susceptibility genotype i (p_i) in the population, $fracS_i \approx p_i$. In the endemic equilibrium, however, highly susceptible individuals have more chance to get infected, so the *susceptible* fraction with susceptibility genotype i differs from the genotype frequency of susceptibility genotype i in the population, $fracS_i \neq p_i$. Thus, in the endemic equilibrium, the average susceptibility of the susceptible individuals is lower than the average susceptibility in a totally susceptible population, therefore, the average susceptibility equals:

$$\bar{\gamma}(t) = \sum_i fracS_i(t) * \gamma_i. \qquad \text{(Equation A2.1)}$$

A lower average susceptibility in the equilibrium leads to a lower reproduction ratio and, therefore, a lower equilibrium prevalence as expected from the initial reproduction ratio. The reproduction ratio at time t, $R(t)$, in a population that is no longer fully susceptible is given by:

$$R(t) = \bar{\gamma}(t) * \bar{\varphi} * \frac{c}{\alpha}. \qquad \text{(Equation A2.2)}$$

Because the susceptibility locus and the infectivity locus were in linkage equilibrium, the *infected* fraction with infectivity genotype j $\left(fracI_j = \frac{I_j}{I}\right)$ will be similar to the total fraction with infectivity genotype j in the population, $fracI_j \approx p_j$.

As we approach the equilibrium, the *susceptible* fraction with susceptibility genotype i at time $t+1$ can be calculated from the *susceptible* fraction with susceptibility genotype i at time t, and the corresponding $R(t)$, by:

$$fracS_i(t+1) = \frac{p_i}{\frac{1}{R(t)} + \frac{\gamma_i}{\bar{\gamma}}\left(1 - \frac{1}{R(t)}\right)}. \qquad \text{(Equation A2.3)}$$

By using Equations A2.1, A2.2, and A2.3 in an iterative process, the *susceptible* fraction with susceptibility genotype i in the endemic equilibrium was found. In the endemic equilibrium three conditions were met:

(i) $\frac{S_i}{N} = \frac{1}{R} fracS_i$ for R > 1

(ii) $\frac{I_i}{N} = \left(1 - \frac{1}{R}\right) fracS_i \frac{\gamma_i}{\bar{\gamma}}$ for R > 1

(iii) $\frac{S_i + I_i}{N} = p_i.$

Therefore, the genotype specific prevalences were known (conditions (i) and (ii)). The basic reproduction ratio, the reproduction ratio in the equilibrium and the genotype-specific prevalences for different effects are in Table A2.1.

In this study, we started the endemic in the equilibrium. The distribution of the *infected* fraction of susceptibility genotypes in endemic equilibrium was obtained by a grid search for the point where conditions (i), (ii) and (iii)were met (note that the fastest way to reach the equilibrium goes through fractions that are in reality not possible, *i.e.*, the path to the equilibrium is not real).

Table A2.1. Basic reproduction ratio and prevalence for different susceptibility effects.

Value susceptibility allele G (γ_G)	Transmission rate parameters for reference type (c)[1]	Basic reproduction ratio[2]		Prevalence			
				Total	Per susceptibility genotype		
		Classic (R_0)	Equilibrium (R)		GG	Gg	gg
0.3	0.8	3.00	2.10	0.52	0.25	0.53	0.79
0.4	0.6	3.03	2.39	0.58	0.36	0.59	0.78
0.5	0.45	3.00	2.59	0.61	0.45	0.62	0.77
0.6	0.35	3.01	2.77	0.64	0.52	0.64	0.75
0.7	0.28	3.07	2.94	0.66	0.58	0.66	0.74
0.8	0.22	3.03	2.98	0.66	0.61	0.67	0.71
0.9	0.18	3.08	3.07	0.67	0.65	0.67	0.70
1.0	0.145	3.045	3.045	0.67	0.67	0.67	0.67

[1] Reference genotype is *ggff*

[2] $p_g = p_f = 0.5$, $\alpha = 0.0476$, $\gamma_g = \varphi_f = 1$ and $\varphi_F = \gamma_G$.

3

Digital Dermatitis in dairy cattle: the contribution of different disease classes to transmission

Floor Biemans[1,2], Piter Bijma[2], Natasja M. Boots[1], Mart C.M. de Jong[1]

[1] Quantitative Veterinary Epidemiology Group, Wageningen University & Research, 6700 AH Wageningen, the Netherlands
[2] Animal Breeding and Genomics, Wageningen University & Research, 6700 AH Wageningen, the Netherlands

Epidemics (2017), in press

Abstract

Digital Dermatitis (DD) is a claw disease mainly affecting the hind feet of dairy cattle. Digital Dermatitis is an infectious disease, transmitted via the environment, where the infectious "agent" is a combination of bacteria. The standardized classification for DD lesions developed by Döpfer et al. (1997) and extended by Berry et al. (2012) has six distinct classes: healthy (M0), an active granulomatous area of 0-2 cm (M1), an ulcerative lesion of >2 cm (M2), an ulcerative lesion covered by a scab (M3), alteration of the skin (M4), and a combination of M4 and M1 (M4.1).

We hypothesize that classes M1, M2, M3, M4, and M4.1 are the potentially infectious classes that can contribute to the basic reproduction ratio (R_0), the average number of new infections caused by a typical infected individual. Here, we determine differences in infectivity between the classes, the sojourn time in each of the classes, and the contribution of each class to R_0.

The analysis is based on data from twelve farms in the Netherlands that were visited every two weeks, eleven times.

We found that 93.89% of the transitions from M0 was observed as a transition to class M4, and feet with another class-at-infection rapidly transitioned to class M4. As a consequence, about 70% of the infectious time was spent in class M4. Transmission rate parameters of class-at-infection M1, M2, M3, and M4 were not significantly different from each other, but differed from class-at-infection M4.1. However, due to the relative large amount of time spent in class M4, regardless of the class-at-infection, R_0 was almost completely determined by this class. The R_0 was 2.36, to which class-at-infection M4 alone contributed 88.5%.

Thus, M4 lesions should be prevented to lower R_0 to a value below one, while painful M2 lesions should be prevented for animal welfare reasons.

3.1 Introduction

Digital Dermatitis (DD) is a claw disease discovered in 1974 in cattle in Italy by Cheli and Mortellaro (Cheli and Mortellaro, 1974). The disease (mainly) affects the hind feet of dairy cattle (Read and Walker, 1998; Sogstad et al., 2005). Round lesions occur along the coronary band of the claws, above the interdigital space next to the heel bulbs (Walker et al., 1995). Lesions can be painful, prone to bleed, develop filiform papillae, and can be surrounded by hyperkeratotic skin with hairs longer than normal (Read and Walker, 1998).

Digital Dermatitis is an infectious disease that is transmitted via the "environment"; environment is defined as any possible pathogen reservoir through which the infection can spread. The infectious "agent" is a combination of bacteria (Rodríguez-Lainz et al., 1996; Read and Walker, 1998; Demirkan et al., 1999; Sogstad et al., 2005; Vink et al., 2009), the most common bacteria present in DD lesions are spirochetes of the genus *Treponema spp.* (Clegg et al., 2015). Digital Dermatitis is associated with lameness; cows that are severely affected bear their weight on the toes of the affected foot, shake the foot as if in pain, and show reluctance to move (Bassett et al., 1990; Collighan and Woodward, 1997; Read and Walker, 1998).

A standardized classification for DD lesions was developed by Döpfer et al. (1997) and was more extensively described by Berry et al. (2012). This classification comprises six distinct classes (M0, M1, M2, M3, M4, and M4.1). Class M0 is described as skin where lesions are macroscopically absent, class M1 as an active granulomatous area of 0-2 cm, class M2 as an ulcerative lesion of >2 cm, class M3 as an ulcerative lesion covered by a scab, class M4 as alteration of the skin with hyperkeratotic lesions that can have a proliferative aspect, and class M4.1 as altered skin (M4) with a painful focus (M1) (Döpfer et al., 1997; Döpfer, 2009; Berry et al., 2012). Class M1, M2, and M4.1 are classes that describe circumscribed, red-greyish, moist, painful, and prone to bleed lesions (Speijers et al., 2010; Berry et al., 2012; Zinicola et al., 2015). Studies on DD tend to focus on these lesions because they can cause lameness.

Here, we investigate what the contribution of the different classes to transmission is with the basic reproduction ratio R_0. The R_0 is the expected number of secondary cases that arise from one typical infectious individual in a fully susceptible population during its entire infectious period (Diekmann et al., 1990). When $R_0 < 1$, a typical infectious individual infects on average less than one other individual and the disease dies out. We hypothesize that classes M1, M2, M3, M4, and M4.1 are the infectious classes that contribute to R_0. When there is variation between classes in infectivity or in sojourn time, the contribution of each class to R_0

may differ. We determine how R_0 is composed by investigating the distribution of the first observed classes after infection, the average sojourn time in each class, and the infectivity of each class.

3.2 Material and methods

3.2.1 Data collection

Between November 2014 and April 2015, twelve farms in the Netherlands were visited eleven times with a two-week interval between visits. Criteria for farms to be selected included a ≥ 20% DD prevalence based on hoof trimming records from the previous year, and the presence of a milking parlour. Two trained observers scored the hind feet of all lactating cows in the milking parlour. Feet were cleaned with a medium pressure water hose, and were macroscopically examined with the use of a strong flashlight and a swivelling mirror (Relun et al., 2011). Feet were scored according to the classification developed by Döpfer et al. (1997) and Berry et al. (2012). Both observers were always present at a farm. An observer either rinsed and scored the feet, or recorded cow ID and the disease class. The role of an observer could alter between farms.

Missing values occurred when a cow was dried off or removed from the population for another reason. Farmers were not informed on the disease status of the cows. Farmers were, however, allowed to identify lesions themselves, and treat cows using their normal routine. Table 3.1 gives an overview of the characteristics of the farms enrolled in the study.

To assess agreement between observers, Cohen's kappa coefficient (Viera and Garrett, 2005) was calculated once immediately before, and two times during data collection. Kappa's coefficient is a measure of the difference between the observed and expected agreement. It is expressed on a -1 to 1 scale, where negative values indicate systematic disagreement between observers, 0 is agreement that would be expected by chance, and 1 is perfect agreement.

3.2.2 Methodology

We calculated R_0 based on methods of Diekmann et al. (2009) and Döpfer et al. (2012b). Digital Dermatitis has multiple infected classes that can all be consecutively presented by a single cow. The *sojourn* time in the infectious classes needs to be taken into account when calculating R_0. The class that is first observed upon infection will be called the *class-at-infection* (Diekmann et al., 2009). After the class-at-infection, a foot may reside in multiple other classes before returning to the susceptible class (M0). The sojourn time in the other classes can depend on the class-at-infection. So for each class-at-infection, the length of the infectious period in each

Table 3.1. Characteristics of the farms enrolled in the study.

Farm	# Cows examined[1]	Outdoor access[2]	Floor type	Manure scraper	# Footbaths[3]	# Observations	# Transitions observed	Average *Δt* (days)	Prevalence (SD)[4]	
									Cow level	Foot level
A	134	Yes	Concrete slatted	No	7	11	2700	14	78.0 (5.4)	69.6 (6.6)
B	105	Yes	Concrete slatted	Yes	0	11	2140	14	56.3 (7.5)	46.9 (7.9)
C	159	No	Concrete and rubber slatted	Yes	5	11	3280	14	49.7 (2.8)	40.2 (1.9)
D	118	Yes	Concrete slatted	Yes	7	11	2380	14	57.8 (5.0)	49.2 (5.1)
E	102	Yes	Concrete slatted	Yes	9	11	2040	13.60	62.8 (5.0)	54.6 (5.4)
F	133	No	Concrete slatted	Yes	10	10	2448	15.56	59.2 (10.0)	48.7 (10.4)
G	100	Yes	Concrete slatted	No	3	11	2000	14	65.6 (8.1)	58.2 (7.6)
H	189	Yes	Concrete slatted	Yes	7	11	3780	14	64.9 (6.2)	56.7 (5.8)
I	104	Yes	Concrete slatted	No	0	11	2080	14	56.4 (5.1)	45.6 (4.9)
J	88	Yes	Concrete slatted	No	0	11	1760	14	65.8 (10.8)	58.1 (10.9)
K	130	Yes	Concrete slatted	Yes	13	9	2144	14	63.6 (9.6)	52.5 (8.5)
L	151	No	Concrete slatted	Yes	3	11	3040	13.90	70.9 (7.2)	62.0 (7.7)

[1] Total number of different cows examined during the study period.

[2] During the study period all cows were housed indoors.

[3] Number of footbaths given during the study period.

[4] Average percentage scored as class M1, M2, M3, M4, or M4.1 and the standard deviation (SD).

class can be unique. During the infectious period feet with class-at-infection *i* have a certain average infectivity, measured by the transmission rate parameter β_i, *i* denoting class-at-infection M*i*. For a foot with class-at-infection M*i,* the β_i is a function of the infectivity of all the classes weighted by the sojourn time in these classes.

We account for the transmission of DD via the "environment" (Laven, 2001). Here, the environment is defined as any possible pathogen reservoir through which transmission can occur, including *e.g.*, the gastrointestinal tract, the nasal cavity, human caretakers or the actual environment. We assume that feet that are infected contribute fully to the current environmental reservoir, while feet that were infected at an earlier stage still contribute partly to the current environmental reservoir. The contribution to the environmental reservoir of feet that were infected earlier is assumed to decrease each interval Δt with factor λ, which may be interpreted as a survival rate. So from a foot that was infectious at *t*, the amount of pathogens that are in the environment at *t*+1is a fraction λ, at *t*+2 a fraction λ^2, at *t*+3 a fraction λ^3, etc.

The R_0 is the expected number of secondary cases that arise from one typical infectious individual in a fully susceptible population during its entire infectious period (Diekmann et al., 1990). In general, R_0 is the product of a transmission rate parameter (β) and the average infectious period (*x*), $R_0 = \beta x$. Because DD has multiple classes-at-infection, each with a possibly unique transmission rate parameter and infectious period, we need to take into account all the classes-at-infection in the calculation of R_0. The R_0 is, therefore, a function of the probability with which class-at-infection M*i* is entered (θ_i), and the transmission rate parameter and infectious period of the classes-at-infection, $R_0 = \sum_i \theta_i \beta_i x_i$. Furthermore, feet that were infectious previously can still contribute to the current environmental reservoir. The total contribution of a foot that was infectious at *t* to the environmental reservoir is the summed contribution of that foot over an infinite number of periods, $1 + \lambda + \lambda^2 + \lambda^3 + \ldots = (1 - \lambda)^{-1}$. Therefore, the full equation for R_0 is,

$$R_0 = (1 - \lambda)^{-1} \sum_i \theta_i \beta_i x_i. \qquad \text{(Equation 3.1)}$$

The elements of the sum represent the contribution of class-at-infection *i* to R_0. Each element of this equation was estimated from the data, as explained below.

3.2.2.1 Distribution of the classes-at-infection (θ_i)

The infectious period starts with class-at-infection M*i* (Diekmann et al., 2009). Given that an infection took place, class-at-infection M*i* is entered with probability θ_i, with *i* = 1, 2, 3, 4, or 4.1. Probability θ_i was estimated as the fraction of feet that were scored as M0 at *t* and as M*i* at *t+1*. Only transitions from M0 to another class are taken into account, so that $\sum_i \theta_i = 1$.

3.2.2.2 Sojourn time (xi)

After the class-at-infection a foot may reside in multiple other classes before returning to class M0. To estimate the sojourn time in each class, the transitions between infected classes over a period *Δt* were compiled in matrix **Σ** with elements $p_{k,l}$ (*k* = *l* = 1, 2, 3, 4, or 4.1) ,

$$\boldsymbol{\Sigma} = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,4.1} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,4.1} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} & p_{3,4.1} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} & p_{4,4.1} \\ p_{4.1,1} & p_{4.1,2} & p_{4.1,3} & p_{4.1,4} & p_{4.1,4.1} \end{pmatrix}.$$

Here, $p_{k,l}$ is the proportion of feet in class M*l* that have moved to class M*k* at the next observation. Throughout, subscripts *k* and *l* indicate the infected classes (M*k* and M*l*) in which a foot may reside, while subscript *i* specifically indicates the class-at-infection (M*i*). For example, $p_{2,1}$ is the proportion of the feet in class M1 that have moved to class M2 at the next observation. Because feet may recover (return to M0), the sum of the elements in a column of matrix **Σ** can be smaller than one, *i.e.,* $\sum_k p_{k,l} \leq 1$.

The sojourn time in each class, given the class-at-infection, equals (Diekmann and Heesterbeek, 2000),

$$\mathbf{X} = (\mathbf{I} - \boldsymbol{\Sigma})^{-1}, \qquad \text{(Equation 3.2)}$$

where, **X** is a matrix with elements $x_{k,i}$.The elements represent the average sojourn time in class M*k* given the class-at-infection M*i* measured in observation intervals. So, the average total duration of the infectious period per class-at-infection (x_i) is given by the sum of the elements in the columns of this matrix, $x_i = \sum_k x_{k,i}$. The **I** is an identity matrix with the same size as **Σ.**

3.2.2.3 Transmission rate parameter (βi)

Because it was assumed that an environmental reservoir contributes to transmission of DD (Laven, 2001), a Susceptible-Infectious-Susceptible model (SIS-model) with an environmental reservoir (E) was formulated. (Figure 3.1) (see e.g., de Rueda et al., 2015).

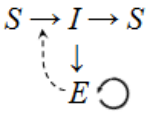


Figure 3.1. Susceptible-Infectious-Susceptible-model (SIS-model), with an extra environment route (E).

In Figure 3.1 *S* is the number of susceptible feet, *I* is the number of infected feet, and *E* is the infectious pressure coming from the environmental reservoir.

In classic epidemiological models the infection rate of susceptible individuals equals $\beta S \frac{I}{N}$, where S + I = N, and β is the transmission rate parameter (Kermack and McKendrick, 1927; Roberts and Heesterbeek, 1993). In these models there is only one infectious class, so the transmission rate parameter is a direct reflection of the infectivity. However, with DD, a foot can go through multiple infectious classes while it is infectious. We allow for possible variation in infectivity between the infectious classes by defining φ_k as the (relative) infectivity of class M*k*. Thus, with multiple infectious classes *k* at time *t*, the infection rate at *t* is,

$$Infection\ rate_t = S_t \sum_k \varphi_k \frac{E_{k,t}}{N_t}, \qquad \text{(Equation 3.3)}$$

where $E_{k,t}$ is the contribution of class *k* to environmental reservoir at time *t*, and N_t the total number of feet (susceptible and infected) at time *t*. With survival rate λ, $E_{k,t} = I_{k,t} + \lambda E_{k,t-1}$, where $I_{k,t}$ is the number of feet of class *k* that are infectious at *t*, and $\lambda E_{k,t-1}$ is the contribution to the environmental reservoir at time *t* of the feet of class *k* that were infectious at an earlier stage. In the analysis, the number of infections (cases) during an interval is connected to the environmental reservoir at the beginning of the interval.

For the sake of readability, the index *t* is dropped from now on. Hence, S, I, I_k, E, E_k and *N* refer to the respective number at the beginning of the interval.

A case was defined as a foot that was susceptible (M0) at the beginning of the interval and infected at the end of the interval. We assume that the number of cases

within an interval *Δt* follows a binomial distribution with binomial total *S, i.e.,* the number of susceptible feet at start of the interval, and the probability to become infected during the interval (Velthuis et al., 2003). From Equation 3.3, the probability (*P*) for a single foot (S = 1) to get infected within period *Δt* is the probability of a non-zero outcome from a Poisson distribution,

$$P = 1 - e^{-\sum_k \varphi_k \frac{E_k}{N} \Delta t}. \qquad \text{(Equation 3.4)}$$

Because the probability to get infected follows from the zero-term of the Poisson distribution (Equation 3.4), a generalized linear mixed model (GLMM) with a complementary log-log link function was used to connect the explanatory variables to the observed variable (Velthuis et al., 2003; Anche et al., 2015; de Rueda et al., 2015),

$$\text{cloglog}(P) = \log(-\log(1-P)) = \log\left(\sum_k \varphi_k \frac{E_k}{E}\right) + \log\left(\frac{E}{N}\Delta t\right), \qquad \text{(Equation 3.5)}$$

where $\frac{E_k}{E}$ is the relative contribution of class *k* to the environmental reservoir, and $\log\left(\frac{E}{N}\Delta t\right)$ is an offset, *i.e.*, an "explanatory variable" with coefficient equal to 1.

To make the model linear in $\log(\varphi_k)$, the arithmetic mean, $\sum_k \varphi_k \frac{E_k}{E}$, was approximated by the corresponding geometric mean, $\prod_k \varphi_k^{\frac{E_k}{E}}$ (Anche et al., 2015),

$$\text{cloglog}(P) \approx \sum_k \frac{E_k}{E} \log(\varphi_k) + \log\left(\frac{E}{N}\Delta t\right). \qquad \text{(Equation 3.6)}$$

Infectivity of class *k* was expressed relative to a reference class, here M1, so that $\varphi'_k = \varphi_1 \varphi'_k$ with $\varphi'_1 = 1$. Therefore, the first term on the right hand side of the model (Equation 3.6) becomes,

$$\sum_k \frac{E_k}{E} \log(\varphi_1 \varphi'_k) = \sum_k \frac{E_k}{E} \left(\log(\varphi_1) + \log(\varphi'_k)\right) =$$

$$\sum_k \frac{E_k}{E} \log(\varphi_1) + \sum_k \frac{E_k}{E} \log(\varphi'_k) =$$

$$\log(\varphi_1) + \sum_k \frac{E_k}{E} \log(\varphi'_k), \qquad \text{(Equation 3.7)}$$

with $\varphi_1' = 1$, so that the summation in the right hand side may exclude the case where $k = 1$.

Now, we have a linear model with an intercept representing $\log(\varphi_1)$, and covariates for the relative contribution of class *k,* where $k \neq 1$, to the environmental reservoir,

$$\text{cloglog}(P) = a_1 + a_2\frac{E_2}{E} + a_3\frac{E_3}{E} + a_4\frac{E_4}{E} + a_{4.1}\frac{E_{4.1}}{E} + \log\left(\frac{E}{N}\Delta t\right), \qquad \text{(Equation 3.8)}$$

where, a_1 is the intercept, which is the log of the infectivity effect of class M1. The covariate $\frac{E_2}{E}$ is the relative contribution of class M2 to the environmental reservoir, *etc.* The a_2, a_3, a_4, and $a_{4.1}$ are the regression coefficients belonging to these covariates, and are the parameters of interest. To account for footbath use in-between observations we also included a dummy variable (0/1) as a fixed effect in the model. Furthermore, a random farm effect was added to the model to account for stochastic differences between farms.

Fitting the GLMM requires the values of E_k, which depend on the survival rate λ. The λ was assumed to be the same for each class. The λ was estimated by evaluating different values and determining the best fit (highest maximum likelihood) for the model (Equation 3.8). A large λ means that the pathogens remain in the environment for a long time. So, even when a foot has recovered, the pathogens from that foot may still reside in the environment. To determine the infectious pressure from the environment, we needed the number of infected feet of class *k* in the weeks prior to the first observation ($E_{k,t=0}$). Since we did not observe the number of infected feet per class in the weeks before the first observation, they had to be estimated. Therefore, we plotted for each farm the number of infected feet of class *k* over the entire observation period, and obtained the corresponding regression equation, $I_k = a + bt$. Where, *a* is the intercept with the y-axis, *b* is the slope and *t* is the time. The intercept was used as the average number of infectious feet of class *k* before observations started ($I_{k,t=0}$). The contribution of class *k* on a farm to the environmental reservoir before the first observation was then estimated as $E_{k,t=0} = \frac{\lambda}{1-\lambda} I_{k,t=0}$.

Because after the class-at-infection a foot may reside in multiple infectious classes, the regression coefficients need to be weighted according to the sojourn time in the different classes. By doing so, for each class-at-infection *i* a transmission rate parameter β_i is estimated for the entire period that a foot is infectious,

$$\hat{\beta}_i = e^{x_1\hat{a}_1+\sum_{k=2}^{4.1}\left(\frac{x_{k,i}}{x_i}(\hat{a}_1+\hat{a}_k)\right)} = e^{\hat{a}_1+\sum_{k=2}^{4.1}\left(\frac{x_{k,i}}{x_i}\hat{a}_k\right)}, \quad \text{(Equation 3.9)}$$

here, $\frac{x_{k,i}}{x_i}$ is the fraction of time spent in class *k* given class-at-infection *i*. The $\hat{a}_1$ and $\hat{a}_k$ are the regression coefficients from the GLM; $\hat{a}_k$ does not include $\hat{a}_1$, because that estimate (of reference class M1) is included in the intercept. Note that, therefore, in Equation 3.9 the sum in the exponent is over *k* = 2 to 4.1.

To determine if there were significant differences between weighted regression coefficients, the contrasts of all pairs were tested against 0. If the 95% confidence interval (CI) of the contrast did not include 0, the weighted estimated regression coefficients were significantly different.

3.2.2.4 Environmental reservoir

Pathogens survived in de environment with a rate λ each period. So the total contribution of a foot to the environmental reservoir, expressed relative to the contribution during a single infectious time period, is $1 + \lambda + \lambda^2 + \lambda^3 \ldots = (1-\lambda)^{-1}$.

3.2.3 Implementation

Data were analysed in Excel and R version 3.4.0. (R Core Team, 2017). The R-package vcd (Meyer et al., 2008) was used to calculate Kappa's coefficient. R-package lme4 (Bates et al., 2014) was used to estimate the transmission rate parameters. This package allows for inclusion of random group effects in the generalized linear mixed model to account for variation between farms. R-package multcomp (Hothorn et al., 2008) was used for linear hypothesis testing of the transmission rate parameter estimates and for the computation of their 95% confidence interval. For the sojourn times and R_0 we generated 95% confidence intervals through bootstrapping of the dataset, the dataset was analysed repeatedly with each time a different farm excluded.

3.2.4 Classes grouped together

In the first analysis, all classes-at-infection were analysed separately. In the second analysis, classes were combined into a simplified classification system because not all classes can be observed with certainty in the milking parlour. In the milking parlour M3 is often confused with M4 (Relun et al., 2011) and M2 with M4.1 (Solano et al., 2017a), so we merged these classes. Furthermore, class M1 was merged with the M2 and M4.1 class as well because of the resemblance with M2 lesions (Relun et al., 2011). Three groups remained: no lesions (M0), active lesions (M1, M2 and

M4.1), and inactive lesions (M3 and M4). In the third analysis, the classes-at-infection of which the transmission rate parameters of the first analysis were not significantly different were grouped.

3.3 Results

Cows on farm A, B, C, E, F and K were scored by observer 1, and cows on farm D, G, H, I, J and L were scored by observer 2. Cohen's kappa coefficient was 0.75 (substantial agreement between observers, 95% CI [0.66, 0.84], n = 204 claws) immediately before data collection; and 0.85 (almost perfect agreement, 95% CI [0.78, 0.93], n = 164 claws) and 0.76 (substantial agreement, 95% CI [0.61, 0.90], n = 52 claws), during data collection.

Table 3.1 gives an overview of the characteristics of the farms enrolled in the study. In total 29,792 transitions were observed. The average time between observations (Δt) was 14.08 days. Farm specific prevalences ranged from 49.7% to 78.0% on cow level and from 40.2% to 69.7% on foot level (Table 3.1). The average prevalence on foot level for each M-class is shown for each farm in Figure 3.2.

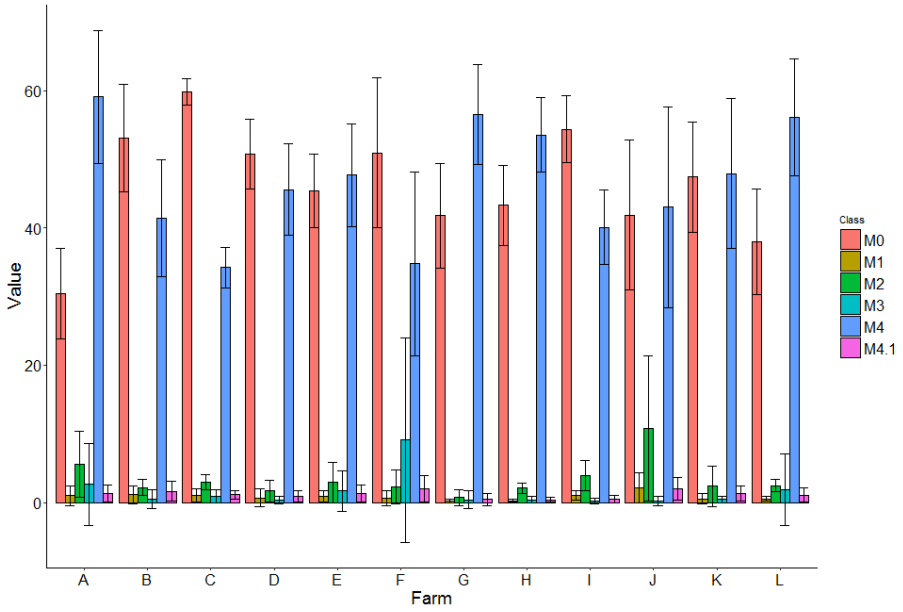


Figure 3.2. Average prevalence and standard deviation of each M-class on foot level per farm.

Table 3.2. The observed frequencies with which the classes-at-infection were entered per farm and for all farms together.

Class-at-infection	Observed frequency θ_i per farm												θ_i for all farms together
	A	B	C	D	E	F	G	H	I	J	K	L	
M1	0.011	0.0107	0.0062	0.0071	0.0316	0.0118	0	0.0073	0.0248	0.0282	0.0068	0.0038	0.0115
M2	0.017	0.0107	0.0309	0	0.0211	0.0178	0.0097	0.0073	0.0248	0.0634	0	0	0.0162
M3	0.005	0	0.0247	0	0.0316	0.2071	0	0	0	0.007	0.0068	0.0266	0.0248
M4	0.964	0.9679	0.9383	0.9929	0.9053	0.7337	0.9903	0.9853	0.9505	0.8803	0.973	0.9544	0.9389
M4.1	0	0.0107	0	0	0.0105	0.0296	0	0	0	0.0211	0.0135	0.0152	0.0086

3.3.1 Class-at-infection (θ_i) and sojourn time (xi)

Table 3.2 summarizes the observed frequencies with which the classes-at-infection were entered per farm and for all farms together. The frequency over all farms was used in the R_0 calculations. Class M4 was the most observed class-at-infection with a frequency of 93.89% for all farms together.

Table 3.3 shows the transitions between classes at consecutive observations. On average 77.11% of the feet that were scored as M0 were also scored as M0 the next observation. Similarly, on average 72.89% of the feet that were scored as M4 were also scored as M4 the next observation. After that, transitions from classes M1, M2, M3, and M4.1 to class M4 were most common (42.28% to 63.19%).

Table 3.3. Observed transitions between Digital Dermatitis classes (M0-M4.1) at consecutive observations. Transitions were summed over all farms and periods. The average observation interval was 14.08 days.

From[1] / To[1]	**NA[2]**	**M0**	**M1[3]**	**M2[3]**	**M3[3]**	**M4[3]**	**M4.1[3]**	***n total***
NA[2]	0.8183	0.0391	0.0521	0.0312	0.0521	0.0373	0.0448	*5,788*
M0	0.0925	0.7711	0.1458	0.0217	0.1458	0.1599	0.0522	*10,969*
M1[3]	0.0015	0.0022	0.0156	0.0217	0.0104	0.0091	0.0299	*166*
M2[3]	0.0067	0.0031	0.1719	0.4119	0.0556	0.0250	0.1381	*747*
M3[3]	0.0022	0.0047	0.0573	0.0569	0.0486	0.0239	0.0597	*418*
M4[3]	0.0762	0.1782	0.5156	0.4228	0.6319	0.7289	0.6269	*11,428*
M4.1[3]	0.0025	0.0016	0.0417	0.0339	0.0556	0.0160	0.0485	*276*
n total	*5,956*	*11,035*	*192*	*738*	*288*	*11,315*	*268*	*29,792*

[1] Fraction transitioning from M*l* to M*k*, columns add up to 1.0.

[2] A transition from or to NA indicates that the animal was not observed the previous or next time.

[3] Within the dashed lines are the elements of matrix **Σ** that describes the transitions between infected classes.

The transitions between infected classes over a period Δt, compiled in matrix **Σ**, are within the dashed lines of Table 3.3. With this matrix, the sojourn times were calculated. Table 3.4 shows the estimated sojourn time in class *k* given the class-at-infection *i* and the 95% confidence intervals for the estimates. Estimate $x_{k,i}$ represents the number of periods of Δt = 14.08 days a foot resides in class *k* given

the class-at-infection *i*. For example, infections that were first observed as class M3 reside on average 3.678 observation periods in class M4.

Table 3.4. Estimated sojourn time in class *k* given the class-at-infection *i* in periods of Δt = 14.08 days, the 95% confidence interval of the estimates is between parentheses.

Sojourn time in *k*	**Class-at-infection *i***				
	M1	**M2**	**M3**	**M4**	**M4.1**
M1	1.067 (1.057-1.077)	0.089 (0.076-0.101)	0.058 (0.049-0.067)	0.054 (0.046-0.062)	0.086 (0.072-0.099)
M2	0.515 (0.433-0.598)	1.962 (1.826-2.098)	0.318 (0.277-0.359)	0.255 (0.219-0.291)	0.489 (0.436-0.542)
M3	0.194 (0.140-0.248)	0.236 (0.182-0.291)	1.175 (1.126-1.224)	0.144 (0.102-0.186)	0.209 (0.153-0.266)
M4	3.601 (3.386-3.815)	4.142 (3.889-4.395)	3.678 (3.414-3.942)	4.755 (4.527-4.984)	4.078 (3.849-4.306)
M4.1	0.137 (0.123-0.151)	0.157 (0.141-0.173)	0.144 (0.114-0.175)	0.100 (0.087-0.113)	1.153 (1.136-1.170)

Figure 3.3 gives the average sojourn time in days in the infectious classes given the class-at-infection. For example, feet with class-at-infection M1 will be on average 1.067 * 14.08 = 15.0 days in class M1. Note that moves between classes do not have to occur in any particular order.

3.3.2 Transmission rate parameters

A survival rate of $\lambda = 0.9$ gave the best fit for the GLMM (highest maximum likelihood). The estimated regression coefficient for footbath use was -0.093 and not significant (P = 0.11) and therefore dropped from the model.

The estimated regression coefficients from the GLMM are in Table 3.5. With Table 3.5 and the matrix of sojourn times **X,** the weighted estimated regression coefficients were calculated (the exponent in Equation 3.9), and summarize in Table 3.6.

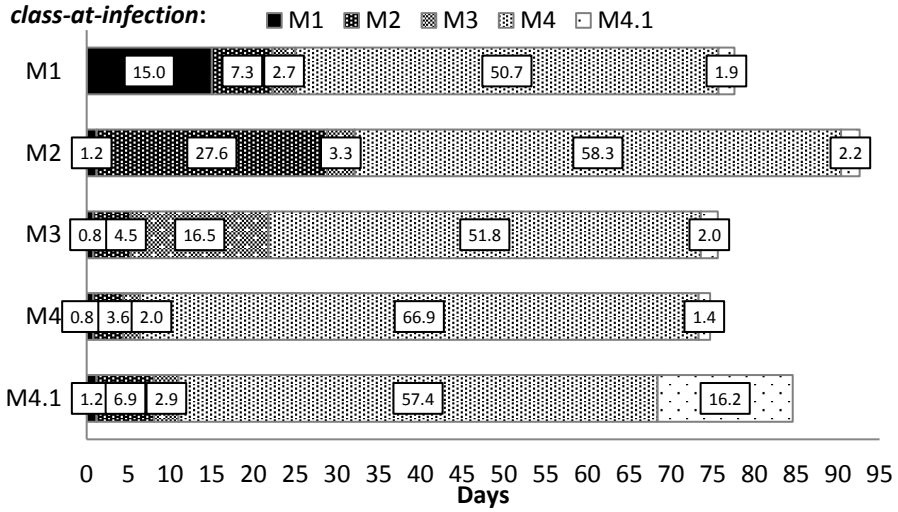


Figure 3.3. Average sojourn time in days of an average infected foot in the different infected classes, given the class-at-infection. Moves between classes do not have to occur in any particular order.

Table 3.5. Estimated regression coefficients.

Regression coefficient	Estimate	Standard error
a_1	-8.803	4.187
a_2	5.027	5.223
a_3	4.751	4.368
a_4	5.510	4.184
$a_{4.1}$	17.529	6.208

The weighted estimated regression coefficient of class-at-infection M4.1 is significantly higher than the other weighted estimated regression coefficients. This is due to a combination of the high estimated regression coefficient of class M4.1 (Table 3.5), and the relatively long sojourn time in class M4.1 of feet with class-at-infection M4.1 (Figure 3.3).

Transmission rate parameters β_i follow by taking the exponent of the weighted regression coefficients b_i (Table 3.6).

Table 3.6. Weighted estimated regression coefficients and estimated transmission rate parameters.

Weighted regression coefficient	Estimate	Standard error	95% confidence interval		Transmission rate parameter β
			lower	upper	
b_1	-4.142[1]	0.708	-5.950	-2.333	0.016
b_2	-3.255[1]	0.385	-4.240	-2.270	0.039
b_3	-3.230[1]	0.230	-3.817	-2.643	0.040
b_4	-3.172[1]	0.071	-3.350	-2.990	0.042
$b_{4.1}$	-1.132[2]	0.619	-2.713	0.449	0.322

[1/2] A different superscript indicates a significant difference in weighted regression coefficient ($P < 0.05$).

3.3.3 Basic reproduction ratio R_0

With the above presented elements R_0 was calculated. Table 3.7 shows the contribution of each class-at-infection to R_0. Class-at-infection M4 contributes over 88%. Furthermore, with a contribution of 2.089, it is the only class-at-infection that contributes >1.0 to R_0. The other classes-at-infection contribute together 2.360-2.089 = 0.290 to R_0. Thus, R_0 can be brought below one, only when the transmission caused by class-at-infection M4 is reduced.

Table 3.7. Contribution of each class-at-infection to the basic reproduction ratio R_0.

Class-at-infection	Contribution to R_0 (95% confidence interval)	Relative contribution (%)
M1	0.010 (-0.022 – 0.042)	0.43
M2	0.041 (0.0164 – 0.066)	1.75
M3	0.053 (0.031 – 0.075)	2.24
M4	2.089 (1.927 – 2.251)	88.53
M4.1	0.167 (-0.093 – 0.426)	7.06
R_0	2.360 (2.060 – 2.661)	

3.3.4 Classes grouped together

In the second analysis, classes were combined into a simplified classification system: no lesions (M0), an active group (M1, M2 and M4.1), and an inactive group (M3 and M4).

Matrix Σ_{merged} describes the transitions within and between the active and inactive lesion classes,

$$\Sigma_{merged} = \begin{bmatrix} 0.3731 & 0.0519 \\ 0.5409 & 0.7509 \end{bmatrix}.$$

With this matrix the sojourn times were calculated (Table 3.8).

Table 3.8. Estimated sojourn time in the active (M1, M2, and M4.1) or inactive (M3 and M4) class given the class-at-infection in periods of Δt = 14.08 days, the 95% confidence interval of the estimates is between parentheses.

Sojourn time	Class-at-infection	
	Active	Inactive
Active	1.945	0.405
	(1.833-2.057)	(0.355-0.455)
Inactive	4.233	4.895
	(3.991-4.455)	(4.666-5.123)

The estimated regression coefficients from the GLMM for the active and inactive group are in Table 3.9. With the estimated regression coefficients and the sojourn times**,** the weighted estimated regression coefficients were calculated. There was no significant difference between the weighted estimated regression coefficients of the active classes and the inactive classes (Table 3.10).

Table 3.9. Estimated regression coefficients when classes M1, M2, and M4.1 (active lesions) and M3 and M4 (inactive lesions) are merged.

Regression coefficient	Estimate	Standard error
a_{active}	-2.960	0.841
$a_{inactive}$	-0.196	0.904

Table 3.10. Weighted estimated regression coefficients and estimated transmission rate parameters when classes M1, M2, and M4.1 (active lesions) and M3 and M4 (inactive lesions) are merged.

Weighted regression coefficient	Estimate	Standard error	95% confidence interval		Transmission rate parameter β
			lower	Upper	
b_{active}	-3.0943	0.234	-3.614	-2.575	0.045
$b_{inactive}$	-3.1411	0.083	-3.325	-2.957	0.043

With the above presented elements $R_{0merged}$ was calculated, $R_{0merged}$ = 2.310. Table 3.11 shows the contribution of the active and inactive classes to R_0, the inactive classes-at-infection contribute over 95%.

Table 3.11. Contribution of each class-at-infection to the basic reproduction ratio R_0.

Class-at-infection	Contribution to R_0 (95% confidence interval)	Relative contribution (%)
Active[1]	0.101 (0.080 – 0.123)	4.39
Inactive[2]	2.208 (2.090 – 2.326)	95.61
R_0	2.310 (2.192- 2.427)	

[1] Classes M1, M2, and M4.1 were merged

[2] Classes M3 and M4 were merged

In the third analysis, classes-at-infection M1, M2, M3, and M4 were grouped together because their weighted estimated regression coefficients were not significantly different (Table 3.6). Their estimated regression coefficients from the GLMM are in Table 3.12.

Table 3.12. Estimated regression coefficients when class-at-infection M1, M2, M3, and M4 are grouped together.

Regression coefficient	Estimate	Standard error
a_{group}[1]	-3.385	0.100
$a_{4.1}$	9.777	3.344

[1] Classes M1, M2, M3, and M4 were grouped together.

Even though class M1, M2, M3, and M4 now have the same estimated regression coefficient, they still had different sojourn times in the infectious classes (matrix $\mathbf{X}$). Table 3.13 summarizes the contribution of each class-at-infection to R_0. Again, R_0 is almost completely determined by class-at-infection M4 that contributes over 89%.

Table 3.13. Contribution of each class-at-infection to the basic reproduction ratio R_0. Classes M1, M2, M3 and M4 were grouped and have the same estimated regression coefficient.

Class-at-infection	Contribution to R_0	Relative contribution (%)
M1	0.027	1.20
M2	0.046	2.01
M3	0.059	2.28
M4	2.028	89.19
M4.1	0.114	5.02
R_0	2.274	

3.4 Discussion

We determined the contribution of different classes-at-infection to R_0. Over 93% of the feet had class-at-infection M4, and feet with another classes-at-infection rapidly transitioned to class M4 as well. As a consequence, about 70% of the infected time was spent in class M4. Transmission rate parameters of class-at-infection M1, M2, M3, and M4 were not significantly different, but differed from the transmission rate parameter of class-at-infection M4.1. However, because over 93% of the infections was first observed as class-at-infection M4, and because the sojourn time in class M4 is relatively long, this class almost completely determined R_0.

Döpfer et al. (2012b) reckoned that only M2 and M4 lesions were infectious, and estimated that M2 lesions were on average two times as infectious as M4 lesions. However, we found in our dataset that class M4.1 had the highest estimated regression coefficient. The contribution of class M4.1 to the transmission rate parameters for classes-at-infection M1, M2, M3, and M4 was, however, minor because the sojourn time in class M4.1 was short.

We found that M4 lesions were persistent, 70% of the infected time was spent in this class. This persistence could be due to the encysted *Treponema spp.* deep in the epidermis of M4 lesions (Döpfer et al., 2012a). *Treponema spp.* are able to penetrate the skin after the epidermis is eroded and the keratin layer is degenerated. The loss of the keratin layer is the first change that can be observed after infection. The loss is probably due to a keratolytic toxin produced by long spiral pathogenic organisms that are observed in the early stage of the infection (Blowey et al., 1994). The loss of the keratin layer is followed by hyperplasia and hypertrophy of the epithelium to over 100 cells/mm (normal epithelium is 5 to 70 cells thick) (Blowey et al., 1994; Döpfer et al., 1997). Next, the central and superficial layers of the epidermis are

eroded, possibly by proteolytic enzymes of *Dichelobacter nodusus*, and infiltrated with spirochetes (Blowey et al., 1994; Rasmussen et al., 2012). At this moment the spirochetes, *Treponema spp.* of phylotype PT1, PT3, PT6, PT8, and *T. brennaborense*, reside in the deep parts of the epidermis (Klitgaard et al., 2008; Rasmussen et al., 2012). These *Treponema spp.* can undergo a morphological change from a spiral to an encysted form (Döpfer et al., 2012a). The change to an encysted form can be a protective mechanism within the host but it is also an important stage in transmission because the cyst increases the bacterium's survival in the environment (Al-Qudah et al., 1983). The optimal conditions for *Treponema spp.* to change to their encysted form are at 37°C in an anaerobic environment (Al-Qudah et al., 1983). It is possible that these conditions are met in the deep parts of the epidermis. Both the location of the *Treponema spp.* deep in the epidermis and their encysted form makes the lesions hard to treat, hence, M4 lesions are persistent.

Krull et al. (2014) observed that early-stage lesions did not quite fit in the M-classification system. They, therefore, developed the Iowa DD scoring system based on morphological appearance and bacterial presence. The Iowa DD scoring system distinguishes normal skin (stage 0), lesion onset (stage 1), developing lesions (stage 2), classical ulceration (stage 3), and chronic lesions (stage 4) (Krull et al., 2014; Krull, 2015). Stage 1 and stage 2 are subdivided into type A and type B; type A lesions have an ulcerated appearance and are located in the interdigital cleft, and type B lesions have a thickened appearance and are located diffusely spread across the heel. Krull et al. (2016) observed that all infections started with an early lesion, *i.e.*, stage 1 or stage 2 (cows were scored every three to four weeks). We, however, observed that the majority of feet that were scored as M0 at *t* and got infected, were scored as M4 at *t*+1 (note that cows were scored every two weeks). Hyperkeratotic lesions of class M4 show similarities with the stage 1A and stage 1B lesion descriptions and pictures in Krull et al. (2014, 2015). We, therefore, suspect that lesions that were scored as M4 in our study would have been classified as stage 1A or stage 1B lesions with the Iowa DD scoring system. This would imply that M4 lesions are not (only) chronic lesions but can be early lesions as well.

To simplify analysis classes M1 to M4.1 are often divided into two categories, a category with active lesions and a category with inactive lesions. The active lesions category consists of classes that describe circumscribed, red-greyish, moist, painful, and prone to bleed lesions (M1, M2, and M4.1) (Speijers et al., 2010; Berry et al., 2012; Zinicola et al., 2015),the inactive lesions category consists of classes M3 and M4. Studies on DD tend to focus on active lesions because these lesions can be painful for the animal, and thus cause lameness. Because cows that show signs of lameness can be easily identified, it is plausible that active lesions are more often

treated than inactive lesions (Ettema et al., 2007). We showed, however, that inactive lesions contribute over 95% to R_0. So, to stop transmission, the inactive lesions that contribute most to R_0 should be prevented.

We scored twelve herds, eleven times each, with a two-week interval between scorings. This resulted in one of the biggest datasets on DD transmission, with a relatively high number of scorings at relatively short intervals compared to other studies on DD transmission. Datasets used in other studies contained data from five herds, scored eight times, every three weeks (Döpfer et al., 2012b), thirty cows, scored eleven times, every four weeks (Berry et al., 2012), three herds, scored twelve times, every week, (Nielsen et al., 2012), or 138 cows, scored four times, every week (Holzhauer et al., 2008). Although, the study of Relun et al. (2012) has more data (52 farms, scored seven times), the interval between scoring was longer, with scorings every four weeks.

We chose a two-week scoring interval to maximize the number of farms in the study, while minimizing the risk of missing moves between classes. Our analysis showed, however, that classes M1, M2, M3, and M4.1 lasted on average less than two weeks (Figure 3.3). So, weekly scoring would be preferred to avoid missing transitions between classes (Tremblay et al., 2016). However, on all farms there was a (dynamic) endemic equilibrium prevalence. Missing transitions on endemically affected farms has less consequences compared to missing transitions on farms that undergo an epidemic outbreak (Tremblay et al., 2016). On an endemically affected farm, the distribution of the infected classes at a certain point in time reflects, on average, the distribution of the infected classes over the entire infectious period, *i.e.*, if on average a foot resides in class M1 two days out of the 80 days a foot is infectious (2.5% of the time), then on average 2.5% of the infected feet will be scored as class M1 at a random moment in time because of the endemic equilibrium. Still, a short(er) scoring interval provides more information as more moves between classes within an individual can be observed.

The risk on DD is affected by factors like lactation number, days in milk, the number of rear feet that are infected, days since treatment, etc. (Argáez-Rodríguez et al., 1997; Holzhauer et al., 2006; Relun et al., 2013; Krull et al., 2016). Heifers have the highest risk of getting infected, and after the second lactation this risk declines with each parity (Rodríguez-Lainz et al., 1996; Read and Walker, 1998; Somers et al., 2005). The risk of getting infected is also higher the first months after calving (Argáez-Rodríguez et al., 1997; Holzhauer et al., 2006). Furthermore, cows that are infected on one rear foot have a higher risk of developing a lesion on the other foot as well (Relun et al., 2013; Krull et al., 2016). Including these factors in the model (Equation 3.8) would be methodically very challenging because all analyses were done on a

herd level. However, we were able to include a variable for footbath use in the period between observations (exact data on the days since treatment or footbath were not complete). The estimate for this factor had no significant effect and was hence dropped from the model.

We scored the cows in the milking parlour without lifting their feet. With this scoring method we cannot observe lesions located in the interdigital space (Relun et al., 2011). Previous studies determined the sensitivity (SE) and specificity (SP) of parlour scoring by comparing this method to scoring in the trimming chute. When lesions are located on the heel bulb, the SE ranged from 0.79 to 0.93 and the SP ranged from 0.67 to 0.92 for absence (M0) versus presence (M1, M2, M3, or M4) of DD (Relun et al., 2011; Solano et al., 2017a). However, if lesions were located elsewhere on the foot, the SE was below 0.64 (Solano et al., 2017a). In the second analysis we merged lesion classes that are often confused. We choose to merge classes into an active and an inactive group based on previous studies (Speijers et al., 2010; Berry et al., 2012; Zinicola et al., 2015). Merging classes is not trivial, another option is merging class M1 with M0, and class M3 with M4 and M4.1. The analysis of these groups also showed that the group that contained class M4 determined most (97.3%) of R_0. We choose, however, to merge different classes because we did not want to assume *a priori* that lesion stage M1 is not infectious. To stop the transmission of DD, R_0 should be lowered to a value below one. We found an R_0 of 2.36, 95% CI [2.06, 2.66], this value is within the range of R_0 values (from 0.463 to 3.273) estimated by Döpfer et al. (2012b). The R_0 was almost entirely determined by class M4 lesions (88.53%), so an R_0 below one can only be achieved by prevention of M4 lesions. An easy preventive measure that is often used on herd level is a footbath (Laven and Proven, 2000). Footbaths effectively reduce the prevalence of active DD lesions on farms with a high prevalence of active DD lesions (Solano et al., 2017b). However, a footbath does not effectively reduce the prevalence of other lesion classes, like class M4 (Solano et al., 2017b). So, footbaths do control clinical disease by reducing the severity of active lesions, but they do not prevent the occurrence of new lesions or the recurrence of existing lesions (Speijers et al., 2010; Teixeira et al., 2010). It is therefore unlikely that by only using footbaths, DD will be eliminated from the herd (Laven and Proven, 2000; Speijers et al., 2010). Topical treatment of class M4 lesions with antibiotics is also difficult because the *Treponema spp.* reside deep in the epidermis, a place that might not be reached (Mumba et al., 1999). Therefore, further research on effective prevention of class M4 is crucial (Döpfer et al., 2012a). In conclusion, active lesions should be prevented for animal welfare reasons, while class M4 lesions should be prevented to lower R_0 to a value below one for the disease to die out.

3.5 Conclusions

We showed that, regardless of the class-at-infection, class M4 determines almost the entire basic reproduction ratio because of the high occurrence and a long duration of this lesion type, not because it has a higher infectivity. The endemic prevalence is a function of the transmission measured by the parameter R_0. So, to lower the endemic equilibrium prevalence, focus should be on prevention of class M4.

3.6 Acknowledgements

We thank CRV for their help with finding the farmers, and the farmers for their participation. This work is part of the research programme “Meer met minder” with project number 847.13.004, which is financed by the Netherlands Organisation for Scientific Research (NWO).

3.7 References

Al-Qudah, A., Mostratos, A., Quesnel, L., 1983. A proposed life cycle for the Reiter treponeme. Journal of applied bacteriology 55, 417-428.

Anche, M.T., Bijma, P., De Jong, M.C., 2015. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47, 1-15.

Argáez-Rodríguez, F.d.J., Hird, D.W., de Anda, J.H., Read, D.H., Rodríguez-Lainz, A., 1997. Papillomatous digital dermatitis on a commercial dairy farm in Mexicali, Mexico: Incidence and effect on reproduction and milk production. Preventive Veterinary Medicine 32, 275-286.

Bassett, H., Monaghan, M., Lenhan, P., Doherty, M., Carter, M., 1990. Bovine digital dermatitis. Veterinary Record 126, 164-165.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Berry, S.L., Read, D.H., Famula, T.R., Mongini, A., Döpfer, D., 2012. Long-term observations on the dynamics of bovine digital dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. The Veterinary Journal 193, 654-658.

Blowey, R., Done, S., Cooley, W., 1994. Observations on the pathogenesis of digital dermatitis in cattle. Veterinary Record 135, 117-115.

Cheli, R., Mortellaro, C., 1974. La dermatite digitale del bovino, 8th international conference on diseases of cattle; Piacenza, Milan, Italy, pp. 208-213.

Clegg, S., Mansfield, K., Newbrook, K., Sullivan, L., Blowey, R., Carter, S., Evans, N., 2015. Isolation of digital dermatitis treponemes from hoof lesions in wild North American elk (Cervus elaphus) in Washington State, USA. Journal of clinical microbiology 53, 88-94.

Collighan, R.J., Woodward, M.J., 1997. Spirochaetes and other bacterial species associated with bovine digital dermatitis. FEMS Microbiology Letters 156, 37-41.

de Rueda, C.B., de Jong, M.C., Eblé, P.L., Dekker, A., 2015. Quantification of transmission of foot-and-mouth disease virus caused by an environment contaminated with secretions and excretions from infected calves. Veterinary research 46, 1.

Demirkan, I., Walker, R., Murray, R., Blowey, R., Carter, S., 1999. Serological evidence of spirochaetal infections associated with digital dermatitis in dairy cattle. The Veterinary Journal 157, 69-77.

Diekmann, O., Heesterbeek, J., Roberts, M., 2009. The construction of next-generation matrices for compartmental epidemic models. Journal of the Royal Society Interface, rsif20090386.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. John Wiley & Sons.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A., 1990. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of mathematical biology 28, 365-382.

Döpfer, D., 2009. Digital Dermatitis: The dynamics of digital dermatitis in dairy cattle and the manageable state of disease, CanWest Veterinary Conference, Banff AB, Canada.

Döpfer, D., Anklam, K., Mikheil, D., Ladell, P., 2012a. Growth curves and morphology of three Treponema subtypes isolated from digital dermatitis in cattle. The Veterinary Journal 193, 685-693.

Döpfer, D., Holzhauer, M., Boven, M.v., 2012b. The dynamics of digital dermatitis in populations of dairy cattle: Model-based estimates of transition rates and implications for control. The Veterinary Journal 193, 648-653.

Döpfer, D., ter Huurne, A., Cornelisse, J., Van Asten, A., Koopmans, A., Meijer, F., Schukken, Y., Szakall, I., Klee, W., Bosma, R., 1997. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and Campylobacter faecalis. Veterinary Record 140, 620-623.

Ettema, J.F., Capion, N., Hill, A.E., 2007. The association of hoof lesions at claw trimming with test-day milk yield in Danish Holsteins. Preventive veterinary medicine 79, 224-243.

Holzhauer, M., Bartels, C.J., Döpfer, D., van Schaik, G., 2008. Clinical course of digital dermatitis lesions in an endemically infected herd without preventive herd strategies. The Veterinary Journal 177, 222-230.

Holzhauer, M., Hardenberg, C., Bartels, C., Frankena, K., 2006. Herd-and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. Journal of dairy science 89, 580-588.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biometrical journal 50, 346-363.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. The Royal Society, pp. 700-721.

Klitgaard, K., Boye, M., Capion, N., Jensen, T.K., 2008. Evidence of multiple Treponema phylotypes involved in bovine digital dermatitis as shown by 16S rRNA gene analysis and fluorescence in situ hybridization. Journal of clinical microbiology 46, 3012-3020.

Krull, A.C., 2015. Bovine digital dermatitis: natural lesion development and experimental induction.

Krull, A.C., Shearer, J.K., Gorden, P.J., Cooper, V.L., Phillips, G.J., Plummer, P.J., 2014. Deep sequencing analysis reveals temporal microbiota changes associated with development of bovine digital dermatitis. Infection and immunity 82, 3359-3373.

Krull, A.C., Shearer, J.K., Gorden, P.J., Scott, H.M., Plummer, P.J., 2016. Digital dermatitis: Natural lesion progression and regression in Holstein dairy cattle over 3 years. Journal of dairy science 99, 3718-3731.

Laven, R., 2001. Control of digital dermatitis in cattle. In practice - London - British Veterinary Association 23, 336-341.

Laven, R., Proven, M., 2000. Use of an antibiotic footbath in the treatment of bovine digital dermatitis. Veterinary Record 147, 503-506.

Meyer, D., Zeileis, A., Hornik, K., 2008. vcd: Visualizing categorical data. R package version.

Mumba, T., Kruitwagen, C., Dreher, M., Gaastra, W., van der Zeijst, B., 1999. Detection of spirochetes by polymerase chain reaction and its relation to the course of digital dermatitis after local antibiotic treatment in dairy cattle. Journal of Veterinary Medicine, Series B 46, 117-126.

Nielsen, B.H., Thomsen, P.T., Green, L.E., Kaler, J., 2012. A study of the dynamics of digital dermatitis in 742 lactating dairy cows. Preventive Veterinary Medicine 104, 44-52.

R Core Team, 2017. R: A Language and Environment for Statistical Computing, 3.4.0 ed. R Foundation for Statistical Computing, Vienna, Austria.

Rasmussen, M., Capion, N., Klitgaard, K., Rogdo, T., Fjeldaas, T., Boye, M., Jensen, T.K., 2012. Bovine digital dermatitis: possible pathogenic consortium consisting of Dichelobacter nodosus and multiple Treponema species. Veterinary microbiology 160, 151-161.

Read, D.H., Walker, R.L., 1998. Papillomatous digital dermatitis (footwarts) in California dairy cattle: clinical and gross pathologic findings. Journal of Veterinary Diagnostic Investigation 10, 67-76.

Relun, A., Guatteo, R., Roussel, P., Bareille, N., 2011. A simple method to score digital dermatitis in dairy cows in the milking parlor. J Dairy Sci 94, 5424-5434.

Relun, A., Lehebel, A., Bareille, N., Guatteo, R., 2012. Effectiveness of different regimens of a collective topical treatment using a solution of copper and zinc chelates in the cure of digital dermatitis in dairy farms under field conditions. Journal of dairy science 95, 3722-3735.

Relun, A., Lehebel, A., Bruggink, M., Bareille, N., Guatteo, R., 2013. Estimation of the relative impact of treatment and herd management practices on prevention of digital dermatitis in French dairy herds. Preventive veterinary medicine 110, 558-562.

Roberts, M., Heesterbeek, H., 1993. Bluff your way in epidemic models. Trends in microbiology 1, 343-348.

Rodríguez-Lainz, A., Hird, D.W., Carpenter, T.E., Read, D.H., 1996. Case-control study of papillomatous digital dermatitis in Southern California dairy farms. Preventive Veterinary Medicine 28, 117-131.

Sogstad, Å., Fjeldaas, T., Østerås, O., Forshell, K.P., 2005. Prevalence of claw lesions in Norwegian dairy cattle housed in tie stalls and free stalls. Preventive veterinary medicine 70, 191-209.

Solano, L., Barkema, H., Jacobs, C., Orsel, K., 2017a. Validation of the M-stage scoring system for digital dermatitis on dairy cows in the milking parlor. Journal of dairy science 100, 1592-1603.

Solano, L., Barkema, H., Pickel, C., Orsel, K., 2017b. Effectiveness of a standardized footbath protocol for prevention of digital dermatitis. Journal of dairy science 100, 1295-1307.

Somers, J., Frankena, K., Noordhuizen-Stassen, E., Metz, J., 2005. Risk factors for digital dermatitis in dairy cows kept in cubicle houses in The Netherlands. Preventive veterinary medicine 71, 11-21.

Speijers, M., Baird, L., Finney, G., McBride, J., Kilpatrick, D., Logue, D., O'Connell, N., 2010. Effectiveness of different footbath solutions in the treatment of digital dermatitis in dairy cows. Journal of dairy science 93, 5782-5791.

Teixeira, A., Machado, V., Caixeta, L., Pereira, R., Bicalho, R., 2010. Efficacy of formalin, copper sulfate, and a commercial footbath product in the control of digital dermatitis. Journal of dairy science 93, 3628-3634.

Tremblay, M., Bennett, T., Döpfer, D., 2016. The DD Check App for prevention and control of digital dermatitis in dairy herds. Preventive veterinary medicine 132, 1-13.

Velthuis, A., De Jong, M., Kamp, E., Stockhofe, N., Verheijden, J., 2003. Design and analysis of an Actinobacillus pleuropneumoniae transmission experiment. Preventive veterinary medicine 60, 53-68.

Viera, A.J., Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. Fam Med 37, 360-363.

Vink, W., Jones, G., Johnson, W., Brown, J., Demirkan, I., Carter, S., French, N., 2009. Diagnostic assessment without cut-offs: Application of serology for the modelling of bovine digital dermatitis infection. Preventive veterinary medicine 92, 235-248.

Walker, R.L., Read, D.H., Loretz, K.J., Nordhausen, R.W., 1995. Spirochetes isolated from dairy cattle with papillomatous digital dermatitis and interdigital dermatitis. Veterinary Microbiology 47, 343-355.

Zinicola, M., Lima, F., Lima, S., Machado, V., Gomez, M., Döpfer, D., Guard, C., Bicalho, R., 2015. Altered Microbiomes in Bovine Digital Dermatitis Lesions, and the Gut as a Pathogen Reservoir. PloS one 10, e0120504.

4

Genetic variance components of host susceptibility, infectivity and R_0 for Digital Dermatitis in dairy cattle

Floor Biemans[1,2], Mart C.M. de Jong[1], Piter Bijma[2]

[1] Quantitative Veterinary Epidemiology Group, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands

[2] Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands

In preparation

Abstract

For an infectious disease, the probability that a susceptible individual gets infected depends on its own susceptibility, on the number of infected contact individuals ("group mates"), and on the infectivity of those group mates. Together, susceptibility, infectivity, and the duration of the infectious period determine the basic reproduction ratio (R_0). Breeding strategies that aim to reduce the prevalence should focus on reducing R_0, preferably to a value below one. Here we estimate genetic variance components for host susceptibility, infectivity, and R_0 for Digital Dermatitis (DD), an endemic infectious claw disease.

We obtained phenotype data for 1513 Holstein-Friesian cows of twelve Dutch dairy farms. The cows were scored for DD disease status every two weeks for eleven times. The genotype data consisted of 75904 SNPs for 1401 of the phenotyped cows. Using four generalized linear mixed models, we modelled the probability that a susceptible individual became infected in an observation interval. All models included a genetic effect for susceptibility; model 2 and 4 also included a genetic effect for infectivity, while model 1 and 2 included a random interaction between farm and period. All models corrected for the variation in exposure of the susceptible individuals to infectious group mates.

Models without an infectivity effect showed significantly lower bias, while models with an infectivity effect tended to have slightly higher accuracy. The estimated genetic variation was substantially higher for susceptibility than for infectivity. The GEBV for R_0 of the model without infectivity and without the interaction term ranged from 0.62 to 6.68. The estimated additive genetic standard deviation for R_0 was large, ~1.28, while the mean R_0 was 2.36, so only about one genetic standard deviation greater than 1. After correcting for bias, the GEBV for R_0 showed large variation, six cows had a GEBV smaller than one, and the approximate accuracy of the GEBV was ~0.6. These results show that genetic selection against DD is very promising; there is substantial heritable variation and a meaningful accuracy can be obtained from a limited amount of data.

4.1 Introduction

Disease transmission in a population is a dynamic process. The probability that a susceptible individual gets infected depends on its own susceptibility, the number of infected contact individuals ("group mates") and on the infectivity of those group mates. The composition of the infectious fraction in the population will vary over time as some individuals get infected while others recover. To make optimal genetic inference, this variation should be taken in to account. However, current studies on genetic variability underlying infectious diseases tend to focus on individual differences in susceptibility (or resistance) only (Woolhouse et al., 1998; Springbett et al., 2003). Even without heterogeneity in infectivity, there is still the variation in number of infected individuals to which susceptible recipients are exposed. This variation is due to chance, and due to differences in susceptibility of the group members. Moreover, most likely there is variation in infectivity as well, demonstrated for example by superspreaders, being individuals that infect substantially more individuals compared to a typical infectious individual (Stein, 2011).

An individual's infectivity affects the disease status of other individuals, rather than its own disease status. If the observed variation in infectivity has a genetic component, then that infectivity is an indirect genetic effect (IGE). IGEs can have a considerable effect on the rate and direction of evolution by natural selection, and on response to selective breeding (Griffing, 1967; Bijma and Wade, 2008). Hence, IGEs can and should be used for genetic improvement of populations, whenever they play a role. Thus, to make optimal use of all variation that exists with respect to diseases transmission, both host susceptibility and host infectivity should be taken into account. A key question is, thus, whether variation in (especially) infectivity can effectively be estimated from data on individual disease status.

Together, susceptibility, infectivity, and the duration of the infectious period determine the basic reproduction ratio (R_0). The R_0 is the average number of secondary cases caused by a typical infectious individual in a fully susceptible population (May and Anderson, 1987). The R_0 contains information on the ability of an infection to establish itself in the population (May and Anderson, 1987). The threshold value is one; if $R_0 < 1$ an infectious individual will infect on average less than one susceptible individual and the disease will die out with certainty. If $R_0 > 1$, a disease can affect a substantial proportion of the population. For endemic diseases in homogeneous populations, the prevalence in the equilibrium follows from R_0 and equals $1 - \frac{1}{R_0}$. Breeding strategies that aim to reduce the prevalence level should thus focus on reducing R_0, preferably to a value below one. Because the R_0 depends

on both susceptibility and infectivity, genetic variation in both those traits should be taken into account.

In this study, we focus on an endemic infectious disease called Digital Dermatitis (DD). DD is a claw disorder that affects (mainly) the hind feet of dairy cattle (Read and Walker, 1998; Sogstad et al., 2005). Typically, a round lesion forms above the interdigital space next to the heel bulbs (Walker et al., 1995). These lesions can be painful and prone to bleed, and can develop filiform papillae or be surrounded by hyperkeratotic skin with hairs longer than normal (Read and Walker, 1998). Cows that are severely affected bear their weight on the toes of the affected foot, shake the foot as if in pain, and show reluctance to move (Bassett et al., 1990; Collighan and Woodward, 1997; Read and Walker, 1998). Therefore, DD has an impact on the welfare of cows and causes economic losses for the farmer (Bruijnis et al., 2012a, b).

The prevalence of DD is affected by many factors, such as herd size, lactation stadium, flooring system, climate, and breed (Holzhauer et al., 2006). Optimizing management strategies is one way to reduce the DD prevalence on a dairy farm (Wells et al., 1999). Another way is to improve claw health through genetic selection (Van der Waaij et al., 2005; van der Linde et al., 2010; Van der Spek et al., 2013). Variation in host susceptibility for DD exists (Capion et al., 2012) and part of this variation is genetic (Van der Waaij et al., 2005). Whether there is genetic variation in host infectivity is unknown.

The objective of this research is to quantify the genetic variation in host susceptibility, host infectivity and R_0 for DD. Thereafter, the susceptibility and infectivity of each animal are estimated, and these estimates are used to calculate individual estimated breeding values for R_0. Additionally, for model validation, models with and without genetic variation for infectivity are compared for their ability to predict whether a susceptible animal gets infected.

4.2 Material and methods

4.2.1 Phenotype data

Phenotypes for DD were collected on twelve dairy farms in the Netherlands, between November 2014 and April 2015. Two observers (author FB being one of them) visited these farms eleven times with a two-week interval between visits (Biemans et al., 2017a). On each farm, one of the observers rinsed and scored the feet using the method of Relun et al. (2011), while the other recorded cow ID and the DD-status of the cow. All feet received a score from the standardized classification developed by Döpfer et al. (1997) and Berry et al. (2012). This classification comprises six distinct classes M0, M1, M2, M3, M4, and M4.1, where M0 is skin without macroscopic lesions, M1 is a small lesion of 0-2 cm, M2 is a lesion

of >2 cm, M3 is a lesion covered by a scab, M4 is irregular skin with dyskeratotis or surface proliferation, and M4.1 is a small lesion (M1) in addition to irregular skin (M4). A foot scored as M0 was classified as susceptible, while a foot scored as M1, M2, M3, M4, or M4.1 was classified as infected and infectious.

Farmers were not informed on the DD status of the cows, but were allowed to identify lesions themselves and treat cows using their normal routine. Phenotypes were collected on 1513 cows, of which 1401 cows were genotyped (see below). On average, a cow was scored 8.7 times. Table 4.1 gives an overview of some characteristics of the farms enrolled in the study.

4.2.2 Genotype data

The Holstein-Friesian cows were genotyped with the Eurogenomics 10K chip. In the final analysis only cows with a call rate of >0.85 were included (n = 1401). Before imputation, quality control was performed on the data following the standard procedure of breeding company CRV. A genome-wide marker (SNP) was included only when the following criteria were met: 1) observed frequency that deviated <0.15 from expected Hardy Weinberg frequency; 2) minor allele frequency >0.025. Furthermore, inconsistent genotypes between parents and offspring were set to missing. The SNPs that passed the quality control were imputed to a set of 76438 SNPs based on the Illumina BovineSNP50 chip and a custom chip from breeding company CRV, with a reference population of >1000 animals with genotypes on both chips. Thereafter, quality control was performed on the imputed data. A SNP was included in the final analysis only when the following criteria were met: 1) no strong deviation from Hardy Weinberg equilibrium (p-value $>1*10^{-15}$); 2) missing rate <0.05; 3) minor allele frequency >2%. In total 75904 SNPs passed the quality control and were included in the final analysis.

4.2.3 Models

In this section, we develop a generalized linear mixed model (GLMM) to estimate genetic parameters for susceptibility and infectivity. To develop the GLMM, we need to find the probability that a susceptible individual becomes infected in an observation interval. In the following, therefore, we first present an epidemiological model, then derive the infection probability from this model, and finally present the resulting GLMM.

Table 4.1. Characteristics of the farms enrolled in the study.

Farm	# Cows examined[1]	Cows genotyped[1]	# Observations[2]	Average *Δt* (days)	# Foot baths[3]	Prevalence (SD)[4]	
						Cow level	Foot level
A	134	116	11	14	7	78.0 (5.4)	69.6 (6.6)
B	105	101	11	14	0	56.3 (7.5)	46.9 (7.9)
C	159	162	11	14	5	49.7 (2.8)	40.2 (1.9)
D	118	116	11	14	7	57.8 (5.0)	49.2 (5.1)
E	102	90	11	13.6	9	62.8 (5.0)	54.6 (5.4)
F	133	112	10	15.56	10	59.2 (10.0)	48.7 (10.4)
G	100	98	11	14	3	65.6 (8.1)	58.2 (7.6)
H	189	180	11	14	7	64.9 (6.2)	56.7 (5.8)
I	104	75	11	14	0	56.4 (5.1)	45.6 (4.9)
J	88	88	11	14	0	65.8 (10.8)	58.1 (10.9)
K	130	116	9	14	13	63.6 (9.6)	52.5 (8.5)
L	151	147	11	13.9	3	70.9 (7.2)	62.0 (7.7)
Total	1513	1401	129	14.07	64	62.6 (7.5)	53.9 (11.0)

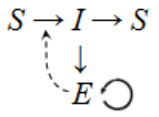
[1] Total number of different cows on a farm.

[2] Total number of farm visits.

[3] Number of footbaths given during the study period.

[4] Average percentage scored as infected (class M1, M2, M3, M4, or M4.1) with the standard deviation (SD).

DD is an endemic infectious disease with infections reoccurring in the same animals. The disease is persistent even in small cattle herds. We therefore used a stochastic compartmental susceptible-infected-susceptible-model (SIS-model) with an environmental route (E) (see e.g., de Rueda et al., 2015) to model disease transmission,



.

In this model, infection of a susceptible individual occurs randomly with a probability per observation period, depending on model parameters, the number of infected claws in the group, and the infection pressure coming from the environment. We modelled the probability that a susceptible individuals gets infected with a generalized linear mixed model (GLMM) as described previously in, for example, Velthuis et al. (2003), Anche et al. (2015), and Biemans et al. (2017b).

In the SIS-model with an environmental route (E), the expected rate with which susceptible individuals get infected is $\beta S\frac{I+E}{N}$, where I is the number of infectious claws, S the number of susceptible claws, and $S + I = N$ the total number of claws in a group (twice the number of cows). The *E* is the infection pressure coming from the environmental reservoir, expressed as the equivalent number of currently infected individuals (*i.e.*, *I* and *E* are on the same scale). The β is the transmission rate parameter that contains information on the contact rate and transmission probability between individuals (Roberts and Heesterbeek, 1993).

Because our interest is in genetic variation in susceptibility and infectivity among individuals, we consider the pairwise β between a susceptible and an infected individual. This pairwise β depends on the susceptibility genotype of the susceptible individual and the infectivity genotype of the infectious individual. Thus, the transmission rate parameter β_{ij} from an infectious individual *j* with infectivity φ_j to susceptible individual *i* with susceptibility γ_i is

$$\beta_{ij} = c\gamma_i\varphi_j, \qquad \text{(Equation 4.1)}$$

where *c* is the overall contact rate.

The expected rate with which susceptible individual *i* gets infected when exposed to *all* infectious claws in the group is the sum of the rates with each infected claw, and thus depends on the susceptibility of individual i, the number of infectious claws in the group, and their average infectivity:

$$Infection\ rate_i = c\ \gamma_i \frac{\sum_j \varphi_j}{I_g} \frac{E + I_{tot}}{N}, \qquad \text{(Equation 4.2)}$$

where $\beta_i = c\ \gamma_i \frac{\sum_j \varphi_j}{I_g}$, which is the pairwise β_{ij} averaged over the infectious group mates *j* of focal individual *i*.

Here, we distinguished between the claws of infectious individuals that were genotyped, I_g, and the total number of infectious claws, I_{tot}. We could estimate infectivity only for the genotyped individuals, but also non-genotyped infectious individuals contributed to infection. To account for all infectious individuals, we assumed that the claws of the non-genotyped cows had the same average infectivity as the claws of the genotyped cows ($\sum_j \varphi_j / I_g$). Thus, in the total number of infectious claws, both the claws from genotyped and non-genotyped infectious individuals are included, $I_g \subseteq I_{tot}$.

Strictly speaking, in the $\frac{\sum_j \varphi_j}{I_g}$ term, we should average the infectivity over all claws that contribute to the current infection pressure, which is the sum of the currently infectious claws and the infection pressure from previous infectious claws via the environment. In other words, all claws that contribute to $E_g + I_g$ should also be included in the $\frac{\sum_j \varphi_j}{I_g}$ term. However, in the statistical software we did not manage to keep track of all those infectivity genotypes. For this reason, only the currently infected claws were included in the $\frac{\sum_j \varphi_j}{I_g}$ term. Thus, our estimates of genetic variation in infectivity utilize only part of the variation in the infection pressure. This is issue is further addressed in the Discussion.

The infection rate varied over time, depending on the average infectivity at *t*, the start of the interval Δt. The probability for cow *i* to get infected (be a case) when exposed to *all* infectious claws during Δt follows from assuming a Poisson process within Δt. It is the probability of a non-zero outcome from a Poisson distribution. Following from Equation 4.2,

$$P_i(t) = 1 - e^{-c\gamma_i \left(\frac{\sum_j \varphi_j}{I_g(t)}\right) \frac{E(t) + I_{tot}(t)}{N(t)} \Delta t}, \qquad \text{(Equation 4.3)}$$

where, $P_i(t)$ is the probability that cow *i* is a case in interval Δt.

The number of cases within an interval follows a binomial distribution with a probability that follows from a Poisson process. Therefore, the complementary log-

log is the appropriate link function to connect the explanatory variables to the expected value of the observed variable (McCullagh, 1984). From Equation 4.3,

$$\text{cloglog}(P_i(t)) = \log(c) + \log(\gamma_i) + \log\left(\frac{\sum_j \varphi_j}{I_g(t)}\right) + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right).$$

(Equation 4.4)

The last term in Equation 4.4, $\log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right)$, is an offset, *i.e.*, an "explanatory variable" with coefficient equal to 1. The offset accounts for the infectious pressure coming from the environment at t ($E(t)/N(t)$), the fraction of infectious cows (genotyped and non-genotyped) at t ($I_{tot}(t)/N(t)$), and the length of the interval (Δt).

The infectious pressure coming from the environment was calculated as described in detail in Biemans et al. (2017a). In short, claws that were infected at an earlier stage still contribute partly to the current environmental reservoir. The contribution was assumed to decrease each interval Δt with factor λ, which may be interpreted as a survival rate of the pathogen. So the number of pathogens from a claw that was infectious at t that are in the environment is a fraction λ at t+1, a fraction λ^2 at t+2, a fraction λ^3 at t+3, etc. With survival rate λ, the values for $E(t)$ were calculated as,

$$E(t) = \lambda(I_{tot}(t-1) + E(t-1)), \qquad \text{(Equation 4.5)}$$

where $I_{tot}(t-1)$ is the total number of infectious claws at t - 1, and $E(t-1)$ is the environmental reservoir at t - 1.

Because we did not observe the number of infected claws in the period before the first (t = 1) observation, they were estimated with a linear model. We fitted the model to the number of infected claws over the observation period. The intercept of the model was used as the average number of infectious claws before observations started ($I_{tot}(t \leq 0)$). Thereafter, the value for the environmental reservoir before the first observation was estimated as, $E(t=0) = \frac{\lambda}{1-\lambda} I_{tot}(t=0)$, which was used in Equation 4.5.

Equation 4.4 is linear in the logarithm of susceptibility, but not in the logarithm of infectivity. Therefore, in the models in which the infectivity of the group mates was included (model 2 and 4, see below), we first moved the number of claws of the genotyped infectious cows into the offset, and then linearized the equation for infectivity following (Biemans et al., 2017b),

$$\text{cloglog}(P_i(t)) = \log(c) + \log(\gamma_i) + \log\left(\sum_j \varphi_j\right) + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)\, I_g(t)}\Delta t\right)$$

$$\approx \log(c) + \log(\gamma_i) + \sum_j \log\left(\varphi_j\right) + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right). \qquad \text{(Equation 4.6)}$$

4.2.4 Implementation

Using a GLMM, we modelled the expectation of the number of cases over the number of susceptible feet of animal *i* within interval *Δt*, $p_{ikl}(t) = E\left(\frac{C_i(t)}{F_i(t)}\right)$. Only the hind feet of the cows were scored, so a susceptible cow could have one or two susceptible feet (*F*) at the start of an interval, that were zero, one, or two cases by the end of the interval. Thus, the number of cases *C* (0, 1 or 2) for each susceptible animal followed a binomial distribution with binomial total *F* (1 or 2).

Table 4.2. Overview of the fixed and random effects that are included in the four models.[1]

Model	Random effects		
1	Gen. susceptibility focal ind.	-	Farm*period
2	Gen. susceptibility focal ind.	Gen. infectivity herd mates	Farm*period
3	Gen. susceptibility focal ind.	-	-
4	Gen. susceptibility focal ind.	Gen. infectivity herd mates	-

[1] All models contained fixed effects for farm, period, parity, and months in milk; and a non-genetic random animal effect for the susceptible animal.

We tested four models (Table 4.2). Model 1 included a random genetic effect for susceptibility only, and a random interaction between farm and period,

$$\text{cloglog}(p_{iklt}(t)) = c_0 + Farm_k + Period_t + Parity_l + c_1 MIM +$$

$$Farm_k.Period_t + Animal_i + \log(\gamma_i) +$$

$$\log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right), \qquad \text{(Model 1)}$$

where c_0 is the intercept. The fixed effects included were for farm ($Farm_k$ with k = A to L), period ($Period_t$ with t = 1 to 10), parity ($Parity_l$ with l = 1, 2, or >2), and months in milk (*MIM*, a continuous variable). The random effects were the interaction between farm and period ($Farm_k.Period_t$ with k = A to L and t = 1 to 10), a non-genetic animal effect for animal *i* to account for repeated observations in different periods

(*Animal*$_i$), and an additive genetic effect for susceptibility of animal *i* (log(***γ***) $\sim$ N($\mathbf{0}, \mathbf{G}\sigma_a^2$), where **G** is the genomic relationships matrix among animals).

Model 2 included random genetic effects for both susceptibility and infectivity,

$$\text{cloglog}(p_{ijklt}(t) = c_0 + Farm_k + Period_t + Parity_l + c_1 MIM + Farm_k.Period_t + Animal_i + \log(\gamma_i) + \textstyle\sum_j \log(\varphi_j) + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right), \quad \text{(Model 2)}$$

where $\sum_j \log(\varphi_j)$ are the random genetic effects for infectivity of the infectious group mates *j* of animal *i*, with (log(***φ***) $\sim$N($\mathbf{0}, \mathbf{G}\sigma_a^2$)).

We expected that the interaction between farm and period could be partly confounded with the genetic effect for infectivity, because previous IGE-studies showed that omitting group-effects may substantially inflate estimated genetic parameters for infectivity (Bergsma et al., 2008). To investigate this issue, we dropped the farm by period interaction from models 1 and 2, giving models 3 and 4,

$$\text{cloglog}(p_{iklt}(t)) = c_0 + Farm_k + Period_t + Parity_l + c_1 MIM + Animal_i + \log(\gamma_i) + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right), \quad \text{(Model 3)}$$

$$\text{cloglog}(p_{ijklt}(t)) = c_0 + Farm_k + Period_t + Parity_l + c_1 MIM + Animal_i + \log(\gamma_i) + \textstyle\sum_j \log(\varphi_j) + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right), \quad \text{(Model 4)}$$

4.2.5 Data analyses

The **G**-matrix was computed using method 1 of VanRaden (2008), using the calc_grm software (Calus and Vandenplas, 2016). We fitted the four models with ASReml v4.1.0 (Gilmour, 2015). Model fit was assessed with Akaike information criterion (AIC). The individual estimates for susceptibility and infectivity from ASReml refer to $\widehat{\log(\gamma_i)}$ and $\widehat{\log(\varphi_i)}$. The estimates are on the log scale because of the complementary log-log link function (Equation 4.6). The susceptibility and infectivity estimates from ASReml are zero on average, $\overline{\log(\gamma)} = \overline{\log(\varphi)} = 0$, so $\bar{\gamma} \approx \bar{\varphi} \approx 1$. By taking the exponent of the estimates from ASReml, we obtained the genomic

estimated breeding values (GEBV) for susceptibility and infectivity relative to a typical (average) individual that has a GEBV of 1.

4.2.6 Cross validation

We validated the GEBV to obtain their bias and accuracy. We, therefore, performed a twelve-fold cross-validation on all four models. In each analysis the dependent variable, *i.e.*, the cases (*C*), from one farm were censored from the dataset. For each susceptible animal *i* of the censored farm at *t*, we predicted the number of cases over the number of susceptible feet $(C_i(t)/F_i(t))$ based on information of the other eleven farms. The value for $C_i(t)/F_i(t)$ that was predicted by the models is referred to as the predicted probability $\widehat{P}_i(t)$.

However, as the fixed effects were nonlinear on the normal scale because they were estimated with a complementary log-log link function, correction of the observed records for fixed effects was not straightforward. To solve this issue, we translated both the predicted probabilities and the observed records to a standard (*i.e.*, average) farm. Subsequently, we validated the models using the weighted correlation and regression of observed records on predicted probabilities (see Appendix for detailed methods).

4.2.7 Basic reproduction ratio

With the estimate breeding values for susceptibility and infectivity, we calculated individual breeding values for the basic reproduction ratio ($\hat{A}_{R_0,i}$). The $\hat{A}_{R_0,i}$ is the product of the relative susceptibility ($\hat{\gamma}_i$), the relative infectivity ($\hat{\varphi}_i$), the contact rate (c), and the average duration of the infectious period ($1/\alpha$) (Anche et al., 2014),

$$\hat{A}_{R_0,i} = \hat{\gamma}_i \hat{\varphi}_i c/a. \qquad \text{(Equation 4.7)}$$

In model 1 and 3 variation in infectivity is not estimated, hence in these models $\hat{\varphi}_i = 1$. With an average R_0 for DD of 2.36 on these farms (Biemans et al., 2017a) and the average product of the estimated relative susceptibility and relative infectivity, the value for *c/a* was calculated as $c/\alpha = \frac{2.36}{\overline{\hat{\gamma}_i \hat{\varphi}_i}}$.

4.3 Results

4.3.1 Model fit

The fit of model 3 was better compared to model 1 (lower AIC), and the fit of model 4 was better compared to model 2 (footnote of Table 4.3). The AIC of model 1 could only be compared to the AIC of model 3 because the dataset of these models was

the same, *i.e.*, the offset in model 1 and 3 was different from the offset in models 2 and 4 because only genotyped individuals could be included in the infectivity term. Similarly, the AIC of model 2 could only be compared to the AIC of model 4.

4.3.2 Fixed effects

The farm effect was significant (P < 0.05) in model 1 and 3, but not in model 2 and 4. For all four models there was a significant effect for period, parity, and months in milk. The probability of getting infected during an interval increased the first six periods and stabilized thereafter. The transmission rate parameter increased with increasing parity; it was 21% higher for parity 2 compared to parity 1, and 69% higher for parities >2 compared to parity 1. For months in milk, the transmission rate parameter decreased by 4% with every month in milk.

4.3.3 Estimated variance components

Table 4.3 shows the estimated variance components and their standard error (SE) on the log scale. The estimated variance is approximately the same on the normal scale, because $var(\ln(x)) \approx var(x)$ around $\ln(x) = 0$ (x = 1) (Hosmer et al., 2008). For model 1, 2, and 4, the estimated genetic variance for susceptibility was about 0.55. In all these models a genetic effect for infectivity and/or the interaction term between farm and period was included. For model 3, the estimated genetic variance for susceptibility was lower, about 0.49. Similarly, the variance of the non-genetic random animal effect was about 0.95 for model 1, 2, and 4, and lower, about 0.92, for model 3. In model 3 there was no genetic effect for infectivity and no interaction between farm and period included.

Table 4.3. Estimated variance components and their standard error (SE) for the genetic effect of susceptibility and infectivity, the interaction between farm and period, and the animal effect for the four models.

Model	Estimated variance (SE) of the random terms			
	Susceptibility	**Infectivity**	**Farm*period**	**Animal**
1[1]	0.5554 (0.1417)	-	0.2618 (0.0502)	0.9493 (0.1300)
2[2]	0.5552 (0.1408)	0.0044 (0.0021)	0.0992 (0.0703)	0.9479 (0.1300)
3[1]	0.4896 (0.1313)	-	-	0.9218 (0.1234)
4[2]	0.5542 (0.1414)	0.0075 (0.0015)	-	0.9486 (0.1299)

[1] The difference in AIC between model 1 and model 3 was: $AIC_1 - AIC_3 = 169.18$

[2] The difference in AIC between model 2 and model 4 was: $AIC_2 - AIC_4 = 3.94$

4.3.4 Relative host susceptibility and infectivity

The genomic estimated breeding values (GEBV) for susceptibility from model 1, 2 and 4 were approximately the same and ranged from 0.26 to 3.45, while susceptibility GEBV from model 3 ranged from 0.28 to 3.06. A cow with a susceptibility GEBV of 0.26 is about four times less susceptible than an average cow, while a cow with a susceptibility GEBV of 3.45 is about 3.5 times more susceptible than an average cow.

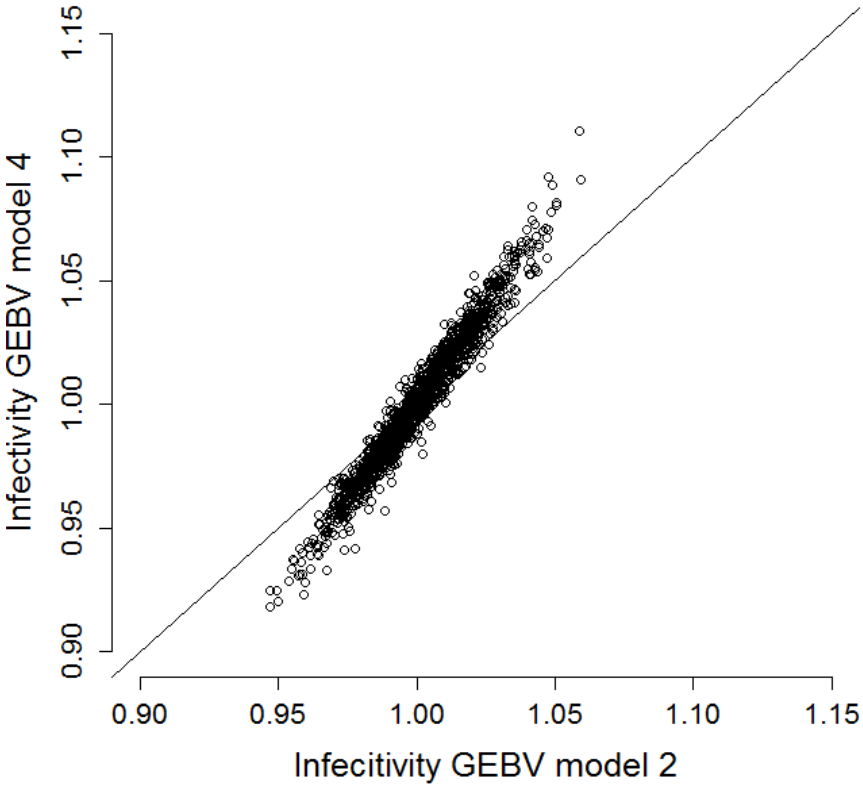


Figure 4.1. Genomic estimated breeding values (GEBV) for infectivity from model 2 versus model 4. Each point represents one cow. The line shows y = x.

The infectivity GEBV showed less variation compared to the susceptibility GEBV. Figure 4.1 shows the infectivity GEBV from model 2 plotted against the infectivity GEBV from model 4. The infectivity GEBV ranged from 0.95 to 1.06 for model 2, and

from 0.92 to 1.11 for model 4. Thus, the infectivity GEBV from model 2 show less variation than those from model 4. Part of the variation that is attributed to the genetic infectivity effect in model 4 is attributed to interaction between farm and period in model 2. This suggests that the infectivity GEBV from model 4 may include both a genetic and a non-genetic component, and may therefore be inflated, while the GEBV from model 2 may better represent the true breeding values.

4.3.5 Cross-validation

In the cross-validation, the number of cases over the number of susceptible feet ($C_i(t)/F_i(t)$) was predicted for each susceptible animal on a censored farm. We used these predictions to validate the GEBVs to obtain their bias and accuracy. To facilitate comparison, both predictions and observations were translated to a standard farm and were averaged by animal (Appendix).

Figure 4.2 and Table 4.4 show the weighted linear regression and correlation coefficients between the average observed number of cases over the number of susceptible feet (C/F) and the average predicted probability. Bias was smallest for models without the infectivity effect (Models 1 and 3) because regression coefficients of these models were closest to one. The weighted correlations coefficients were higher for models with the infectivity effect (Model 2 and 4). Thus, models without an infectivity effect showed significantly lower bias, while models with an infectivity effect tended to have slightly higher accuracy.

Table 4.4. Weighted linear regression and correlation coefficients between the average observed number of cases over the number of susceptible feet (C/F) and the average predicted probability for the observations.

Model	Linear regression		Correlation coefficient (SE)
	Intercept (SE)	Regression coefficient (SE)	
1	0.151 (0.030)	0.815 (0.128)	0.200 (0.027)
2	0.197 (0.021)	0.655 (0.103)	0.216 (0.027)
3	0.140 (0.032)	0.847 (0.135)	0.198 (0.027)
4	0.206 (0.020)	0.621 (0.096)	0.223 (0.027)

4.3.6 Basic reproduction ratio

Figure 4.3 shows the GEBV for the basic reproduction ratio (R_0), calculated from susceptibility GEBV from model 3. The mean susceptibility GEBV was 1.081, which is slightly above one because of the transformation from the log scale to the normal scale. With an average R_0 for DD on these farms of 2.36 (Biemans et al., 2017a), it

follows that *c/a* is estimated at 2.182. GEBV for R_0 ranged from 0.62 to 6.68. There were 38 cows with a $\hat{A}_{R_0,i} < 1$ out of the 1401 cows included.

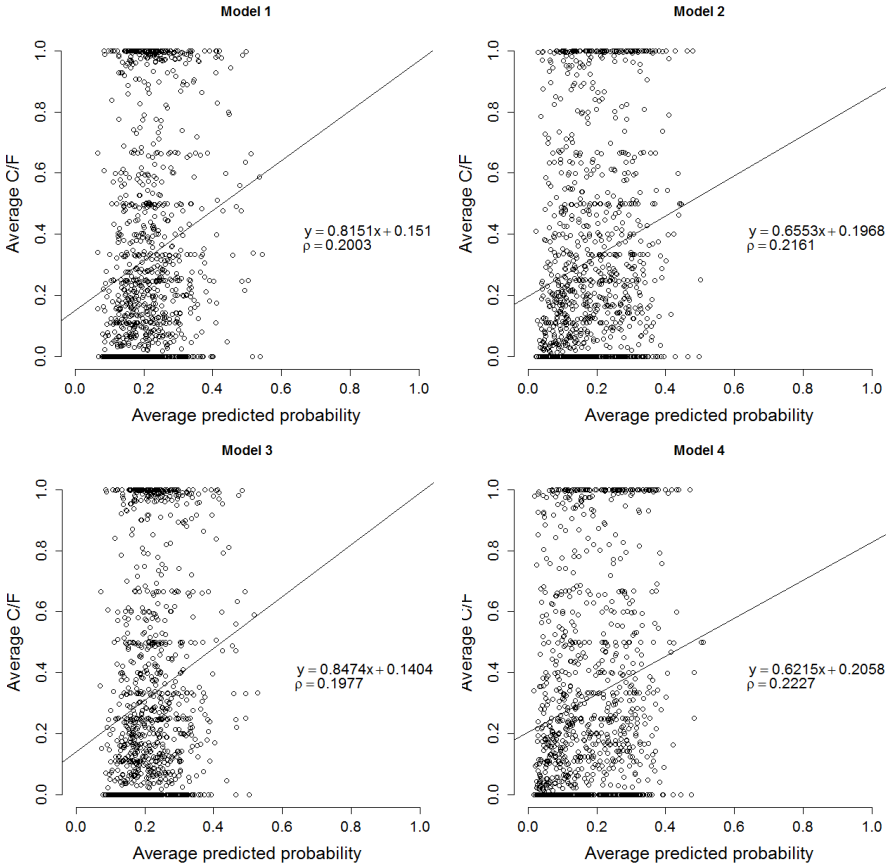


Figure 4.2. Weighted linear regression and correlation coefficients between the average observed number of cases over the number of susceptible feet (C/F) and the average predicted probability for the observations.

Results in Figure 4.3 have not been corrected for the bias observed in the probability that an individual is infected (Figure 4.2, Model 3, b = 0.8474). Assuming that the bias in GEBV for susceptibility is approximately proportional to the observed bias in infection probability implies that GEBV for R_0 in Figure 4.3 have to be shrunken, relative to their mean, by a factor 0.8474. It can be shown that this assumption is reasonable because predicted probabilities in Figure 4.2 are much smaller than 1, suggesting that multiple infections of the same susceptible claw

within a period are relatively unlikely. Resulting GEBV for R_0 range from 0.88 to 6.02. Hence, after correcting for the bias in GEBV, the range in GEBV for R_0 is still very large, and there are still six individuals with a GEBV for R_0 below 1. (Note that DD is absent when R_0 in a population is smaller than one.)

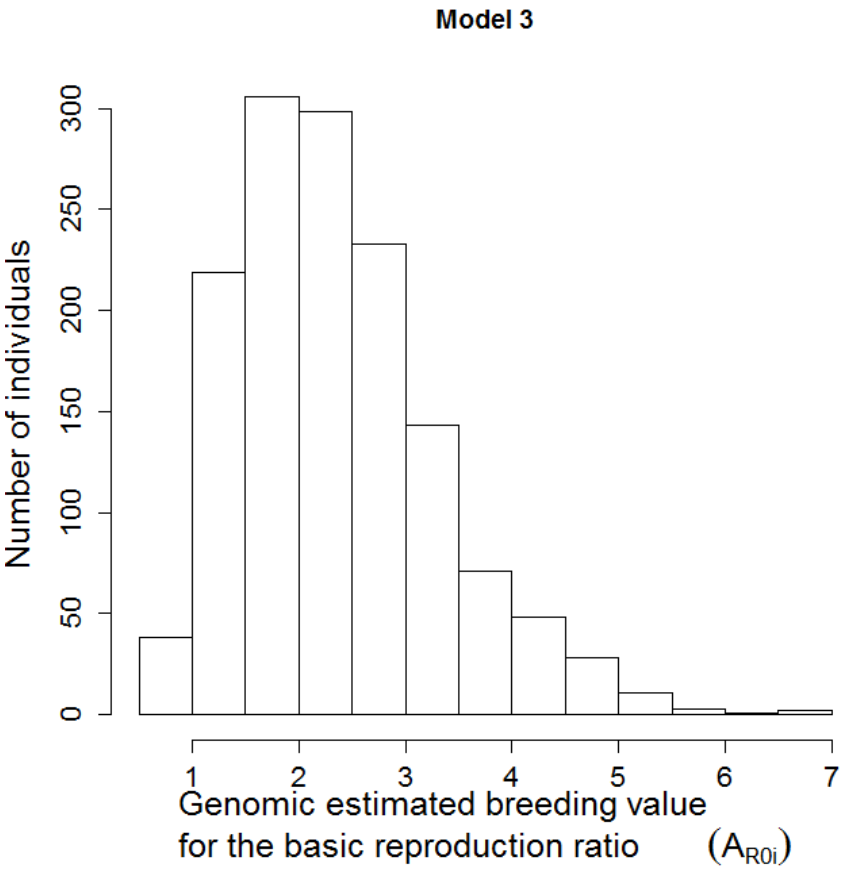


Figure 4.3. Histogram of the individual GEBV for the basic reproduction ratio for all genotyped animals, based on results from model 3.

The additive genetic variance of R0 is given by $\sigma^2_{A_{R_0}} = \left(\bar{\gamma}^2\sigma^2_{\varphi} + \bar{\varphi}^2\sigma^2_{\gamma} + \sigma^2_{\gamma}\sigma^2_{\varphi}\right)\left(\frac{c}{\alpha}\right)^2 \approx \left(\sigma^2_{\gamma} + \sigma^2_{\varphi} + \sigma^2_{\gamma}\sigma^2_{\varphi}\right)\left(\frac{c}{\alpha}\right)^2$ (Anche et al., 2014). Substituting the estimate from Model 3 (Table 4.3), taking the square root and shrinking the result by 0.8474 to account for the bias, yields an estimated additive genetic standard deviation for R0 of 1.28. This result shows that the current R0 of 2.36 (Biemans et al.,

2017a) is only about one genetic standard deviation greater than 1. Hence, this suggests that a genetic improvement of R0 by about one genetic standard deviation would be sufficient to eradicate DD.

The approximate accuracy of the GEBV for R_0 follows from the ratio of the standard deviation of GEBV over the additive genetic standard deviation of R_0. After shrinking GEBV by 0.8474 to account for the bias, this accuracy equals ~60%. Hence, despite the relatively small reference population of ~1400 individuals in this experiment, GEBV for R_0 have meaningful accuracy.

4.4 Discussion

We estimated the additive genetic variation in host susceptibility, infectivity, and R_0 for DD in dairy cattle. Furthermore, we calculated GEBVs for susceptibility, infectivity, and R_0 for each animal. Four models were compared for their ability to predict whether a susceptible animal gets infected. All four models included a genetic effect for susceptibility; model 2 and 4 also included a genetic effect for infectivity, while model 1 and 2 included an interaction term between farm and period. All models corrected for the variation in exposure of the susceptible individuals to infectious group mates via the offset of the model. The estimated genetic variation was substantially higher for susceptibility than for infectivity. The estimated additive genetic standard deviation for R_0 was large, ~1.30, and the mean R_0 (2.36) was only about one genetic standard deviation greater than the important threshold value of 1. Furthermore, GEBV for R_0 (corrected for bias) showed large variation, six animals had a GEBV smaller than 1, and the approximate accuracy of GEBV was ~0.6. These results show that genetic selection against DD is very promising; there is substantial heritable variation and a meaningful accuracy can be obtained from a limited amount of data.

Farm, parity, period, and months in milk of the focal cow were included in the models as fixed effect. The transmission rate parameter was 21% higher for parity 2 compared to parity 1, and 69% higher for parity >2 compared to parity 1. The prevalence also increased with parity. This is in contrast with most previous studies where DD was most prevalent in first and second parity cows (Argáez-Rodríguez et al., 1997; Read and Walker, 1998). For months in milk, the transmission rate parameter decreased with 4% per month in milk. This is in agreement with Argáez-Rodríguez et al. (1997) who found that cows had the highest risk of DD in the first and third month of lactation, after which the risk decreased. The effect of parity and months in milk on the infectivity of a cow was not considered because incorporating these factors in the summed effect of the infectious group mates was difficult.

The infectivity estimates showed considerably less variation compared to the susceptibility estimates. However, this does not necessarily mean that there is indeed less variation in infectivity than in susceptibility. Because of technical difficulties, we included only the genetic effects of claws that were infectious at the start of the observation interval (*t*). However, the total infectious pressure was composed both of claws that were infectious at the start of the observation interval, and of claws that were infectious earlier. With an estimated survival rate (λ) of the pathogen of 0.9 (Biemans et al., 2017a), 90% of the total infectious pressure originates from claws that were infectious before the start of the observation interval. This suggests that we may have disregarded the majority of the heritable variation in infectivity. Hence, the relevance of genetic variation in infectivity may be substantially larger than suggested by estimates presented here (Table 4.3).

Unlike variation in susceptibility, variation in infectivity must be estimated indirectly. Infectivity estimates are based on the number of susceptible group mates that become infected and on differences in genotype among the infected group mates at different points in time. When there are multiple infected group mates, the accuracy of the infectivity estimates decreases (Anacleto et al., 2015). Especially in large groups, like this study (~100 cows), more records and groups are needed to estimate genetic variation in infectivity accurately. This issue is very similar to the estimation of indirect genetic effects from large groups (Bijma, 2010).

We included a random interaction between farm and period in the first two models, to account for non-genetic effects of infectivity. This interaction serves to avoid overestimation of the genetic variance in infectivity (Anche, 2016), similar to the inclusion of a random group effect in the analysis of indirect genetic effects (Bijma et al., 2007; Bergsma et al., 2008). The genetic effect for infectivity and the interaction term were partly confounded, because both effects reflect the number of susceptibles that became a case within a certain period on a certain farm. However, confounding is not complete because of genetic relationships between the infectious animals across farms and periods. Nevertheless, our results suggest that inclusion of a random farm.period effect is essential to avoid overestimation of the genetic variation due to infectivity.

The estimated variances for susceptibility (genetic and non-genetic) were lower for model 3 that did not include a genetic effect for infectivity nor an interaction between farm and period. Anacleto et al. (2015) showed that estimates for susceptibility are less accurate when genetic variation in infectivity is not accounted for. Indeed, we also found a slightly higher correlation in the cross-validation when infectivity was in the model, but this was accompanied by an inflation of the GEBV,

as shown by the regression coefficients in Figure 4.2. However, inflation of GEBVs can be remedied by shrinking them based on results of cross-validation, whereas a reduction in correlation cannot. Therefore, even when infectivity is not the trait of interest, it might be beneficial to include infectivity in the model to accurately estimate susceptibility GEBV (Anacleto et al., 2015).

In the cross-validation, we estimated a weighted correlation of about 0.2 between the observed and predicted number of cases over the number of susceptible feet. This value can be used to approximate the accuracy of the GEBV ($r_{g,\hat{g}}$) (Calo et al., 1973),

$$r_{g,\hat{g}} \approx \frac{r_{p,\hat{g}}}{\sqrt{h^2}}, \qquad \text{(Equation 4.8)}$$

where, $r_{p,\hat{g}}$ is the correlation between the observations and the predictions, and h^2 is the heritability of the trait. Heritability estimates for Digital Dermatitis from previous studies range from 0.05 to 0.29, depending on the model used (Schöpke et al., 2015). Assuming a heritability of 0.28, the accuracy of the predicted number of cases is 0.38. Note, this value represents the estimated correlation between the predicted number of cases for an individual (its "GEBV") and its true expected number of cases given its genes (its true "breeding value"). This value is somewhat smaller than the approximate accuracy of the breeding values for R0 presented above.

In general, studies on genetic variability of infectious diseases commonly focus on individual differences in susceptibility only, and those differences are estimated with a linear model that ignores variation in exposure among individuals. In this study, we used a GLMM to estimate genetic variability in susceptibility. Estimates were corrected for variation in exposure via the offset, and variation in infectivity of group mates was included. Further work is needed to quantify the benefits of such GLMMs over simpler linear models, to better account for the full genetic variation in infectivity via the environment, and to include genetic variation in the duration of the infectious period.

4.5 Conclusions

Genetic variance components for susceptibility and infectivity for Digital Dermatitis were estimated with four generalized linear mixed models. The model that included only a genetic effect for susceptibility and no interaction between farm and period had the best fit and predictive ability. Even though there was not a significant difference in infectivity among cows in this model, variation in exposure was still

accounted for via the offset. Furthermore, for each animal its relative susceptibility and relative infectivity compared to an average cow in this study were estimated. These estimates were used to calculate the individual GEBVs for the basic reproduction ratio. Estimated breeding values for R_0 ranged from 0.62 to 6.68, and the mean R_0 (2.36) was only about one genetic standard deviation greater than 1. Furthermore, genomic estimated breeding values for R_0 (corrected for bias) showed large variation, six cows had an GEBV smaller than 1, and the approximate accuracy of GEBV was ~0.6. These results show that lowering transmission of DD with selective breeding is very promising.

4.6 Acknowledgements

We thank Natasja Boots for her help with the phenotype collection, the farmers for their participation, and breeding company CRV for providing the genotype data. This work is part of the research programme “Meer met Minder” with project number 847.13.004, which is financed by the Netherlands Organisation for Scientific Research (NWO).

4.7 References

Anacleto, O., Garcia-Cortés, L.A., Lipschutz-Powell, D., Woolliams, J.A., Doeschl-Wilson, A.B., 2015. A novel statistical model to estimate host genetic effects affecting disease transmission. Genetics 201, 871-884.

Anche, M., de Jong, M., Bijma, P., 2014. On the definition and utilization of heritable variation among hosts in reproduction ratio R0 for infectious diseases. Heredity 113, 364-374.

Anche, M.T., 2016. Estimating host genetic effects on susceptibility and infectivity to infectious diseases and their contribution to response to selection. Wageningen University.

Anche, M.T., Bijma, P., De Jong, M.C., 2015. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47, 1-15.

Argáez-Rodríguez, F.d.J., Hird, D.W., de Anda, J.H., Read, D.H., Rodríguez-Lainz, A., 1997. Papillomatous digital dermatitis on a commercial dairy farm in Mexicali, Mexico: Incidence and effect on reproduction and milk production. Preventive Veterinary Medicine 32, 275-286.

Bassett, H., Monaghan, M., Lenhan, P., Doherty, M., Carter, M., 1990. Bovine digital dermatitis. Veterinary Record 126, 164-165.

Bergsma, R., Kanis, E., Knol, E.F., Bijma, P., 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics 178, 1559-1570.

Berry, S.L., Read, D.H., Famula, T.R., Mongini, A., Döpfer, D., 2012. Long-term observations on the dynamics of bovine digital dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. The Veterinary Journal 193, 654-658.

Biemans, F., Bijma, P., Boots, N.M., de Jong, M.C., 2017a. Digital Dermatitis in dairy cattle: the contribution of different disease classes to transmission. Epidemics In press.

Biemans, F., Jong, M.C., Bijma, P., 2017b. A model to estimate effects of SNPs on host susceptibility and infectivity for an endemic infectious disease. Genetics Selection Evolution 49, 53.

Bijma, P., 2010. Estimating indirect genetic effects: precision of estimates and optimum designs. Genetics 186, 1013-1028.

Bijma, P., Muir, W.M., Ellen, E.D., Wolf, J.B., Van Arendonk, J.A., 2007. Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. Genetics 175, 289-299.

Bijma, P., Wade, M., 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. Journal of evolutionary biology 21, 1175-1188.

Bruijnis, M., Beerda, B., Hogeveen, H., Stassen, E., 2012a. Assessing the welfare impact of foot disorders in dairy cattle by a modeling approach. Animal 6, 962-970.

Bruijnis, M., Beerda, B., Hogeveen, H., Stassen, E., 2012b. Foot disorders in dairy cattle: impact on cow and dairy farmer. Animal Welfare 21, 33-40.

Calo, L., McDowell, R., VanVleck, L.D., Miller, P., 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. Journal of Animal Science 37, 676-682.

Calus, M.P.L., Vandenplas, J., 2016. Calc_grm – a program to compute pedigree, genomic, and combined relationship matrices, ABGC, Wageningen UR Livestock Research.

Capion, N., Boye, M., Ekstrøm, C.T., Jensen, T.K., 2012. Infection dynamics of digital dermatitis in first-lactation Holstein cows in an infected herd. Journal of Dairy Science 95, 6457-6464.

Collighan, R.J., Woodward, M.J., 1997. Spirochaetes and other bacterial species associated with bovine digital dermatitis. FEMS Microbiology Letters 156, 37-41.

de Rueda, C.B., de Jong, M.C., Eblé, P.L., Dekker, A., 2015. Quantification of transmission of foot-and-mouth disease virus caused by an environment contaminated with secretions and excretions from infected calves. Veterinary research 46, 1.

Döpfer, D., ter Huurne, A., Cornelisse, J., Van Asten, A., Koopmans, A., Meijer, F., Schukken, Y., Szakall, I., Klee, W., Bosma, R., 1997. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and Campylobacter faecalis. Veterinary Record 140, 620-623.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. , 2015. ASReml User Guide Release 4.1 Structural Specification, 4.1 ed. VSN International Ltd, Hemel Hempstead, UK.

Griffing, B., 1967. Selection in reference to biological groups I. Individual and group selection applied to populations of unordered groups. Australian Journal of Biological Sciences 20, 127-140.

Holzhauer, M., Hardenberg, C., Bartels, C., Frankena, K., 2006. Herd-and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. Journal of dairy science 89, 580-588.

Hosmer, D.W., Lemeshow, S., May, S., 2008. Appendix 1: The Delta Method. Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition, 355-358.

May, R.M., Anderson, R.M., 1987. Transmission dynamics of HIV infection. Nature 326, 137.

McCullagh, P., 1984. Generalized linear models. European Journal of Operational Research 16, 285-292.

Read, D.H., Walker, R.L., 1998. Papillomatous digital dermatitis (footwarts) in California dairy cattle: clinical and gross pathologic findings. Journal of Veterinary Diagnostic Investigation 10, 67-76.

Relun, A., Guatteo, R., Roussel, P., Bareille, N., 2011. A simple method to score digital dermatitis in dairy cows in the milking parlor. J Dairy Sci 94, 5424-5434.

Roberts, M., Heesterbeek, H., 1993. Bluff your way in epidemic models. Trends in microbiology 1, 343-348.

Schöpke, K., Gomez, A., Dunbar, K., Swalve, H., Döpfer, D., 2015. Investigating the genetic background of bovine digital dermatitis using improved definitions of clinical status. Journal of Dairy Science.

Sogstad, Å., Fjeldaas, T., Østerås, O., Forshell, K.P., 2005. Prevalence of claw lesions in Norwegian dairy cattle housed in tie stalls and free stalls. Preventive veterinary medicine 70, 191-209.

Springbett, A., MacKenzie, K., Woolliams, J., Bishop, S., 2003. The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165, 1465-1474.

Stein, R.A., 2011. Super-spreaders in infectious diseases. International Journal of Infectious Diseases 15, e510-e513.

van der Linde, C., de Jong, G., Koenen, E.P.C., Eding, H., 2010. Claw health index for Dutch dairy cattle based on claw trimming and conformation data. Journal of Dairy Science 93, 4883-4891.

Van der Spek, D., Van Arendonk, J., Vallée, A., Bovenhuis, H., 2013. Genetic parameters for claw disorders and the effect of preselecting cows for trimming. Journal of dairy science 96, 6070-6078.

Van der Waaij, E., Holzhauer, M., Ellen, E., Kamphuis, C., De Jong, G., 2005. Genetic parameters for claw disorders in Dutch dairy cattle and correlations with conformation traits. Journal of dairy science 88, 3672-3678.

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. Journal of dairy science 91, 4414-4423.

Velthuis, A., De Jong, M., Kamp, E., Stockhofe, N., Verheijden, J., 2003. Design and analysis of an Actinobacillus pleuropneumoniae transmission experiment. Preventive veterinary medicine 60, 53-68.

Walker, R.L., Read, D.H., Loretz, K.J., Nordhausen, R.W., 1995. Spirochetes isolated from dairy cattle with papillomatous digital dermatitis and interdigital dermatitis. Veterinary Microbiology 47, 343-355.

Wells, S., Garber, L., Wagner, B., 1999. Papillomatous digital dermatitis and associated risk factors in US dairy herds. Preventive Veterinary Medicine 38, 11-24.

Woolhouse, M., Stringer, S., Matthews, L., Hunter, N., Anderson, R., 1998. Epidemiology and control of scrapie within a sheep flock. Proceedings of the Royal Society of London. Series B: Biological Sciences 265, 1205-1210.

4.8 Appendix

4.8.1 Cross-validation

In the twelve-fold cross-validation, we predicted for each susceptible animal of the censored farm in each period the number of cases over the number of susceptible feet (C_i/F_i) based on information of the eleven other farms. In general, the predicted probability for the number of cases over the number of susceptible feet can be calculated with the estimated effects. Because of the complementary log-log link function, these effects need to be back calculated to the original scale,

$$\widehat{P}_i(t) = 1 - e^{-e^{\sum \widehat{FE} + \widehat{\log(\gamma_i)} + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right)}}, \qquad \text{(Equation A4.1)}$$

when the genetic effect for infectivity was not included (model 1 and 3), and

$$\widehat{P}_i(t) = 1 - e^{-e^{\sum \widehat{FE} + \widehat{\log(\gamma_i)} + \sum_j \widehat{\log(\varphi_j)} + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right)}}, \qquad \text{(Equation A4.2)}$$

when the genetic effect for infectivity was included (model 2 and 4). In Equations A4.1 and A4.2, $\widehat{P}_i(t)$ is the predicted probability for the number of cases over the number of susceptible feet for animal *i* in the period from *t* to *t+1*, $\sum FE$ is the sum of the estimates for the fixed effects, *i.e.*, the estimates for the intercept, farm *k*, period *t*, parity *l*, and months in milk. The $\widehat{\log(\gamma_i)}$ is the estimated genetic effect for susceptibility on the log scale and $\sum_j \widehat{\log(\varphi_j)}$ is the sum of the estimated genetic effects for infectivity of the infectious group mates of the susceptible focal individual on the log scale. The last term in Equations A4.1 and A4.2 is the offset. In the twelve-fold cross-validation, the cases (*C*) of one entire farm were censored from the dataset, therefore, the random effects for the interaction between farm and period, and non-genetic animal effect for animal *i* could not be estimated for this farm. So, these random effects did not contribute to the predicted probabilities, they are, therefore, not included in Equations A4.1 and A4.2.

To validate the estimated genetic effects, we wanted to estimate the probability that an animal would be a case during an interval independent of the fixed effects in the model. We achieved this independence by standardizing both $\widehat{P}_i(t)$ and $C_i(t)/F_i(t)$. We obtained the regression coefficients for the fixed effects for each of the four models that were applied to the full dataset were no data was censored. With these regression coefficients we calculated the average value of summed fixed effects $\left(\overline{\sum FE}\right)$. The $\overline{\sum FE}$ can be interpreted as a standard/average farm and a

standard/average cow. This $\overline{\sum FE}$ was used in Equations A4.1 and A4.2 instead of the estimated fixed effects,

$$\widehat{P}_i(t)^* = 1 - e^{-e^{\overline{\sum FE} + \widehat{\log(\gamma_i)} + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right)}}, \quad \text{(Equation A4.3)}$$

for model 1 and 3, and

$$\widehat{P}_i(t)^* = 1 - e^{-e^{\overline{\sum FE} + \widehat{\log(\gamma_i)} + \sum_j \widehat{\log(\varphi_j)} + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right)}}, \quad \text{(Equation A4.4)}$$

for model 2 and 4. Here, $\widehat{P}_i(t)^*$ is the predicted probability for the number of cases over the number of susceptible feet for animal *i* in a period, as if it were an average cow with an average parity and months in milk, during a standard period, on a standard farm. Note that the genetic susceptibility and infectivity did differ between cows, and thus had an effect on the predictions.

Similarly, we wanted to standardize the observed cases over the number of susceptible claws $(C_i(t)/F_i(t))$, so that they would be independent of the fixed effects that contributed to that observation. The observations were transformed to the complementary log-log scale so that they were linear in the effects,

$$\log(-\log\left(1 - {C_i(t)}/{F_i(t)}\right) =$$

$$\sum FE + \log(\gamma_i) + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right), \quad \text{(Equation A4.5)}$$

for model 1 and 3, and,

$$\log(-\log\left(1 - {C_i(t)}/{F_i(t)}\right) =$$

$$\sum FE + \log(\gamma_i) + \sum_j \log(\varphi_j) + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right), \quad \text{(Equation A4.6)}$$

for model 2 and 4.

Next, the summed fixed effects $\sum FE$ in Equation A4.5 and A4.6 were replaced with the average value of the summed fixed effects $(\overline{FE})$, and back calculated to the original scale to obtain the observed number of cases over the number of susceptible feet independent of the fixed effects $((C_i(t)/F_i(t))^*)$, for all models,

$$(C_i(t)/F_i(t))^* = 1 - e^{-e^{\left(\left(\log\left(-log\left(1-{}^{C_i(t)}/_{F_i(t)}\right)\right)\right)-\sum FE+\overline{\sum FE}\right)}}. \quad \text{(Equation A4.7)}$$

Here, $(C_i(t)/F_i(t))^*$ is the observed number of cases over the number of susceptible feet for animal *i* in a period, as if observed on an average cow with an average parity and months in milk, on a standard farm. Again, the genetic susceptibility and infectivity did differ between cows (see Equations A4.5 and A4.6), and affected the dependent variable.

4.8.2 Correlation between observations and predictions

We calculated weighted correlation coefficients between the average "corrected" observed number of cases over the number of susceptible feet ($(\overline{C_i(t)/F_i(t))^*}$) and the average "corrected" predicted probabilities ($\overline{\widehat{P}_i(t)^*}$). The summed corrected observations and predictions were averaged over the number of times an animal was susceptible at the start of an interval. The number of times an animal was susceptible was used as the weight.

5

A genome-wide association study for susceptibility and infectivity of dairy cattle to Digital Dermatitis

Floor Biemans[1,2], Mart C.M. de Jong[1], Piter Bijma[2]

[1] Quantitative Veterinary Epidemiology Group, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands

[2] Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands

In preparation

Abstract

One approach to fight transmission of infectious diseases is selection and breeding for livestock with desirable traits that reduce disease prevalence. Disease prevalence in a population depends on the susceptibility and infectivity of the individuals. Knowledge of the genetic background of those traits would, therefore, facilitate efficient selection for lower prevalence. Here we investigate the genetic background of host susceptibility and infectivity for Digital Dermatitis (DD), an infectious claw disease in dairy cattle, using a genome-wide association study (GWAS).

We obtained phenotype data on 1513 Holstein-Friesian cows of twelve Dutch dairy farms. The cows were scored for DD disease status and class (M0 to M4.1) every two weeks for eleven times. The genotype data consisted of 75,904 SNPs for 1401 of the phenotyped cows. We performed a GWAS with two models. First, a linear mixed model that associated SNPs with host susceptibility only. With this model we estimated genetic parameters for ten different DD-related host traits. Second, we used a generalized linear mixed model (GLMM) that associated SNPs with both host susceptibility and infectivity while taking the variation in exposure of susceptible cows to infectious herd mates into account.

For the linear model there were no significant SNPs (FDR < 0.05), but there were 135 suggestive SNPs (FDR < 0.30) for eight traits: for the presence of active lesion in the observation period; the fraction of observations a cow had an active lesion; the fraction of observations a cow had a M0, M1, M2, M4, or M4.1 on at least one claw; and the fraction of observations a cow was DD-free. Heritability estimates ranged from 0.093 to 0.367. For the GLMM there were no significant and no suggestive SNPs. SNP effects on susceptibility of the linear model had a correlation coefficient of only 0.70 with SNP effects on susceptibility of the GLMM, indicating that both models capture partly different effects.

5.1 Introduction

One approach to fight transmission of infectious diseases is selection and breeding for livestock with desirable traits that affect disease prevalence. Disease transmission is affected by two sets of host traits, those affecting susceptibility and those affecting infectivity. Susceptibility is the relative risk that an individual gets infected when exposed to a typical infectious individual, or to infectious material excreted by a typical infectious individual. Infectivity is the relative propensity of an infected individual to infect a typical susceptible individual.

In livestock genetic improvement, studies on genetic effects related to infectious diseases generally link the disease status of the host to the genotype of the host (Woolhouse et al., 1998; Springbett et al., 2003). Thereby they capture the genetic effects on susceptibility only, while the variation in exposure of susceptible individuals to infectious group mates and the variation in infectivity among those group mates is ignored. The infectivity of infected group mates, however, may contain a heritable component. In other words, infectivity may be an indirect genetic effect (Moore et al., 1997; Anche et al., 2014), *i.e.*, a heritable effect of an individual on the phenotype of another individual (Muir, 2005). When infectivity contains a heritable component, selection for lower infectivity can also be used to improve populations by selective breeding (Lipschutz-Powell et al., 2012; Anche et al., 2014; Anacleto et al., 2015).

Here we focus on host susceptibility and infectivity for Digital Dermatitis (DD), an infectious claw disease in dairy cattle. In infected cattle, round lesions that are sometimes painful form along the coronary band of the claws (Walker et al., 1995). Mainly the hind claws are affected by the disease (Read and Walker, 1998; Sogstad et al., 2005). Digital Dermatitis is transmitted via the environment, and the infectious "agent" is a combination of different bacteria (Rodríguez-Lainz et al., 1996; Read and Walker, 1998; Demirkan et al., 1999; Sogstad et al., 2005; Vink et al., 2009). Lesions can be divided into six distinct classes: skin where lesions are macroscopically absent (M0), a small lesion of 0-2 cm (M1), a lesion of >2 cm (M2), a lesion covered by a scab (M3), altered skin with dyskeratotis or surface proliferation (M4), and a small lesion in addition to altered skin (M4.1) (Döpfer et al., 1997; Döpfer, 2009; Berry et al., 2012). Classes M1, M2, and M4.1 are sometimes referred to as the active classes because they describe circumscribed, red-greyish, moist, painful, and prone to bleed lesions (Speijers et al., 2010; Berry et al., 2012; Zinicola et al., 2015).

The genetic variants associated with host susceptibility and infectivity for DD can be detected in a genome wide association study (GWAS) with the use of molecular markers such as SNPs. Results of previous GWAS studies on host susceptibility for DD

are inconsistent. Significant peaks have been detected on BTA26 (Scholey, 2011), on BTA1, 5, 8, 14 and 26 (Malchiodi et al., 2015), not at all (van der Spek et al., 2015), or on BTA3, 8, and 29 (Oberbauer et al., 2016). All these studies used a linear model to find associations between SNPs and host susceptibility. In addition, these studies did not take into account variation in exposure of susceptible individuals to infected herd mates, and variation in infectivity among those herd mates.

Here we perform a GWAS with two different models. The first model is a linear mixed model that associates SNPs with host susceptibility only. With this model we estimate genetic parameters for ten different DD-related host traits. The second model is a generalized linear mixed model (GLMM) that associates SNPs with both host susceptibility and infectivity. In the GLMM, we also take the variation in exposure of the susceptible individuals to infectious herd mates into account.

5.2 Material and methods

5.2.1 Phenotype data

Phenotypes were collected on twelve dairy farms in the Netherlands, between November 2014 and April 2015. Two observers visited the farms eleven times, every other week. One observer rinsed and scored the claws with the method of Relun et al. (2011), and the other observer recorded the cow ID and the DD-status of the cow. All hind claws were scored with the standardized classification developed by Döpfer et al. (1997) and extended by Berry et al. (2012). This classification comprises six distinct classes M0, M1, M2, M3, M4, and M4.1, Where, M0 is skin without macroscopic lesions, M1 is a small lesion of 0-2 cm, M2 is a lesion of >2 cm, M3 is a lesion covered by a scab, M4 is irregular skin with dyskeratotis or surface proliferation, and M4.1 is a small lesion (M1) in addition to irregular skin (M4). A claw scored as M0 was classified as susceptible, and a claw scored as M1, M2, M3, M4, or M4.1 was classified as infected and infectious.

Farmers were allowed to identify and treat lesions but were not informed on the DD status of the cows by the observers. Table 5.1 gives an overview of some characteristics of the farms enrolled in the study. Phenotypes were collected on 1513 cows, of which 1401 cows were also genotyped. The average number of observations per cow was 8.7, because some cows were removed from, or introduced into, the herd during the study.

Table 5.1. Characteristics of the farms enrolled in the study.

Farm	# Cows examined[1]	Cows genotyped[1]	# Observations[2]	Average Δt (days)	Prevalence (SD)[3]	
					Cow level	Foot level
A	134	116	11	14	78.0 (5.4)	69.6 (6.6)
B	105	101	11	14	56.3 (7.5)	46.9 (7.9)
C	159	162	11	14	49.7 (2.8)	40.2 (1.9)
D	118	116	11	14	57.8 (5.0)	49.2 (5.1)
E	102	90	11	13.6	62.8 (5.0)	54.6 (5.4)
F	133	112	10	15.56	59.2 (10.0)	48.7 (10.4)
G	100	98	11	14	65.6 (8.1)	58.2 (7.6)
H	189	180	11	14	64.9 (6.2)	56.7 (5.8)
I	104	75	11	14	56.4 (5.1)	45.6 (4.9)
J	88	88	11	14	65.8 (10.8)	58.1 (10.9)
K	130	116	9	14	63.6 (9.6)	52.5 (8.5)
L	151	147	11	13.9	70.9 (7.2)	62.0 (7.7)
Total	1513	1401	129	14.07	62.6 (7.5)	53.9 (11.0)

[1] Total number of different cows on a farm.

[2] Total number of farm visits.

[3] Average percentage scored as infected (class M1, M2, M3, M4, or M4.1) with the standard deviation (SD).

5.2.2 Genotype data

The cows (Holstein-Friesian) were genotyped with the Eurogenomics 10K chip. In the final analysis only cows with a call rate of >0.85 were included (n = 1401). Quality control was performed on the data. A genome-wide marker (SNP) was included only when the following criteria were met: 1) observed frequency that deviated <0.15 from expected Hardy Weinberg equilibrium frequency; 2) minor allele frequency >0.025. Furthermore, inconsistent genotypes between parents and offspring were set to missing. The SNPs that passed the quality control were imputed to a set of 76438 SNPs based on the Illumina BovineSNP50 chip and a custom chip from breeding company CRV. Both chips had a reference population of >1000 animals with genotypes. After imputation, quality control was performed on the imputed data. A SNP was included in final analyses only when the following criteria were met: 1) no strong deviation from Hardy Weinberg equilibrium (p-value $>1*10^{-15}$); 2) missing rate <0.05; 3) minor allele frequency >0.02. In total 75904 SNPs passed the quality control and were included in the final analysis.

5.2.3 Traits

With the linear model, we analysed five DD-related host traits (Table 5.2) that reflect only the susceptibility of cows. The first trait was disease status (0 = both claws susceptible, 1 = one claw susceptible, one claw infected, 2 = both claws infected) of each cow at each observation *k* (*k* = 1 to 11). A cow was infected when at least one claw received a score different from M0. The second trait was disease status for active lesions (0 = no active lesions observed (only M0, M3, or M4 lesions), 1 = at least one active lesion (M1, M2, or M4.1) observed) during the entire observation period. The third trait was the fraction of observations that a cow had at least one active lesion ($Fraction_{active}$). We chose the fraction, rather than the number, of observations because not all cows were scored the same number of times. The fourth trait was the fraction of observations a cow had a lesion of class M*i* ($Fraction_{Mi}$ with *i* = 0, 1, 2, 3, 4, or 4.1). The fifth trait was the fraction of observations both claws were scored as M0, *i.e.*, the fraction of observations a cow was free of DD ($Fraction_{free}$). Table 5.3 illustrates the DD-related host traits that were calculated from the M-class scores that a cow received at each observation.

Table 5.2. DD-related host traits that were analysed with the linear model.

Trait		Dependent variable	# records per cow
1.	Disease status	0 = both claws susceptible 1 = one claw susceptible, one claw infected 2 = both claws infected	11[1]
2.	Active lesion observed	0 = no active lesions observed 1 = at least one active lesion observed	1
3.	Fraction of observations with an active lesion[2]	$Fraction_{active} = \frac{\#\ observations\ with\ active\ lesion\ on\ at\ least\ one\ claw}{\#\ observation\ total}$	1
4.	Fraction of observations with a lesion of class M*i*[3]	$Fraction_{Mi} = \frac{\#\ observations\ with\ class\ Mi\ on\ at\ least\ one\ claw}{\#\ observations\ total}$	1
5.	Fraction of observations with M0 on both claws	$Fraction_{free} = \frac{\#\ observations\ with\ class\ M0\ on\ both\ claws}{\#\ observations\ totatl}$	1

[1] Cows had a maximum of eleven records.

[2] Class M1, M2, and M4.1 are active lesions.

[3] *i* = 0, 1, 2, 3, 4, or 4.1.

Table 5.3. Example of M-class scores that a cow received at each observation, and the DD-related host traits that were calculated from these scores.

Cow ID	Farm ID	Hind claw	M-class at observation											Total number of observations
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	
2047	L	Left	4	4.1	2	4	2	1	0	4	4	3	NA	10
		Right	3	4	2	0	2	1	0	1	4	3	NA	10

Trait		Dependent variable for cow 2047										
1	*Observation*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
	Disease status	2	2	2	1	2	2	0	2	2	2	NA
2	Active lesion observed = 1											
3	$Fraction_{active}$ = 5/10 = 0.50											
4	$Fraction_{M0}$ = 2/10 = 0.20											
	$Fraction_{M1}$ = 2/10 = 0.20											
	$Fraction_{M2}$ = 2/10 = 0.20											
	$Fraction_{M3}$ = 2/10 = 0.20											
	$Fraction_{M4}$ = 5/10 = 0.50											
	$Fraction_{M4.1}$ = 1/10 = 0.10											
5	$Fraction_{free}$ = 1/10 = 0.10											

With the GLMM we analysed whether the susceptible claws of a cow had become infected during an observation interval. Only the hind claws of the cows were scored, so a susceptible cow had one or two susceptible claws at the start of an interval (F = 1 or 2), that were zero, one, or two cases by the end of the interval (C = 0, 1 or 2). The dependent variable was the number of cases over the number of susceptible claws, C/F, for each cow in each observation interval. The dataset for the GLMM consisted of 6099 cows that were susceptible at the start of an interval.

5.2.4 Linear Model

With the linear model, we investigated the association between SNPs and each of the five traits (Table 5.2),

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}, \qquad \text{(Model 1)}$$

where $\boldsymbol{y}$ is a vector of observations on a trait, **X** is the incidence matrix for the fixed effects; **b** is a vector of fixed effects for SNP-genotype (0, 1, or 2), farm (A to L), and parity (1, 2, or >2); **Z** is the incidence matrix for the additive genetic effects; and **a** is

a vector of additive genetic effects, with $\mathbf{a} \sim \mathrm{N}(0, \mathbf{G}\sigma_{\mathrm{a}}^2)$, where **G** is the additive genetic relationship matrix of the 1401 genotyped animals and σ_a^2 is the additive genetic variance. For the first trait, disease status, we also included fixed effects for observation (1 to 11) and the interaction between farm and observation; and a random cow effect to account for repeated observations on a cow. Months in milk was not significant and therefore not included in the model.

5.2.5 Generalized linear mixed model (susceptibility and infectivity)

With the GLMM, we investigated the association between SNPs and host susceptibility and infectivity for DD. In contrast to the linear model, this GLMM is founded in epidemiological principles. In the GLMM, the variation in exposure of a susceptible individual to the infectious individuals, and the (genetic) variation in infectivity of the infectious individuals was taken into account. The use of a GLMM to analyse binary data on disease status has previously been described in, for example, Velthuis et al. (2003), Anche et al. (2015), and Biemans et al. (2017b).

To define the GLMM, we need to find the probability that a susceptible claw becomes infected during an interval. To find this probability, we modelled DD-transmission with a stochastic compartmental susceptible-infected-susceptible-model (SIS-model) with an environment route E (see *e.g.*, de Rueda et al., 2015),

In this model, two events can occur; a susceptible claw can get infected, and an infected claw can recover. These events occur randomly with a probability per unit of time depending on the model parameters, the number of infectious claws, and the infection pressure coming from the environment. The expected rate with which susceptible claws get infected is

$$\beta S \frac{E+I}{N},$$

where S the number of susceptible claws, I is the number of infectious claws, and *N* the total number of claws in a group (twice the number of cows), so that $S + I = N$. The *E* is the infection pressure coming from the environment, expressed in "currently infected individual equivalents", so that *E* + *I* represents the total infection pressure expressed as the equivalent number of currently infected individuals. The β is the

transmission rate parameter, a population specific parameter that contains information on the contact rate, and on the susceptibility and infectivity of the population.

To model genetic variation in susceptibility and infectivity between individuals, we consider the pairwise transmission rate parameter β_{ij} between a susceptible claw of (focal) individual *i* and an infectious claw of its herd mate *j*. The transmission rate parameter β_{ij} from a single infectious claw of individual *j* with infectivity φ_j to one susceptible claw of individual *i* with susceptibility γ_i is:

$$\beta_{ij} = c\gamma_i\varphi_j, \qquad \text{(Equation 5.1)}$$

where *c* is the average contact rate. In this expression, average susceptibility and infectivity are defined to be one, $\bar{\gamma} = \bar{\varphi} = 1$, so that *c* represents the transmission rate parameters for a typical pair of individuals *i* and *j*. In other words, *c* refers to an *effective* contact rate, not to a physical contact rate.

When a claw of *i* is exposed to all infectious claws, the expected rate of infection for this claw depends on the susceptibility of *i*, on the number of infectious claws in the group, and on their average infectivity. The total transmission rate equals the sum of the rates due to each infectious claw,

$$Transmission\ rate_i = c\ \gamma_i \frac{\sum_j \varphi_j}{I_g} \frac{E + I_{tot}}{N}. \qquad \text{(Equation 5.2)}$$

Because not all cows were genotyped, we distinguished between the number of claws of infectious individuals that are genotyped, I_g, and the number of claws of infectious individuals in total, I_{tot}, in Equation 5.2. The total number of infectious claws includes the claws from both genotyped and non-genotyped infectious individuals, $I_{tot} \geq I_g$. We made the distinction because differences in infectivity could only be estimated for the genotyped infectious individuals, while in the transmission rate we wanted to account for the non-genotyped infectious individuals as well. Therefore, all infectious individuals are included in the last term of Equation 5.2, while the summation term, $\sum_j \varphi_j / I_g$, was calculated over claws of genotyped individuals only, and thus represents the average infectivity of the claws of genotyped infectious individuals.

The probability that a susceptible claw becomes infected in an interval depends on the number of infectious claws of herd mates at the start (*t*) of the interval, and on their average infectivity. This probability follows from assuming a Poisson process

within the interval (Δt), and is the probability of a non-zero outcome from a Poisson distribution. This probability follows from Equation 5.2 as,

$$P_i(t) = 1 - e^{-c\gamma_i\left(\frac{\sum_j \varphi_j}{I_g(t)}\right)\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t}, \qquad \text{(Equation 5.3)}$$

where, $P_i(t)$ is the probability that a claw of individual *i* becomes infected (is a case) in an interval *Δt*, and *t* indicates the start of the interval. Thus, the number of cases, $C_i(t)$ = 0, 1, or 2, follows a binomial distribution, with binomial total $F_i(t)$ = 1 or 2, and probability given by Equation 5.3.

Because the probability of being a case follows a Poisson process, the complementary log-log is the appropriate link function to connect the explanatory variables to the expected probability to become infected (Equation 5.3),

$$\text{cloglog}(P_i(t)) = \log(c) + \log(\gamma_i) + \log\left(\frac{\sum_j \varphi_j}{I_g(t)}\right) + \log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right). \qquad \text{(Equation 5.4)}$$

The last term $\log\left(\frac{E(t)+I_{tot}(t)}{N(t)}\Delta t\right)$ is the offset, *i.e.*, an "explanatory variable" with a fixed regression coefficient of 1. The offset accounts for the infection pressure coming from the environment at *t*, $E(t)/N(t)$, the fraction of infectious claws at *t*, $I_{tot}(t)/N(t)$, and the length of the interval, *Δt*.

Details of the calculation of the infection pressure coming from the environment are described in Biemans et al. (2017a). In short, claws that are currently infected ($I_{tot}(t)$) contributed fully to the current environmental reservoir, while claws that were infected at an earlier stage contributed partly to the current environmental reservoir ($E(t)$). The contribution was assumed to decrease each interval *Δt* with factor *λ*, which may be interpreted as a survival rate of the pathogen in the environment. The survival rate of DD pathogens was estimated to be 0.9 (Biemans et al., 2017a). So, the infection pressure from the environment due to a single claw that is infectious at time t equalled 0.9 at *t*+1, 0.9^2 at *t*+2, 0.9^3 at *t*+3, etc. Thus values for $E(t)$ were calculated as,

$$E(t) = 0.9\,(E(t-1) + I_{tot}(t-1)), \qquad \text{(Equation 5.5)}$$

which is the total contribution of the claws that were infectious at earlier stages to the environmental reservoir at time t.

For the first interval, calculation of Equation 5.5 requires information on the number of infectious claws before observations started. Since this number was unknown, we estimated it with linear regression within herd of the number of infected claws on the time point, where *t* = 1 represents the first observation. The intercept of the model was used as the average number of infectious claws before the first observation ($I_{tot}(t=0)$), and the value for the environmental reservoir before the first observation was estimated as $E(t=0) = \frac{0.9}{1-0.9} I_{tot}(t=0)$ (Biemans et al., 2017a).

Equation 5.4 is not linear in infectivity. To solve this issue, we first moved the number of infectious claws of cows that were genotyped to the offset,

$$\text{cloglog}(P_i(t)) = \log(c) + \log(\gamma_i) + \log\left(\sum_j \varphi_j\right) + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right).$$

Subsequently, we approximated $\log\left(\sum_j \varphi_j\right)$ by $\sum_j \log\left(\varphi_j\right)$, which is equivalent to approximating a geometric mean by the corresponding arithmetic mean (Anche et al., 2015; Biemans et al., 2017b), so that

$$\text{cloglog}(P_i(t)) \approx \log(c) + \log(\gamma_i) + \sum_j \log\left(\varphi_j\right) + \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right).$$

(Equation 5.6)

The final model to test the association between a SNP and the probability of infection was based on Equation 5.6, with additional fixed effects for farm, interval, parity, and months in milk, and random effects for the interaction between farm and interval, and animal. The final GLMM was,

$$\begin{aligned}\text{cloglog}\left(P_{ijklm}(t)\right) = {} & c_0 + Farm_k + Interval_m + Parity_l + c_1 MIM + \\ & c_2 SNP_{SUS,i} + c_3 SNP_{INF,\bar{j}} + Farm_k.Interval_m + \\ & Animal_i + \log(\gamma_i) + \sum_j \log\left(\varphi_j\right) + \\ & \log\left(\frac{E(t)+I_{tot}(t)}{I_g(t)N(t)}\Delta t\right).\end{aligned}$$

(Model 2)

Here, $P_{ijklm}(t)$ is the expectation of the number of cases over the number of susceptible claws of animal *i* within Δt, $E\left(\frac{C_i(t)}{F_i(t)}\right)$. The c_0 is the intercept; $Farm_k$ is a fixed farm effect (*k* = A to L); $Interval_m$ is a fixed interval effect (*m* = 1 to 10); $Parity_l$ is a fixed effect for parity (*l* = 1, 2, or >2); c_1 is the fixed regression coefficient for months in milk (*MIM* a continuous variable); c_2 is fixed the regression coefficients for susceptibility, where $SNP_{SUS,i}$ represents the number of reference alleles of the susceptible cow *i*, and takes values 0, 1, or 2; and c_3 is the fixed regression coefficient for infectivity, where $SNP_{INF,\bar{j}}$ represents the average number of reference alleles of the infectious claws on a farm, and takes real values between 0 and 2 (Biemans et al., 2017b). The random effects were the interaction between farm and interval ($Farm_k.interval_m$ with *k* = A to L and *m* = 1 to 10), a non-genetic animal effect to account for repeated observations ($Animal_i$), an additive polygenic effect for susceptibility of animal *i* ($\log(\gamma_i)$, with $\log(\boldsymbol{\gamma}) \sim \mathrm{N}(\mathbf{0}, \mathbf{G}\sigma^2_{\mathrm{a}})$, where **G** is the genomic relationships matrix among animals), and an additive polygenic effect for infectivity of the infectious group mates *j* of animal *i* ($\sum_j log(\varphi_j)$, with $\log(\boldsymbol{\varphi}) \sim \mathrm{N}(\mathbf{0}, \mathbf{G}\sigma^2_{\mathrm{a}})$). The last term in model 2 is the offset.

5.2.6 Analyses

We fitted the linear model for disease status (Table 5.2, trait 1) with ASReml v4.1.0 (Gilmour et al., 2015). The G-matrix used in ASReml was computed using method 1 of VanRaden (2008), implemented with calc_grm (Calus and Vandenplas, 2016). We fitted the linear models for the other traits (Table 2) with GCTA (Yang et al., 2011) for reasons of computing time. In GCTA, the chromosome on which the candidate SNP was located was excluded from the calculation of the genomic relationship matrix. We fitted the generalized linear mixed model (model 2) with ASReml v4.1.0 (Gilmour, 2015). SNPs were fitted individually in succession, starting with the first SNP on BTA1, following the genome, and ending with the last SNP on BTA29. The significance threshold for both models was adjusted for multiple testing, using the package qvalue (Storey and Tibshirani, 2003) in R v3.4.0 (R Core Team, 2017) to obtain the false discovery rate (FDR). If FDR ≤ 0.30 the association was called suggestive, and if FDR ≤ 0.05 the association was called significant. Manhattan plots and quantile-quantile plots were created using package qqman (Turner, 2014). Phenotypic correlations between traits were calculated in ASReml v4.1.0. A correlation plot of the SNP effects was created with the R package PerformanceAnalytics (Peterson et al., 2014).

5.3 Results

5.3.1 Trait comparison

In Table 5.4 are the mean value and the standard deviation of all traits and the total number of observations in de dataset. Most traits consisted of one observation per cow (1401 observations in total) except disease status and C/F.

Table 5.4. Mean trait value with the standard deviation and the total number of observations in de dataset.

Dependent variable	Mean	Standard deviation	# Observations
Disease status	1.08	0.90	12195
Active lesions observed	0.32	0.47	1401
$Fraction_{active}$	0.09	0.18	1401
$Fraction_{M0}$	0.57	0.41	1401
$Fraction_{M1}$	0.01	0.05	1401
$Fraction_{M2}$	0.06	0.14	1401
$Fraction_{M3}$	0.02	0.07	1401
$Fraction_{M4}$	0.56	0.38	1401
$Fraction_{M4.1}$	0.02	0.07	1401
$Fraction_{free}$	0.39	0.40	1401
C/F[1]	0.21	0.38	6099

[1] Dependent variable in the generalized linear mixed model.

In Table 5.5 are the phenotypic correlations between the traits analysed with the linear model. Mean disease status had a high phenotypic correlation with $fraction_{M0}$ (-0.95), $fraction_{M4}$ (0.93), and $fraction_{free}$ (-0.95). Furthermore, there were high correlations between $fraction_{active}$ and $fraction_{M2}$ (0.88), and between $fraction_{M4}$ and $fraction_{free}$ (-0.95).

Table 5.5. Phenotypic correlations between the traits analysed with the linear model.

	Active lesions observed	**Fraction$_{active}$**	**Fraction$_{M0}$**	**Fraction$_{M1}$**	**Fraction$_{M2}$**	**Fraction$_{M3}$**	**Fraction$_{M4}$**	**Fraction$_{M4.1}$**	**Fraction$_{free}$**
Mean disease status[1]	0.48	0.4	**-0.95**	0.15	0.32	0.3	**0.93**	0.28	**-0.95**
Active lesions observed		0.72	-0.41	0.42	0.54	0.2	0.35	0.47	-0.49
Fraction$_{active}$			-0.33	0.43	**0.88**	0.14	0.21	0.5	-0.42
Fraction$_{M0}$				-0.11	-0.27	-0.29	**-0.82**	-0.26	**0.81**
Fraction$_{M1}$					0.15	0.05	0.08	0.1	-0.18
Fraction$_{M2}$						0.1	0.14	0.14	-0.34
Fraction$_{M3}$							0.17	0.13	-0.29
Fraction$_{M4}$								0.2	**-0.95**
Fraction$_{M4.1}$									-0.27

[1] For comparison of the traits one observation per cow was needed. The first trait consisted of repeated measurements so we calculated the mean disease status of cows over the entire observation period.

5.3.2 Fixed effects

In the linear model there was a significant effect (P < 0.05) for farm, observation, the interaction between farm and observation, and parity. The probability to be infected with DD increased with observation and with parity (Table 5.6).

Table 5.6. Fixed effects estimates from the linear model on disease status (Trait 1, Table 5.2).

Variable		Coefficient[1]	Standard error[1]
Intercept		1.374	0.102
Farm	A	0.000	0.000
	B	-0.483	0.132
	C	-0.346	0.116
	D	-0.373	0.126
	E	-0.310	0.133
	F	-0.464	0.128
	G	-0.339	0.127
	H	-0.017	0.133
	I	-0.158	0.149
	J	-0.319	0.147
	K	-0.419	0.129
	L	-0.136	0.117
Observation	1	0.000	0.000
	2	0.082	0.076
	3	0.079	0.077
	4	-0.073	0.076
	5	0.157	0.077
	6	0.287	0.076
	7	0.085	0.075
	8	0.259	0.076
	9	0.148	0.076
	10	0.307	0.076
	11	0.271	0.077
Parity	1	-0.395	0.043
	2	-0.157	0.040
	>2	0.000	0.000

[1] Estimates were averaged over 75904 analyses.

In the GLMM, the farm effect was not significant. The main difference between farms was the number of the infectious individuals, which was accounted for by the offset. There were significant effects (P < 0.05) for interval, parity, and months in milk. The probability of getting infected with DD during an interval increased the first six observations and stabilized thereafter. The transmission rate parameter increased with increasing parity, it was 21% higher for parity 2 cows compared to parity 1 cows, and 69% higher in parity >2 cows compared to parity 1 cows. For months in milk, the transmission rate parameter decreased with 4% as months in milk increased with one. Higher-order effects of months in milk were non-significant.

5.3.3 SNP effects linear model (susceptibility only)

For all traits, the quantile-quantile plots and the allele frequency and SNP effects of the suggestive SNPs are in the Supplementary material (Figure S5.1 and Table S5.1).

In the first analysis, we investigated the association of SNPs with the disease status of each cow at an observation (Table 5.2). Figure 5.1 shows the Manhattan plots of the GWAS for the disease status. There were no significant and no suggestive SNPs.

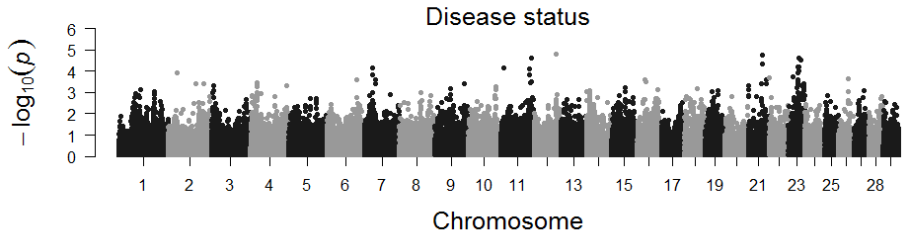


Figure 5.1. Genome Wide Association Study for disease status for Digital Dermatitis (Trait 1, Table 5.2). Plotted is the position on the chromosome in base pairs against the $-\log_{10}$ P-value for each SNP. The data were analysed with a linear model (model 1).

In the second analysis, we investigated the association of SNPs with the presence of an active lesion during the entire observation period (Trait 2, Table 5.2). The top panel of Figure 5.2 shows the Manhattan plot of the GWAS for the presence of an active lesion. There were no significant SNPs, but there were three suggestive SNPs, two on BTA4 and one on BTA21.

In the third analysis, we investigated the association of SNPs with the fraction of observations a cow had an active lesion (Trait 3, Table 5.2). The bottom panel of Figure 5.2 shows the Manhattan plot of the GWAS for $\text{fraction}_{\text{active}}$. There were no significant SNPs but there were two suggestive SNPs, one on BTA1 and one on BTA14.

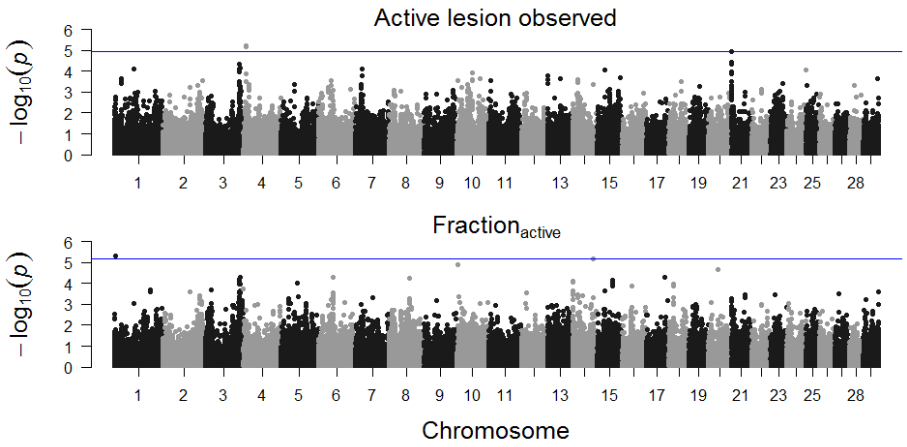


Figure 5.2. Genome Wide Association Study for the presence of an active Digital Dermatitis lesion during the entire observation period (top) and the fraction of observations a cow had an active lesion (bottom) (Trait 3 and 4, Table 5.2). Plotted is the position on the chromosome in base pairs against the $-\log_{10}$ P-value for each SNP. The data were analysed with a linear model (model 1). The false discovery rate was 0.30 for suggestive SNPs (above the blue line).

In the fourth analysis, we investigated the association of SNPs with the fraction of observations a cow was scored with a lesion of class M*i* (*i* = 1, 2, 3, 4, or 4.1) (Trait 4, Table 5.2). Figure 5.3 shows the Manhattan plots of the GWAS for the fractions. There were no significant SNPs but there were suggestive SNPs for all fractions except for M3. For $\text{fraction}_{\text{M0}}$ there were 41 suggestive SNPs, of which 22 were on BTA14. For $\text{fraction}_{\text{M1}}$ there were eight suggestive SNPS, of which six were on BTA2. For $\text{fraction}_{\text{M2}}$ there was one suggestive SNP, for $\text{fraction}_{\text{M4}}$ there were three suggestive SNPs, and for $\text{fraction}_{\text{M4.1}}$ there was one suggestive SNP.

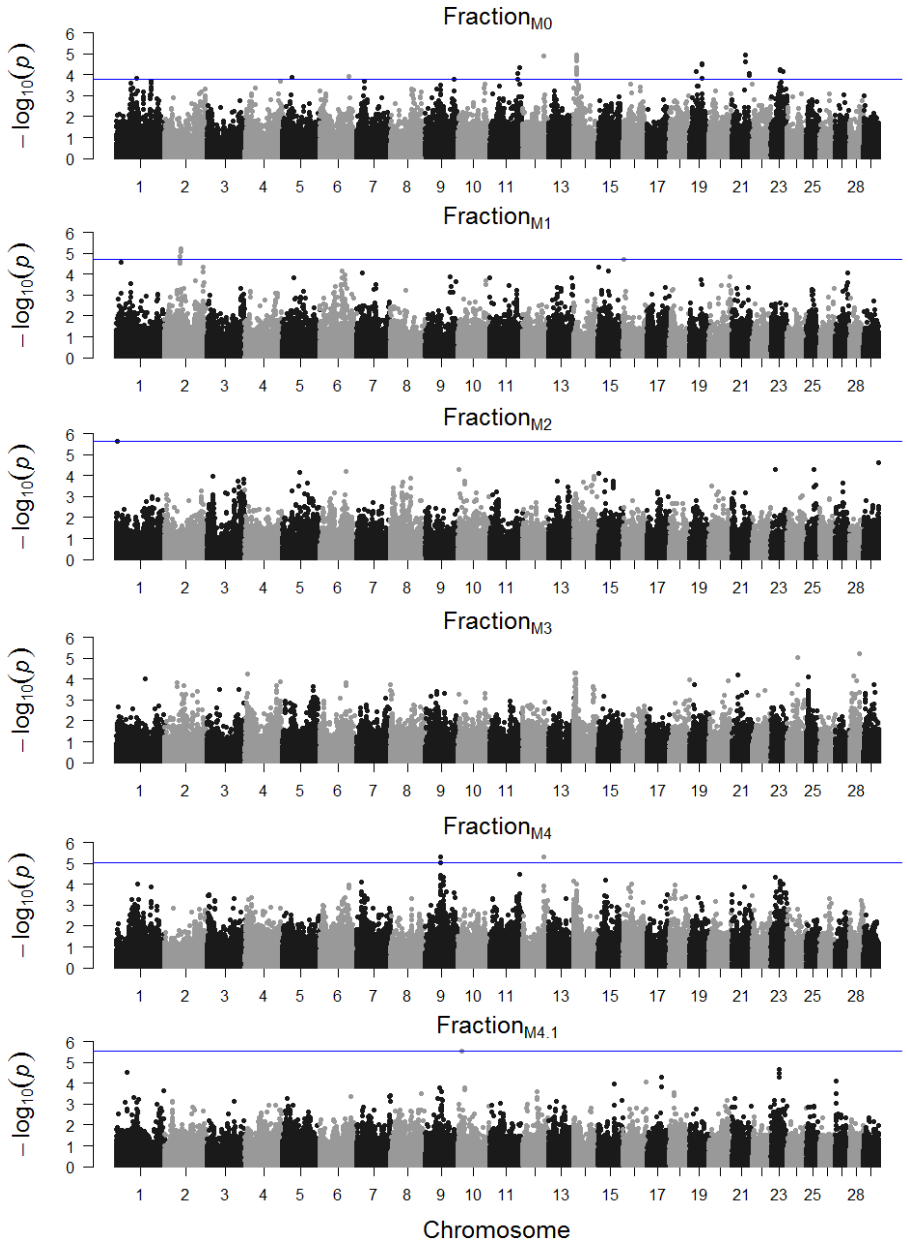


Figure 5.3. Genome Wide Association Study for the fraction of observations a cow received a score of class M0, M1, M2, M3, M4, or M4.1 on at least one claw (Trait 4, Table 5.2). Plotted is the position on the chromosome in base pairs against the $-\log_{10}$

P-value for each SNP. The data were analysed with a linear model (model 1). The false discovery rate was 0.30 for suggestive SNPs (above the blue line).

In the fifth and final analysis, we investigated the association of SNPs with the fraction of observations a cow was susceptible, *i.e.*, scored as M0 on both claws (Trait 5, Table 5.2). Figure 5.4 shows the GWAS for the fraction. There were no significant SNPs but there were 76 suggestive SNPs, of which 16 SNPs were located on BTA23.

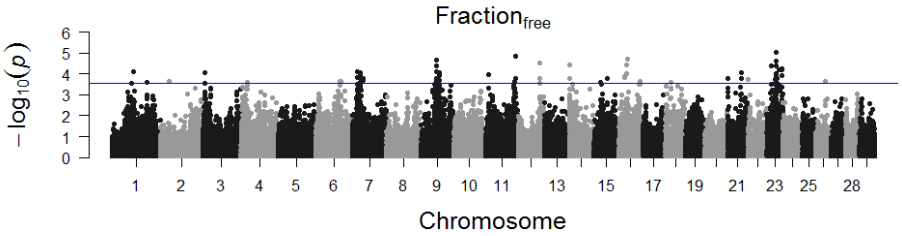


Figure 5.4. Genome Wide Association Study for the fraction of observations a cow was Digital Dermatitis free, *i.e.*, scored as M0 on both claws (Trait 5, Table 5.2). Plotted is the position on the chromosome in base pairs against the $-\log_{10}$ P-value for each SNP. The data were analysed with a linear model (model 1). The false discovery rate was 0.30 for suggestive SNPs (above the blue line).

Eight SNPs had a suggestive association with multiple traits (Table 5.7). Seven of these SNPs were associated with the fraction of the observations a cow was DD-free. Two of these SNPs were also associated with the fraction of observations a cow was scored as M4, and five with the fraction of observations a cow was scored as M0.

5.3.4 Generalized linear mixed model (susceptibility and infectivity)

In the GLMM, SNPs were associated with both host susceptibility and infectivity for DD, taking variation in exposure of susceptible individuals into account. Figure 5.5 shows the GWAS for susceptibility and infectivity. There were no significant and no suggestive SNPs. The $-\log_{10}$ P-values ranged from 0 to 5.11 for susceptibility, and from 0 to 4.35 for infectivity. SNP effects ranged from -0.77 to 0.79 for susceptibility and from -13.84 to 14.5 for infectivity. Overall, the standard error for the effects were large, particularly for infectivity effects.

Table 5.7. Details on SNPs that had a suggestive association with multiple DD-related host traits (Table 5.2).

BTA	BP	MAF	Trait 1	Effect (SE)	% var(A) explained	Trait 2	Effect (SE)	% var(A) explained	Trait 3	Effect (SE)	% var(A) explained
1	5629481	0.067	$Frac_{M2}$	0.0517 (0.0110)	16.99	$Frac_{active}$	0.0612 (0.0134)	12.31	-	-	-
9	51639398	0.339	$Frac_{Free}$	0.0658 (0.0160)	3.65	$Frac_{M4}$	-0.0673 (0.0151)	4.55	-	-	-
9	51639576	0.339	$Frac_{Free}$	0.0681 (0.0161)	3.9	$Frac_{M4}$	-0.0693 (0.0151)	4.82	-	-	-
11	99663236	0.433	$Frac_{Free}$	0.0661 (0.0152)	4.04	$Frac_{M0}$	0.0634 (0.0155)	3.75	-	-	-
12	71831958	0.305	$Frac_{Free}$	-0.0708 (0.0169)	3.99	$Frac_{M0}$	-0.0757 (0.0173)	4.62	$Frac_{M4}$	0.0730 (0.0159)	5.06
23	30712904	0.455	$Frac_{Free}$	0.0635 (0.0155)	3.75	$Frac_{M0}$	0.0637 (0.0158)	3.82	-	-	-
23	31741848	0.41	$Frac_{Free}$	0.0583 (0.0153)	3.09	$Frac_{M0}$	0.0625 (0.0156)	3.59	-	-	-
23	40284986	0.109	$Frac_{Free}$	0.0957 (0.0239)	3.33	$Frac_{M0}$	0.0968 (0.0244)	3.45	-	-	-

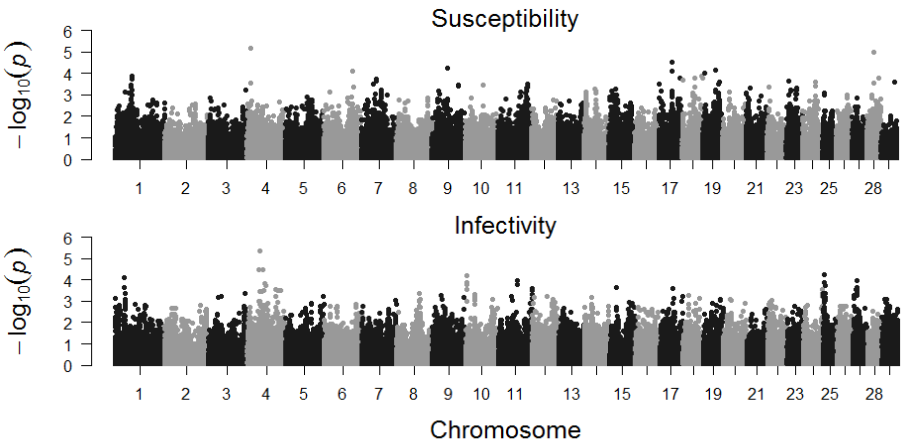


Figure 5.5. Genome Wide Association Study on host susceptibility and host infectivity for Digital Dermatitis. Plotted is the position on the chromosome in base pairs against the $-\log_{10}$ P-value for each SNP. The data were analysed with a generalized linear mixed model (Model 2).

5.3.5 Model and trait comparison

Figure 5.6 shows the correlation matrix of the estimated SNP effects for all traits. The first ten traits were analysed with the linear model and the last two traits with the GLMM. The SNP effects of disease status had a high correlation with fraction_{M0} (-0.94), fraction_{M4} (0.92), and fraction_{free} (-0.94). The SNP effects of disease status analysed with the linear model and susceptibility analysed with the GLMM had a correlation of 0.70. Furthermore, the SNP effects of Fraction_{M4} had a high negative correlation with fraction_{free} (-0.95) and with fraction_{M0} (-0.85), while fraction_{free} had a high positive correlation with fraction_{M0} (0.84). Additionally, fraction_{active} had a moderate to high correlation with fraction_{M1} (0.43), $\text{fraction}_{M4.1}$ (0.48), and fraction_{free} (-0.43), while it had a high correlation with fraction_{M2} (0.89). The SNP effects for susceptibility had a moderately high correlation with disease status, fraction_{M0}, fraction_{M4}, and fraction_{free}, whereas infectivity effects had a low correlation (between -0.14 and 0.12) with all other traits.

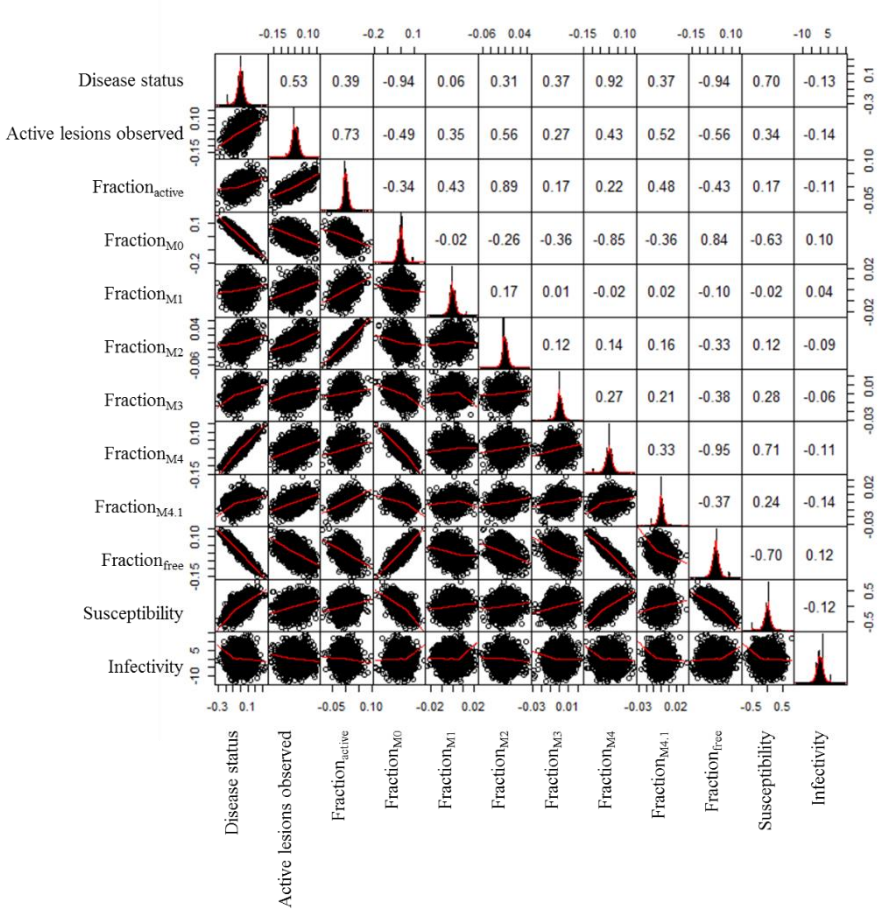


Figure 5.6. Visualization of the correlation matrix between the estimated SNP effects for all traits analysed. The diagonal shows histograms of the SNP effects, above the diagonal are the Pearson correlations coefficients for the SNP effects for different traits, and below the diagonal are bivariate scatterplots with a fitted line. The first ten traits (Table 5.2) were analysed with a linear model (Model 1), the last two traits were analysed with a generalized linear mixed model (Model 2).

5.3.6 Heritabilities

Heritability estimates ranged from 0.093 (fraction$_{m1}$) to 0.367 (fraction$_{free}$) (Table 5.8). The standard error of the estimates was small.

Table 5.8. Estimated heritabilities for the traits analysed with the linear model

Trait	σ_A^2 (SE)	h^2 (SE)
Disease status[1]	0.2073 (0.0015)	0.28 (0.01)
Active lesions observed	0.0368 (0.0107)	0.18 (0.05)
$\text{Fraction}_{\text{active}}$	0.0039 (0.0014)	0.13 (0.05)
$\text{Fraction}_{\text{M0}}$	0.0516 (0.0088)	0.15 (0.05)
$\text{Fraction}_{\text{M1}}$	0.0002 (0.0001)	0.09 (0.04)
$\text{Fraction}_{\text{M2}}$	0.0020 (0.0009)	0.10 (0.04)
$\text{Fraction}_{\text{M3}}$	0.0006 (0.0002)	0.16 (0.05)
$\text{Fraction}_{\text{M4}}$	0.0446 (0.0073)	0.35 (0.05)
$\text{Fraction}_{\text{M4.1}}$	0.0000 (0.0001)	0.01 (0.03)
$\text{Fraction}_{\text{free}}$	0.0532 (0.0085)	0.37 (0.05)

[1] averaged over 75904 analyses with standard deviation between parentheses.

5.4 Discussion

We investigated the genetic background of digital dermatitis using GWAS with two different models. There were 1401 cows included in the analysis. First, we used a linear model to associate SNPs with different host susceptibility traits. There were no significant SNPs (FDR < 0.05), but there were 135 suggestive SNPs (FDR < 0.30) for eight traits: for active lesion observed during the entire observation period, the fraction of observations with an active lesion; the fraction of observations a cow was scored with an M0, M1, M2, M4, or M4.1 on at least one claw; and the fraction of observations a cow was DD-free. Second, we used a generalized linear mixed model to associate SNPs with host susceptibility and infectivity for DD while taking variation in exposure of the susceptible cows into account. There were no significant and no suggestive SNPs in this analysis. The SNP effects for the trait disease status (linear model) had a correlation of 0.70 with the trait susceptibility (GLMM). Heritability estimates ranged from 0.09 for $\text{fraction}_{\text{m1}}$ to 0.37 for $\text{fraction}_{\text{free}}$. Our results suggest that DD is highly polygenic.

5.4.1 Linear model (susceptibility only)

With the linear model we analysed the fraction of observations a cow presented a lesion of class M*i* (*i* = 0, 1, 2, 3, 4, or 4.1). We chose the fraction, rather than the number, of observations a cow presented a certain lesion as dependent variable to account for the number of times a cow was scored. Observations in which a cow was scored with the same lesions on both claws were only counted once. This way of counting lesion classes works well for infected claws, *i.e.*, M1, M2, M3, M4, or M4.1,

because a cow scored with one of these classes is indeed infected at that observation. However, for the class M0 this is not the case. A cow with M0 on one claw and an infectious class on the other claw would count as an observation with M0, so as an observation at which the cow is susceptible, while in reality the cow is also infected. Thus, the fraction of observations with M0 is actually the fraction of observations a cow was not infected on at least one of both claws.

For the fraction of observations scored as M0 on at least one claw, there were 41 suggestive SNPs, of which 22 SNPs were located on BTA14 between base pairs 10119721 and 10293408 (Figure 5.3). In this region, three different genes are located, HHLA1, OC90, and EFR3A (Table 5.9). The HHLA1 gene has orthologs in human and mice, but the function and expression of the gene is unknown in mammals (Kowalski et al., 1999). The OC90 gene has orthologs in human and mice where it causes secretion of an inner ear protein and is a protein component of otoconia (Kowalski et al., 1999). In mice OC90 is expressed in the inner ear and affects the balance (Zhao et al., 2008). The ortholog of the EFR3A gene in human is associated with autism spectrum disorders (Gupta et al., 2014). Based on the orthologs there seems to be no logical direct relation between the genes found in this region and DD in dairy cattle.

For the fraction of observations scored as M0 on at least one claw (Figure 5.3), the SNPs located at base pair 41.646.786 and 41.771.165 of BTA19 had a significant ($P < 0.0005$) association with DD status in a previous study, but that study was based on 47 cows only (Scholey, 2011).

Table 5.9. Genes on BTA14 in the region with suggestive SNPs, between base pairs 10119721 and 10293408, for the fraction of observations a cow received score M0.

Region (BP)	Strand	Gene symbol	Gene description	Gene type
10,106,503-10,130,802	Forward	HHLA1	HERV-H LTR-associating 1	Protein coding
10,106,503-10,127,283	Forward	HHLA1	HERV-H LTR-association 1	Protein coding
10,142,426-10,166,694	Forward	OC90	Otoconin 90	Protein coding
10,171,320-10,250,757	Reverse	EFR3A	EFR3 homolog A	Protein coding

The phenotypic correlation and the correlation of the SNP effects was high for $fraction_{M0}$, $fraction_{M4}$, and $fraction_{free}$. There is a strong relation between these traits because claws were either susceptible (M0) or M4 the majority of the time they were infected (Biemans et al., 2017a). Furthermore, five SNPs that were associated with $fraction_{free}$ were also associated with $fraction_{M0}$. The estimates for these SNPs were approximately the same. Also, two SNPs associated with $fraction_{free}$ were also associated with $fraction_{M4}$. As expected, the estimates of the suggestive SNPs were opposite (positive vs negative) for these traits. A SNP on BTA12 was associated with all of these three traits (Table 5.7). This SNPs is close to muscle blind like splicing regulator 2 (MBNL2) (Bae et al., 2010). Multiple species have this gene. It is a protein coding gene and plays a role in myotonic dystrophy in humans (Carpentier et al., 2014). There seems to be no direct relation with DD in dairy cattle.

5.4.2 Generalized linear mixed model (susceptibility and infectivity)

In the second analysis, we used a generalized linear mixed model (GLMM) to analyse the probability that a cow would get infected in the interval between two observations. The susceptibility of the focal cow and the infectivity of her group mates were in the model as explanatory variables. Furthermore, variation in exposure of susceptible cows to infectious group mates was accounted for in the offset. There were no SNPs associated with host susceptibility and host infectivity (FDR < 0.30).

We assumed that cows get infected via the environment. The total infectious pressure coming from the environment was composed of claws that were infectious at the start of an interval, and of claws of cows that were infectious at an earlier stage. In the offset we accounted for the total infectious pressure from all previous infections observed during the entire experiment. With this offset, the susceptibility estimates were corrected for the total variation in infectious pressure. However, the estimated infectivity effects of SNPs were based only on those claws that were infectious at the start of the interval; we did not consider the genotype of cows that were infectious at an earlier stage. Not considering those genotypes reduces the power to estimate infectivity effects of SNP, especially when the survival of infectious material in the environment is long. For DD, 90% of the infectivity was through claws that were infectious at an earlier stage. The power to associate SNPs with infectivity will increase when the claws of all cows that contribute to the infection pressure are considered. However, in the statistical software we did not manage to keep track of all those infectivity genotypes (of infectious claws and of previous infectious claws via the environment). Future GLMM on DD transmission should be extended to incorporate the infectivity of these claws as well.

5.4.3 Model comparison

We analysed DD-related host traits with a linear model and a GLMM. Heritability estimates for DD-related host susceptibility traits ranged from 0.09 ($fraction_{m1}$) to 0.37 ($fraction_{free}$). Except for the heritability estimate for $fraction_{free}$, these estimates fall in the same range as heritability estimates for host susceptibility for Digital Dermatitis from previous studies (0.04 to 0.29), depending on the model used (Van der Waaij et al., 2005; van der Spek et al., 2013; Schöpke et al., 2015) . We only estimated heritabilities for the traits analysed with the linear model. Estimating heritabilities for traits analysed with a GLMM is not trivial because of the link function. With a complementary loglog link function, the estimated genetic variance components are on a log-scale. Back transforming these estimates to the original scale is possible (*e.g*., it has been done for GLMM with a logit or a probit link function (Roehe and Kalm, 2000)) but the transformation for a complementary loglog link function needs to be derived before it can be applied to the estimates in this study.

Susceptibility effects of SNPs from the GLMM showed a moderately high correlation with those for disease status and $fraction_{M4}$ from the linear model (~0.70; Figure 5.6). However, this correlation was clearly smaller than correlations between SNP-effects of similar traits analysed with the linear model (*e.g.,* 0.92 between disease status and $fraction_{M4}$ in Figure 5.6). This difference suggests that the GLMM captures partly different information than the linear model. One such difference is that some of the traits analysed with the linear model may also have captured genetic variation in the duration of the infectious period (*e.g.,* disease status and $fraction_{M4}$), which is not captured by the GLMM.

Suggestive SNPs were found with the linear model, but not with the GLMM. Such SNP-effects may reflect true genetic variation, for example in the duration of the infectious period, but may also result from violation of model assumptions. The linear model analyses assumed a Gaussian distribution of residuals, which was clearly not the case. This may have caused the marginal inflation of P-values with the linear model, whereas P-values from the GLMM showed no inflation at all (see QQ-plots in Supplementary Material).

The GLMM allowed us to correct susceptibility estimates for variation in exposure and to include variation in infectivity of the group mates. Anacleto et al. (2015) showed that estimates for susceptibility are less accurate when genetic variation in infectivity is not accounted for. Therefore, even when infectivity is not the trait of interest, it might be beneficial to include infectivity in the model to accurately estimate susceptibility (Anacleto et al., 2015). However, the GLMM needs to be extended to better account for the full genetic variation in infectivity via the environment.

5.5 Conclusion

We associated SNPs with different traits related to DD susceptibility and infectivity of the host. We identified 135 suggestive SNPs with a linear model for DD-related host traits on 20 chromosomes. We used a generalized linear mixed model to identify SNPs for host susceptibility and infectivity, but did not find significant or suggestive associations. In contrast to the linear model, in the GLMM variation in exposure of the susceptible cow to infectious group mates was accounted for, and genetic variation infectivity was estimated as well. The SNP effects for the trait disease status (linear model) had a substantial correlation (0.70) with the trait susceptibility (GLMM). Heritability estimates ranged from 0.09 for fraction_{m1} to 0.37 for fraction_{free}.

5.6 Acknowledgements

We thank Natasja Boots for her help with the phenotype collection, the farmers for their participation, and breeding company CRV for providing the genotype data. This work is part of the research programme “Meer met Minder” with project number 847.13.004, which is financed by the Netherlands Organisation for Scientific Research (NWO).

5.7 References

Anacleto, O., Garcia-Cortés, L.A., Lipschutz-Powell, D., Woolliams, J.A., Doeschl-Wilson, A.B., 2015. A novel statistical model to estimate host genetic effects affecting disease transmission. Genetics 201, 871-884.

Anche, M., de Jong, M., Bijma, P., 2014. On the definition and utilization of heritable variation among hosts in reproduction ratio R0 for infectious diseases. Heredity 113, 364-374.

Anche, M.T., Bijma, P., De Jong, M.C., 2015. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47, 1-15.

Bae, J.S., Cheong, H.S., Kim, L.H., NamGung, S., Park, T.J., Chun, J.-Y., Kim, J.Y., Pasaje, C.F.A., Lee, J.S., Shin, H.D., 2010. Identification of copy number variations and common deletion polymorphisms in cattle. BMC genomics 11, 232.

Berry, S.L., Read, D.H., Famula, T.R., Mongini, A., Döpfer, D., 2012. Long-term observations on the dynamics of bovine digital dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. The Veterinary Journal 193, 654-658.

Biemans, F., Bijma, P., Boots, N.M., de Jong, M.C., 2017a. Digital Dermatitis in dairy cattle: the contribution of different disease classes to transmission. Epidemics In press.

Biemans, F., Jong, M.C., Bijma, P., 2017b. A model to estimate effects of SNPs on host susceptibility and infectivity for an endemic infectious disease. Genetics Selection Evolution 49, 53.

Calus, M.P.L., Vandenplas, J., 2016. Calc_grm – a program to compute pedigree, genomic, and combined relationship matrices, ABGC, Wageningen UR Livestock Research.

Carpentier, C., Ghanem, D., Fernandez-Gomez, F., Jumeau, F., Philippe, J.-V., Freyermuth, F., Labudeck, A., Eddarkaoui, S., Dhaenens, C., Holt, I., 2014. Tau exon 2 responsive elements deregulated in myotonic dystrophy type I are proximal to exon 2 and synergistically regulated by MBNL1 and MBNL2. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1842, 654-664.

de Rueda, C.B., de Jong, M.C., Eblé, P.L., Dekker, A., 2015. Quantification of transmission of foot-and-mouth disease virus caused by an environment contaminated with secretions and excretions from infected calves. Veterinary research 46, 1.

Demirkan, I., Walker, R., Murray, R., Blowey, R., Carter, S., 1999. Serological evidence of spirochaetal infections associated with digital dermatitis in dairy cattle. The Veterinary Journal 157, 69-77.

Döpfer, D., 2009. Digital Dermatitis: The dynamics of digital dermatitis in dairy cattle and the manageable state of disease, CanWest Veterinary Conference, Banff AB, Canada.

Döpfer, D., ter Huurne, A., Cornelisse, J., Van Asten, A., Koopmans, A., Meijer, F., Schukken, Y., Szakall, I., Klee, W., Bosma, R., 1997. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and Campylobacter faecalis. Veterinary Record 140, 620-623.

Gilmour, A., Gogel, B., Cullis, B., Welham, S., Thompson, R., 2015. ASReml user guide release 4.1 structural specification. Hemel Hempstead: VSN International Ltd.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. , 2015. ASReml User Guide Release 4.1 Structural Specification, 4.1 ed. VSN International Ltd, Hemel Hempstead, UK.

Gupta, A.R., Pirruccello, M., Cheng, F., Kang, H.J., Fernandez, T.V., Baskin, J.M., Choi, M., Liu, L., Ercan-Sencicek, A.G., Murdoch, J.D., 2014. Rare deleterious mutations of the gene EFR3A in autism spectrum disorders. Molecular autism 5, 31.

Kowalski, P.E., Freeman, J.D., Mager, D.L., 1999. Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. Genomics 57, 371-379.

Lipschutz-Powell, D., Woolliams, J.A., Bijma, P., Pong-Wong, R., Bermingham, M.L., Doeschl-Wilson, A.B., 2012. Bias, accuracy, and impact of indirect genetic effects in infectious diseases. Frontiers in genetics 3.

Malchiodi, F., Koeck, A., Christen, A., Schenkel, F., Kelton, D., Miglior, F., Network, C.D., 2015. Genetic parameters and genome wide association study of individual hoof lesions in Canadian Holsteins using different contemporary groups.

Moore, A.J., Brodie III, E.D., Wolf, J.B., 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution, 1352-1362.

Muir, W.M., 2005. Incorporation of competitive effects in forest tree or animal breeding programs. Genetics 170, 1247-1259.

Oberbauer, A., Danner, A., Belanger, J., Famula, T., Heguy, J., 2016. 0329 Genome-wide association study for loci associated with digital dermatitis and pododermatitis circumscripta in Holstein cattle. Journal of Animal Science 94, 158-158.

Peterson, B.G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., Lesstel, M., Balkissoon, K., Wuertz, D., 2014. PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.4. 3541.

R Core Team, 2017. R: A Language and Environment for Statistical Computing, 3.4.0 ed. R Foundation for Statistical Computing, Vienna, Austria.

Read, D.H., Walker, R.L., 1998. Papillomatous digital dermatitis (footwarts) in California dairy cattle: clinical and gross pathologic findings. Journal of Veterinary Diagnostic Investigation 10, 67-76.

Relun, A., Guatteo, R., Roussel, P., Bareille, N., 2011. A simple method to score digital dermatitis in dairy cows in the milking parlor. J Dairy Sci 94, 5424-5434.

Rodríguez-Lainz, A., Hird, D.W., Carpenter, T.E., Read, D.H., 1996. Case-control study of papillomatous digital dermatitis in Southern California dairy farms. Preventive Veterinary Medicine 28, 117-131.

Roehe, R., Kalm, E., 2000. Estimation of genetic and environmental risk factors associated with pre-weaning mortality in piglets using generalized linear mixed models. Animal Science 70, 227-240.

Scholey, R.A., 2011. Investigating host genetic susceptibility to bovine digital dermatitis. University of Liverpool.

Schöpke, K., Gomez, A., Dunbar, K., Swalve, H., Döpfer, D., 2015. Investigating the genetic background of bovine digital dermatitis using improved definitions of clinical status. Journal of Dairy Science.

Sogstad, Å., Fjeldaas, T., Østerås, O., Forshell, K.P., 2005. Prevalence of claw lesions in Norwegian dairy cattle housed in tie stalls and free stalls. Preventive veterinary medicine 70, 191-209.

Speijers, M., Baird, L., Finney, G., McBride, J., Kilpatrick, D., Logue, D., O'Connell, N., 2010. Effectiveness of different footbath solutions in the treatment of digital dermatitis in dairy cows. Journal of dairy science 93, 5782-5791.

Springbett, A., MacKenzie, K., Woolliams, J., Bishop, S., 2003. The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165, 1465-1474.

Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100, 9440-9445.

Turner, S.D., 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. BioRxiv, 005165.

van der Spek, D., van Arendonk, J., Bovenhuis, H., 2015. Genome-wide association study for claw disorders and trimming status in dairy cattle. Journal of dairy science 98, 1286-1295.

van der Spek, D., van Arendonk, J., Vallée, A., Bovenhuis, H., 2013. Genetic parameters for claw disorders and the effect of preselecting cows for trimming. Journal of Dairy Science.

Van der Waaij, E., Holzhauer, M., Ellen, E., Kamphuis, C., De Jong, G., 2005. Genetic parameters for claw disorders in Dutch dairy cattle and correlations with conformation traits. Journal of dairy science 88, 3672-3678.

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. Journal of dairy science 91, 4414-4423.

Velthuis, A., De Jong, M., Kamp, E., Stockhofe, N., Verheijden, J., 2003. Design and analysis of an Actinobacillus pleuropneumoniae transmission experiment. Preventive veterinary medicine 60, 53-68.

Vink, W., Jones, G., Johnson, W., Brown, J., Demirkan, I., Carter, S., French, N., 2009. Diagnostic assessment without cut-offs: Application of serology for the modelling of bovine digital dermatitis infection. Preventive veterinary medicine 92, 235-248.

Walker, R.L., Read, D.H., Loretz, K.J., Nordhausen, R.W., 1995. Spirochetes isolated from dairy cattle with papillomatous digital dermatitis and interdigital dermatitis. Veterinary Microbiology 47, 343-355.

Woolhouse, M., Stringer, S., Matthews, L., Hunter, N., Anderson, R., 1998. Epidemiology and control of scrapie within a sheep flock. Proceedings of the Royal Society of London. Series B: Biological Sciences 265, 1205-1210.

Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics 88, 76-82.

Zhao, X., Jones, S.M., Yamoah, E., Lundberg, Y.W., 2008. Otoconin-90 deletion leads to imbalance but normal hearing: a comparison with other otoconia mutants. Neuroscience 153, 289-299.

Zinicola, M., Lima, F., Lima, S., Machado, V., Gomez, M., Döpfer, D., Guard, C., Bicalho, R., 2015. Altered Microbiomes in Bovine Digital Dermatitis Lesions, and the Gut as a Pathogen Reservoir. PloS one 10, e0120504.

5.8 Supplementary material

Table S5.1. Details on suggestive SNPs.

Trait	BTA	BP	MAF	SNP effect	SE SNP effect	Variance explained
Active lesions observed	4	9865958	0.326	-0.0845	0.0187	8.53
	4	9906517	0.340	-0.0849	0.0188	8.78
	21	2534840	0.455	-0.0825	0.0188	9.17
$Fraction_{active}$	1	5629481	0.067	0.0612	0.0134	12.31
	14	67961226	0.030	0.0878	0.0195	11.64
$Fraction_{M0}$	1	69956409	0.195	-0.0721	0.0190	3.10
	5	32338737	0.343	-0.0626	0.0163	3.36
	6	97217326	0.424	0.0616	0.0160	3.53
	9	95836920	0.288	0.0636	0.0169	3.15
	11	93192585	0.164	0.0818	0.0209	3.49
	11	93206531	0.164	0.0787	0.0209	3.23
	11	99663236	0.433	0.0634	0.0155	3.75
	12	71831958	0.305	-0.0757	0.0173	4.62
	14	10119721	0.386	0.0642	0.0157	3.71
	14	10146833	0.408	0.0676	0.0156	4.20
	14	10150707	0.408	0.0669	0.0156	4.11
	14	10167188	0.408	0.0677	0.0156	4.21
	14	10169628	0.407	0.0671	0.0156	4.14
	14	10180119	0.405	0.0680	0.0155	4.24
	14	10187244	0.406	0.0672	0.0156	4.14
	14	10188636	0.408	0.0673	0.0156	4.15
	14	10200713	0.386	0.0642	0.0157	3.71
	14	10206900	0.386	0.0635	0.0157	3.63
	14	10210854	0.396	0.0619	0.0155	3.49
	14	10220237	0.407	0.0671	0.0156	4.14
	14	10222060	0.407	0.0663	0.0156	4.04
	14	10232947	0.408	0.0662	0.0156	4.02
	14	10238596	0.408	0.0674	0.0156	4.17
	14	10244476	0.407	0.0671	0.0156	4.14
	14	10270234	0.359	0.0629	0.0160	3.47

	14	10271079	0.357	0.0634	0.0160	3.51
	14	10281314	0.359	0.0626	0.0160	3.43
	14	10283569	0.359	0.0629	0.0160	3.47
	14	10292310	0.359	0.0629	0.0160	3.46
	14	10293408	0.360	0.0630	0.0160	3.48
	19	21823640	0.349	-0.0676	0.0170	3.95
	19	41646786	0.184	0.0861	0.0206	4.24
	19	41771165	0.185	0.0854	0.0206	4.18
	19	42058156	0.447	-0.0586	0.0154	3.23
	21	47109228	0.049	0.1622	0.0369	4.69
	21	47150404	0.046	0.1620	0.0383	4.36
	21	58737991	0.024	0.1910	0.0493	3.19
	21	58738055	0.024	0.1932	0.0494	3.27
	23	30712904	0.453	0.0637	0.0158	3.82
	23	31741848	0.410	0.0625	0.0156	3.60
	23	40284986	0.109	0.0968	0.0244	3.45
$Fraction_{M1}$	1	18244760	0.106	0.0134	0.0032	14.54
	2	53268617	0.113	0.0133	0.0032	15.32
	2	53268660	0.113	0.0138	0.0032	16.46
	2	53293034	0.112	0.0135	0.0032	15.61
	2	53307187	0.113	0.0132	0.0032	14.99
	2	54236610	0.173	0.0120	0.0026	17.65
	2	54257522	0.174	0.0118	0.0026	17.26
	16	2444029	0.129	0.0126	0.0030	15.42
$Fraction_{M2}$	1	5629481	0.067	0.0517	0.0110	16.99
$Fraction_{M4}$	9	51639398	0.339	-0.0673	0.0151	4.55
	9	51639576	0.339	-0.0693	0.0151	4.82
	12	71831958	0.305	0.0730	0.0159	5.06
$Fraction_{M4.1}$	10	12645355	0.029	0.0344	0.0074	168.68
$Fraction_{free}$	1	64378196	0.093	0.0930	0.0256	2.73
	1	72956333	0.476	-0.0585	0.0148	3.21
	1	1.18E+08	0.049	-0.1315	0.0360	3.01
	2	31211042	0.040	0.1408	0.0381	2.84
	3	9591430	0.479	0.0602	0.0153	3.40
	3	9602457	0.435	0.0601	0.0153	3.34

3	9602819	0.474	0.0554	0.0152	2.87
4	21214710	0.377	0.0557	0.0152	2.74
6	81270446	0.231	0.0664	0.0181	2.95
6	83895649	0.202	0.0686	0.0186	2.85
7	15388000	0.271	-0.0657	0.0166	3.21
7	16790531	0.136	-0.0781	0.0217	2.70
7	16811336	0.144	-0.0808	0.0213	3.02
7	26383794	0.311	-0.0615	0.0162	3.04
7	26383840	0.313	-0.0633	0.0161	3.24
7	26681958	0.091	-0.0937	0.0256	2.72
7	26820260	0.191	-0.0740	0.0192	3.18
7	36831104	0.127	0.0864	0.0233	3.11
7	36832260	0.126	0.0881	0.0234	3.22
9	43713692	0.297	0.0600	0.0167	2.83
9	50889943	0.316	-0.0597	0.0164	2.89
9	50892579	0.316	-0.0606	0.0164	2.98
9	51288088	0.227	-0.0704	0.0178	3.27
9	51309233	0.243	-0.0679	0.0172	3.19
9	51556481	0.403	0.0558	0.0148	2.82
9	51639398	0.339	0.0658	0.0161	3.65
9	51639576	0.339	0.0681	0.0161	3.90
9	60005309	0.295	-0.0623	0.0162	3.03
9	60045256	0.379	-0.0601	0.0153	3.20
9	60350738	0.436	0.0553	0.0153	2.83
9	60376659	0.369	0.0581	0.0156	2.95
11	9746998	0.222	0.0695	0.0180	3.13
11	97783769	0.473	0.0574	0.0155	3.08
11	99265744	0.429	0.0581	0.0154	3.10
11	99663236	0.433	0.0661	0.0152	4.04
12	71831958	0.305	-0.0708	0.0169	3.99
12	71943340	0.456	0.0546	0.0151	2.78
12	71945647	0.456	0.0568	0.0150	3.00
14	931162	0.206	-0.0713	0.0190	3.12
14	996982	0.282	-0.0689	0.0167	3.62
14	11287235	0.460	-0.0557	0.0155	2.90

15	23597899	0.493	0.0545	0.0149	2.79
15	25668779	0.189	-0.0679	0.0188	2.66
15	44432549	0.430	-0.0594	0.0157	3.25
16	16697956	0.259	0.0660	0.0172	3.14
16	16698099	0.257	0.0655	0.0173	3.08
16	24254945	0.302	-0.0712	0.0172	4.01
16	24938091	0.442	-0.0592	0.0152	3.25
16	27691033	0.365	-0.0675	0.0158	3.97
16	66422597	0.310	0.0592	0.0164	2.82
16	70280355	0.157	0.0761	0.0207	2.89
16	70280639	0.158	0.0759	0.0207	2.88
16	70280803	0.160	0.0741	0.0206	2.77
18	18594407	0.046	0.1287	0.0351	2.71
21	2594377	0.386	-0.0619	0.0164	3.42
21	2637648	0.386	-0.0619	0.0164	3.42
21	47109228	0.049	0.1413	0.0360	3.52
21	47150404	0.046	0.1409	0.0375	3.25
22	517975	0.071	0.1141	0.0305	3.21
23	13536059	0.273	-0.0608	0.0169	2.76
23	14634231	0.397	0.0640	0.0156	3.69
23	28788714	0.440	-0.0562	0.0150	2.93
23	28788841	0.222	-0.0637	0.0176	2.63
23	29745302	0.175	-0.0836	0.0198	3.78
23	30712904	0.453	0.0635	0.0155	3.75
23	30958975	0.346	-0.0690	0.0156	4.05
23	31144592	0.377	-0.0613	0.0157	3.32
23	31741848	0.410	0.0583	0.0153	3.09
23	32213476	0.388	-0.0587	0.0153	3.08
23	32759584	0.056	-0.1178	0.0325	2.74
23	32759797	0.056	-0.1214	0.0321	2.95
23	40284986	0.109	0.0957	0.0239	3.33
23	49662342	0.199	-0.0756	0.0188	3.43
23	49664725	0.200	-0.0721	0.0187	3.12
23	49677937	0.200	-0.0751	0.0188	3.38
26	33733727	0.235	0.0679	0.0185	3.12

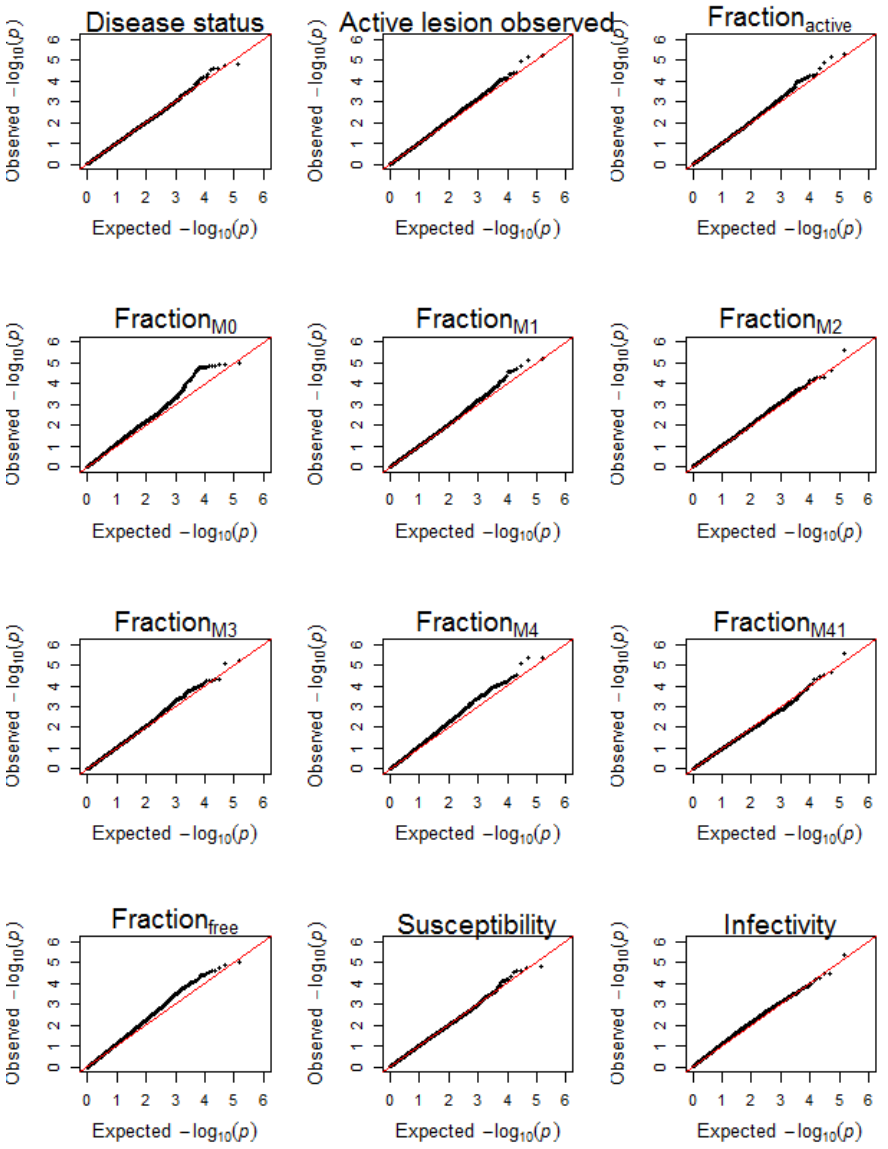


Figure S5.1. Quantile-quantile plots for the p-values from the genome wide association studies for the traits presented in the main text.

6

General discussion

6.1 Introduction

The basic reproduction ratio R_0 is defined as the average number of secondary cases caused by a typical infectious individual in a fully susceptible population (Diekmann et al., 1990). The R_0 contains information on the ability of an infection to establish itself in the population (May and Anderson, 1987). The threshold value is one, if $R_0 < 1$ an infectious individual will infect on average less than one susceptible individual and the disease will die out. If $R_0 > 1$ a major outbreak is possible, and sometimes such a disease persists at an endemic level in a population. Any strategy on reducing the prevalence level in a population, should aim therefore at reducing R_0. Selection and breeding for host individuals with desirable traits with respect to infectious diseases can be combined with other mitigation methods to achieve a reduction in R_0.

Genetic improvement of populations requires estimating breeding values for the individuals in the population, and selecting the top individuals to breed the next generation. The breeding value reflects the genetic potential of an animal to produce offspring superior compared to the current generation. With regard to breeding against infectious diseases, we want to define individual breeding values for R_0. Anche et al. (2014) showed that even though R_0 is property of a population, while a traditional breeding value is property of a single individual, it is possible to define individual breeding values for R_0 based on relative differences in host susceptibility and host infectivity.

Since we cannot directly observe the breeding value of an individual, they need to be estimated based on phenotype data. The quality of the phenotype data affects the accuracy of breeding values. In chapter 2 we quantified the quality of susceptibility and infectivity estimates for different SNP effects and recording intervals based on simulated phenotype data on disease status. We extended the generalized linear model of Anche et al. (2015) and applied it to time series data of an endemic disease. We showed that SNP effects were on average underestimated and therefore conservative. The power to detect SNP effects was high for susceptibility but lower for infectivity. When the total number of observations was limited to eleven, the optimal recording interval in our simulated populations was similar for susceptibility and infectivity, and around 25 to 50% of the length of the average infectious period.

In chapter 3, we applied the generalized linear mixed model to field data on Digital Dermatitis to determine the infectivity of the different disease classes. We estimated the length of the infectious period and the distribution of the first observed classes after infection. With these elements we estimated the contribution

of the different disease classes to R_0. The R_0 for Digital Dermatitis on these farms was estimated to be on average 2.36, to which the chronic state with hyperkeratotic lesions and, sometimes, a proliferative aspect (M4) contributed 88.5%.

In chapter 4, we combined the phenotype data on Digital Dermatitis with genotype data of the same animals to estimate relative differences in host susceptibility and infectivity and individual breeding values for R_0. Genetic variance components for susceptibility and infectivity were estimated with four generalized linear mixed models, while variation in exposure of the susceptible individuals was accounted for via the offset. For each animal, its susceptibility and infectivity relatively to the average cow were estimated. For all models, susceptibility estimates ranged from 0.26 to 3.45, and infectivity estimates ranged from 0.92 to 1.11. The model that included only a genetic effect for susceptibility and no interaction between farm and period had the best fit and showed the least bias, but models with a genetic infectivity effect showed slightly higher accuracy. The susceptibility estimates were used to calculate the individual breeding values for R_0. Estimated breeding values ranged from 0.62 to 6.68, and 2.7% of the cows had a breeding value for $R_0 < 1$.

In chapter 5, we investigated the association between genetic markers (SNPs) and different traits related susceptibility and infectivity for DD. We used a linear model to identify SNPs associated with host susceptibility. The susceptibility traits were disease status at different time points, the fraction of observations a cow was scored with a certain lesion class, and the presence of active lesions. We did not find any significant SNPs, but there were 135 suggestive SNPs located on 20 different chromosomes. Thereafter, we used the generalized linear mixed model to investigate the association between SNPs and susceptibility and infectivity. The susceptibility trait was the effect of the cow on the probability that it got infected over a two-week period. The infectivity trait was the combined effect of the infectious group mates on this probability. We did not find any significant or suggestive associations. The SNP effects for susceptibility in the linear model (the trait disease status) had a substantial correlation (0.70) with the trait susceptibility in the GLMM. Heritability estimates ranged from 0.09 to 0.37 for different susceptibility traits.

These chapters show promising results for lowering DD transmission with selective breeding for R_0. In this discussion, I will focus first on disease traits that were not considered in this thesis. Some traits affect the basic reproduction ratio, like the duration of the infectious period and the “indirect infectivity” of a cow via the environment. Second, I will address breeding against infectious diseases in practice, focussing on the correlation between the estimated breeding value for R_0

and the estimated breeding values for milk production and DD that are currently used. Furthermore, I will address the use of sensor systems for phenotype collection. Finally, I will propose an additional explanation for the lack of power to estimate differences in infectivity.

6.2 Other disease traits

6.2.1 Duration of the infectious period

In this thesis we aimed to develop a method with which individual breeding values for R_0 can be estimated. These breeding values can be used to reduce disease transmission and thus disease prevalence with selective breeding. The parameter that determines the prevalence in the endemic equilibrium is R_0. The R_0 is a function of the transmission rate parameter β and the recovery rate parameter α. The β is a disease specific parameter that is the product of the contact rate and the transmission probability given contact (Roberts and Heesterbeek, 1993). The β depends on the susceptibility genotypes of the susceptible animals γ, the infectivity genotypes of the infectious animals φ, and the average contact rate *c*. In a genetically homogeneous population, all individuals have the same level of susceptibility and infectivity, so there is a single β in the population, $\beta = \gamma\varphi c$. In a genetically heterogeneous population, individuals can have a different level of susceptibility and infectivity, so β may vary between pairs of individuals. Assuming separable mixing, *i.e.*, the susceptibility effect of individuals that are susceptible is independent of the infectivity effect of individuals that are infectious; the average β in the population depends on the average susceptibility and the average infectivity in the population, $\bar{\beta} = \bar{\gamma}\bar{\varphi}c$. The recovery rate parameter α is the probability per unit of time that an infected individual recovers (Diekmann and Heesterbeek, 2000). The α determines the average duration of the infectious period ($\frac{1}{\alpha}$). The R_0 can be calculated as $\mathrm{R}_0 = \frac{\beta}{\alpha}$. Thus, the R_0 can be lowered by lowering (average) susceptibility, infectivity, contact rate, or the duration of the infectious period. In this thesis we focussed solely on (genetic) variation in host susceptibility and host infectivity, and used this to determine variation in R_0. However, also (genetic) variation in the duration of the infectious period may contribute to variation in R_0. Variation in the duration of the infectious period exists for DD. In chapter 3 we showed that the duration of the infectious period depends on the *class-at-infection*. If (part of) this variation is due to genetics, it is an additional source of heritable variation that affects disease transmission.

The duration of the infectious period affects the infectivity of an animal. The longer an animal is infectious, the more susceptible animals it can infect. This effect

on infectivity exists next to infectivity φ that is included in the transmission rate parameter β. To make optimal use of all heritable variation in infectivity, variation in the duration of the infectious period should be taken into account as well.

In the following section I try to estimate the genetic variation in the duration of the infectious period for each cow from the field data, to have an idea about the amount of variation that exists. However, from our dataset it was not possible to determine the duration of the infectious period for each cow directly. Some cows were already infected when data collection started, and the majority of the cows that got infected during data collection was either still infected when data collection ended or removed from the population before a recovery was observed. Since the data are not perfectly fit for estimating individual duration of the infectious period, the following estimates must be interpreted with caution.

We collected the field data over a period of half a year, with a two-week interval between scorings. In total 1513 different cows were observed. Of these cows 220 animals were never scored with an infectious class (M1, M2, M3, M4, or M4.1), 196 animals had an infection on only one foot (the other foot remained uninfected), and the remaining 1097 animals had an infection on both feet (not necessarily simultaneously) at least once. In total 579 cows were scored with an infectious class at every observation. The number of times an animal was scored ranged from one to eleven, with an average of 8.7 scorings. In Table 6.1 are the number of times cows were scored and the fraction of time that cows were scored as infectious. The average fraction of observations a cow was scored as infectious was 0.61. When not taking the 220 cows into account that where never scored as infectious, the average fraction of observations a cow was scored as infectious was 0.71.

I estimated the amount of genetic variation in the duration of the infectious period with ASReml v4.1 (Gilmour et al., 2015), with a simple linear model. The fraction of observations a cow was scored as infectious was the dependent variable,

$$Fraction\ infectious_{ij} = mu + farm_j + Cow_i,$$

where, *Fraction infectious*$_{ij}$ is the fraction of observations that cow *i* on farm *j* was scored as infectious, *mu* is the intercept, *farm*$_j$ is the fixed effect for farm with *j* = A to L, and *Cow*$_i$ is an additive genetic effect for cow *i* ($Cow_i \sim \mathrm{N}(\mathbf{0}, \mathbf{G}\sigma^2_{\mathrm{a}})$, where **G** is the genomic relationships matrix among cows).

The genetic estimate for the fraction of time a cow was infectious relative to an average cow is $\widehat{\alpha_i^{-1}} = \frac{\overline{Fraction\ infectious} + \widehat{Cow_i}}{\overline{Fraction\ infectious}} = \frac{0.71 + \widehat{Cow_i}}{0.71}$, where $\overline{Fraction\ infectious}$ is the average fraction of observations a cow was scored as

infectious without taking the 220 cows that were never infectious into account. The $\widehat{Cow}_i$ is the estimated breeding value for the infectious time of cow i. Estimates for $\widehat{\alpha_i^{-1}}$ ranged from 0.926 to 1.050. With these estimates the breeding value for R_0 was calculated analogous to the calculation in chapter 4,

$$\hat{A}_{R_0,i} = \hat{\gamma}_i \hat{\varphi}_i c \widehat{\alpha_i^{-1}}.$$

With an average R_0 for DD of 2.36 on these farms (chapter 3) and the average product of the estimated relative susceptibility, relative infectivity, and relative infectious time, the value for *c* was calculated. Using the estimates from model 3 in chapter 4 (this model included only a random genetic effect for susceptibility and no interaction between farm and period), $c = \frac{2.36}{\hat{\gamma}_i \hat{\varphi}_i \widehat{\alpha^{-1}}} = 2.178$. Figure 6.1 is a histogram of the estimated breeding values for the basic reproduction ratio for all genotyped animals. Estimated breeding values ranged from 0.59 to 6.66. There were 40 cows with a $\hat{A}_{R_0,i} < 1$ for DD. Three cows that had a $\hat{A}_{R_0,i} > 1$ in chapter 4 now have a $\hat{A}_{R_0,i} < 1$, and one cow that had a $\hat{A}_{R_0,i} < 1$ in chapter 4 now has a $\hat{A}_{R_0,i} > 1$.

Table 6.1. The number of times cows were scored and the fraction of observations cows were scored as infectious.

Fraction	**Times scored**											**Total**
of time infected	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	
0	28	14	13	7	16	10	15	20	17	12	68	220
< 0, 0.1]	-	-	-	-	-	-	-	-	-	18	43	61
< 0.1, 0.2]	-	-	-	-	5	12	18	15	20	10	37	117
< 0.2, 0.3]	-	-	-	3	-	-	7	10	10	10	34	74
< 0.3, 0.4]	-	-	6	-	6	12	-	6	10	10	23	73
< 0.4, 0.5]	-	10	-	7	-	6	5	9	10	9	21	77
< 0.5, 0.6]	-	-	-	-	2	-	10	-	6	4	14	36
< 0.6, 0.7]	-	-	5	-	-	4	-	9	5	6	24	53
< 0.7, 0.8]	-	-	-	1	3	-	11	12	15	8	31	81
< 0.8, 0.9]	-	-	-	-	-	6	11	14	20	16	29	96
< 0.9, 1.0 >	-	-	-	-	-	-	-	-	-	-	46	46
1.0	17	16	18	16	17	25	41	47	73	59	250	579
												1513
Average	0.38	0.52	0.56	0.60	0.47	0.55	0.60	0.60	0.65	0.60	0.64	0.61

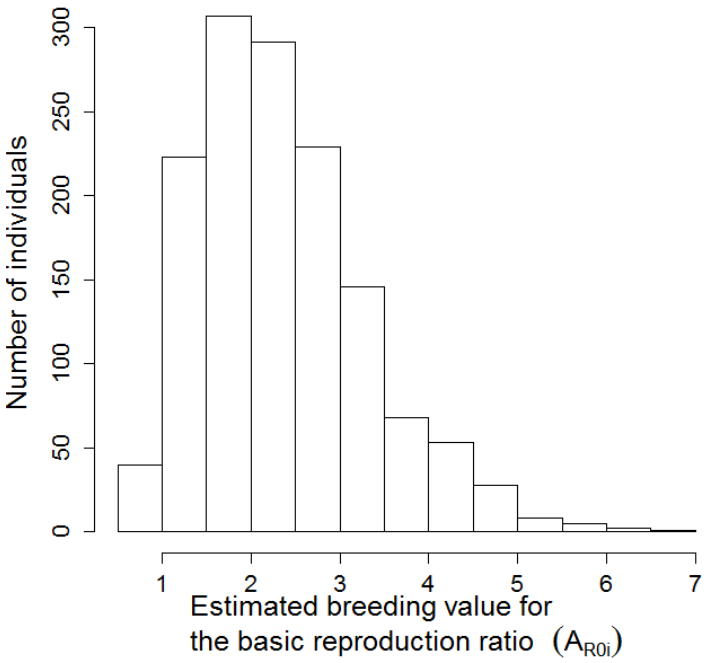


Figure 6.1. Histogram of the individual breeding values for the basic reproduction ratio for all genotyped animals, when genetic estimates for the fraction of time a cow was infectious are taken into account. The relative susceptibility used in the calculation of $\hat{A}_{R_0,i}$ was estimated with Model 3 in chapter 4.

To get insight in the potential genetic variation that exist in duration of the infectious period, I used a linear model with a fixed farm effect to estimate individual values. Including this additional genetic variation had little effect on $\hat{A}_{R_0,i}$. The estimates could improve when additional effects, for example fixed effects for parity and months in milk and a random farm effect, were included in the model. However, more importantly, the dataset should be improved. Now, the estimate for the duration of the infectious time was based on the fraction of observations a cow was scored as infectious. These estimates do not directly reflect the actual time a cow is infectious, especially the estimates for cows that were only observed one or two times. For accurate estimates of the infectious time, the entire infectious period from start to end should be observed for each cow. However, this would be highly unpractical in a field trial because cows need to be scored regularly and for a long

period of time. Automated sensor systems that provide regular data on disease status might be a solution (see below section *Sensor systems*).

Another point related to dataset quality is the scoring accuracy. When diagnostic test are imperfect bias can arise. The quality of a diagnostic test can be assessed by comparing the test with its gold standard, and calculating the sensitivity and specificity of the diagnostic test. Sensitivity (SE) is the ability of a test to correctly identify infected animals, and specificity (SP) is the ability of a test to correctly identify animals that are not infected (Lalkhen and McCluskey, 2008). For SE and SP < 1.00, heritability estimates will be underestimated, and so is the genetic variance (Bishop and Woolliams, 2010). With regard to DD scoring, the milking parlour method of Relun et al. (2011) was compared with scoring in a trimming chute (gold standard). The SE of parlour scoring ranged from 0.79 to 0.93 and the SP ranged from 0.67 to 0.92 for absence (M0) versus presence of DD (M1, M2, M3, or M4) when the lesion was located on the heel bulb (Relun et al., 2011; Solano et al., 2017). However, if lesions were located elsewhere, the SE was below 0.64 (Solano et al., 2017). When looking a specific M-classes, the SE for scoring in the milking parlour is low for class M1, M2, M3, and M4.1 (0 – 0.62), and high for class M4 (0.82) (Solano et al., 2017). Contrary, the SP is high for class M1, M2, M3, and M4.1 (0.96 – 1.00), and moderate for class M4 (0.76) (Solano et al., 2017).

In Chapter 4 and in the calculations on the duration of the infectious period we distinguished claws on absence versus presence of DD. SE and SP estimates for this trait are relatively high but still smaller than one. Therefore, the observed genetic variance might be underestimated. So, the true genetic variance is probably greater, which could result in a higher response to selection than expected from the data.

6.2.2 Infectivity of the environment

In this thesis we assumed that there was an infectious pressure coming from the environment to which susceptible cows were subjected. Claws that were infected at the start of an interval contributed fully to the total infectious, while claws that were infected at an earlier stage could still contribute partly to the total infectious pressure. The contribution was assumed to decrease each interval Δt with factor λ, which was estimated to be 0.9 (Chapter 3). This means that in the endemic equilibrium, the total contribution of claws to infection is $1 + \lambda + \lambda^2 + \lambda^3 + \cdots = \frac{1}{1-\lambda} = \frac{1}{1-0.9} = 10$, to which claws that were infected at an earlier stage contribute $\frac{\lambda}{1-\lambda} = \frac{0.9}{1-0.9} = 9$. Thus, the total infectious pressure is determined for 90% by claws that were infected at an earlier stage and only for 10% by claws that were infected at the start of an interval. However, genetic individual infectivity estimates were

calculated based only on the claws that were infected at the start of an interval, the 10% of the total infectious pressure. By not take into account who contaminated the environment at an earlier stage, we may have disregarded the majority of the heritable variation in infectivity. This has probably no large effect on the quality of the susceptibility estimates, because they are corrected for variation in the total infectious pressure via the offset. However, the quality of the infectivity estimates will improve when all claws that contribute to infection (also those that were infectious earlier) are considered. Hence, the relevance of genetic variation in infectivity may be larger than the estimates presented in chapter 4.

6.2.3 Tolerance, resistance, resilience

Three disease traits that were not considered in this thesis are tolerance, resistance, and resilience. Tolerance is the ability of an animal to maintain performance despite being infected, where infection is the colonization of the host with a pathogen (Gibson and Bishop, 2005). Tolerance is measured as the change in performance following a change in pathogen burden in the animal (Doeschl-Wilson et al., 2012). Selection for tolerance is selection against the side effects of infection that affect the performance of the animal (Rausher, 2001). Resistance, on the other hand, is the ability of a host to resist colonization with a pathogen or to limit the amount of pathogens. Resistance is the opposite of susceptibility, a highly resistant animal has a low susceptibility and vice versa (Knap and Bishop, 2000). Either tolerance or resistance, or a combination of both make an animal resilient (Bisset and Morris, 1996). Resilience is the performance of an animal irrespective of the exact pathogen burden in the animal (Doeschl-Wilson et al., 2012). Selection for resilience is thus selection against the side effects of infection, selection against infection itself, or both. Selection for tolerance does not necessarily reduce disease transmission, while selection for resistance does (Gibson and Bishop, 2005). When defining the breeding goal with regard to infectious diseases it is important to consider whether maintaining performance (select for tolerance), reducing transmission (select for resistance), or both at the same time (select for resilience) is the ultimate goal, and which traits fit this goal best.

6.3 Breeding against infectious diseases in practice

6.3.1 Correlation with other traits

When a disease is rare or can be easily treated or controlled with other measures like adjusting management practises, selective breeding against disease transmission might be uneconomical (Stear et al., 2001). Costs for these type of diseases are relatively low and do not exceed benefits of a high production level.

Therefore, if these disease traits would be included in the breeding goal, they would have a relatively low weight compared to, for example, production traits. However, when a disease is endemic, persistent, and hard to eradicate, it can cost a farmer a lot of time and money especially when these diseases affect productivity of an animal over long periods of time. Furthermore, this type of disease can affect the welfare of an animal for prolonged periods. In this case, selective breeding can be an additional method (next to traditional treatment and control measures) to avoid or reduce cost and improve animal welfare. Whether selective breeding against disease transmission is economically desirable depends on the weight of this trait in the breeding goal and on the correlation of this trait with production traits.

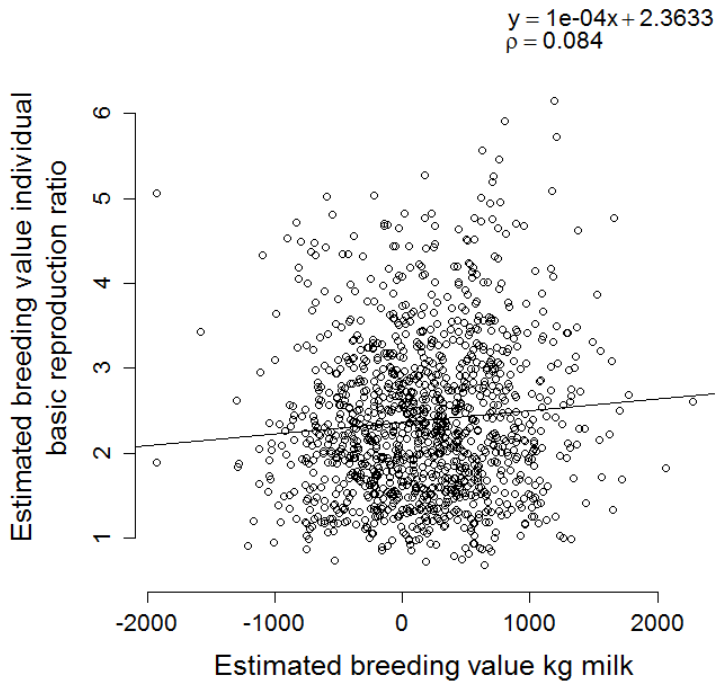


Figure 6.2. Estimated breeding value for milk production in kg provided by the breeding company against the estimated breeding value for the individual basic reproduction ratio estimated in Chapter 4 with Model 3.

Here I will investigate the relation between the estimated breeding value for the R_0 of DD (Chapter 4) and the estimated breeding value for kg milk, estimated by the breeding company.

6.3.1.1 Correlation with milk production

The optimal scenario would be a strong negative correlation, where cows with a good (high) estimated breeding value for milk production would also have a good (low) estimated breeding value for R_0, and vice versa. In the literature, there is no indication that cows with a high milk production have a greater risk at developing DD (Amory et al., 2008). However, cows affected with DD produce on average 122 kg milk less compared to unaffected cows (P = 0.35) (Argáez-Rodríguez et al., 1997). In our dataset, the correlation between the estimated breeding value for the basic reproduction ratio from chapter 4 and the genomic estimated breeding value for kg milk provided by the breeding company was 0.084 (Figure 6.2). So, there is no correlation between the estimated breeding value for R_0 and the estimated breeding value for milk production. Given this value, selection against DD is possible without a loss in milk production.

6.3.1.2 Correlation with breeding value for DD

In Chapter 4 we estimated individual breeding values for the basic reproduction ratio. The susceptibility and infectivity of the host were taken into account while accounting for variation in exposure. Previous studies on genetic variation underlying infectious diseases generally focus on susceptibility only. Disease status of the host is linked to the genotype of the host (Woolhouse et al., 1998; Springbett et al., 2003). The correlation between the breeding value for digital dermatitis estimated by the breeding company and the breeding value for the individual basic reproduction ratio was -0.295 (Figure 6.3). So, there is a weak negative correlation between the estimated breeding value for R_0 and the estimated breeding value for DD. A negative correlation was expected because there is a relation between the estimated breeding value for DD (susceptibility) and for R_0 (in which susceptibility is included). Therefore, cows with a good (high) estimated breeding value for DD also have a good (low) estimated breeding value for R_0. The difference in estimated breeding values occurs first because the estimates were based on different datasets. Furthermore, in the estimated breeding value for DD variation in infectivity and variation in exposure of the susceptible individual are ignored.

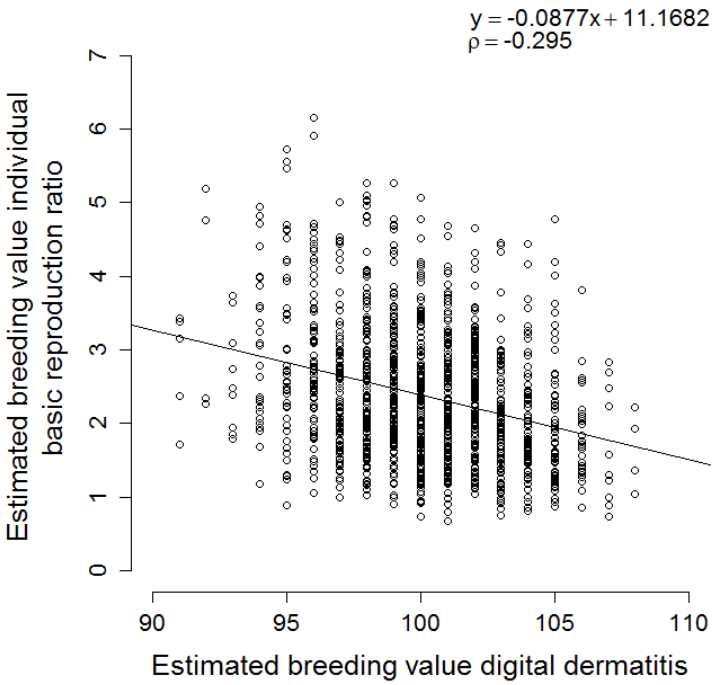


Figure 6.3. Estimated breeding value for digital dermatitis provided by the breeding company against the estimated breeding value for the individual basic reproduction ratio estimated in this Chapter 4 with Model 3.

6.3.2 Sensor systems

The phenotype data that are used in the current breeding program consist of trimming records obtained by professional claw trimmers. Both the farmer and the claw trimmer determine the amount and accuracy of the phenotype data. All cows in a herd are trimmed at the same time, or part of the herd is selected for trimming multiple times a year. The farmer selects cows in need of trimming; some cows with a claw disorder are not selected, for example because they do not show signs of lameness. Furthermore, scoring lameness is subjective, the perception and the skills of the farmer determine the quality (Whay, 2002), and the correlation between the presence of a claw disorder and lameness score is 0.50 at best (Winckler and Willen, 2001). The claw trimmers on the other hand, score the disease status of a claw in the trimming chute, again there is subjectivity in scoring (Holzhauer et al., 2006). Furthermore, cows are only trimmed, and thus scored, by the claw trimmer once or

twice a year. However, in Chapter 2 we showed that both susceptibility and infectivity are underestimated when the interval between observations is large and the number of observations limited. A short observation interval with unlimited observations will increase the accuracy of the susceptibility and infectivity estimates. The accuracy of these estimates determine the accuracy of R_0. So, for accurate breeding value estimation for R_0, phenotype data must be collected regularly and on the whole herd.

Opportunities lie in the increasing number of sensor systems used on Dutch dairy farms. In 2013 about 39% of the farms already had at least one sensor system (Steeneveld and Hogeveen, 2015). The most common sensor systems were a mastitis detection sensor system that analyse the milk, *e.g.*, somatic cell count sensors, milk colour sensors, and electrical conductivity sensors; and a sensor system that detect activity, *e.g.*, pedometers (Steeneveld et al., 2015). These type of sensors are an opportunity to obtain phenotypic data on disease status on a regular basis. However, the actual value of data depends on the sensor system's output (Rutten et al., 2013). For example, the output of mastitis detection sensor systems is informative because it gives an indication on a specific disease; but the output from pedometers or activity meters is too general to be used directly, because general information on walking behaviour or frequency does not indicate a specific disease/claw problem (Rutten et al., 2013).

The development of sensor systems that give detailed information on disease status (*e.g.*, lameness) is ongoing. Examples are Pastell and Kujala (2007), who measured the leg-load distribution of dairy cattle during milking and used a probabilistic neural network model to detect lame cows; Poursaberi et al. (2010), who analysed the back posture of walking cows with video processing techniques to detect lameness in real time; and Van Hertem et al. (2013), who applied a logistic regression model to behavioural and production data to identify lame cows. For treatment purposes, regular data on the lameness status of a cow is useful. However, for breeding value estimation for lameness, data that is more specific on the cause of lameness is needed because lameness can have many causes. Two studies took a first step towards DD detection with sensor systems. First, Alsaaod et al. (2014), who suggested using infrared thermography to detect inflammation of the foot, which could be an indication of a DD infection. By measuring the difference in maximal temperature of the coronary band and the skin between rear and front claws, they could detect cows with DD on the rear claws only with a sensitivity of 89.1% and a specificity of 66.6%. Second, Frössling et al. (2017), who used an ELISA to detect *Treponema phagedenis*–like antibodies in serum and milk samples. For serum samples, the sensitivity was 11.7% and the specificity was 100.0% when results were

based on three different antigens. For bulk milk samples, the sensitivity was 80.0% and the specificity was 100.0%. Thus, the ELISA can distinguish, to some extent, a cow/herd with a DD infection from an uninfected cow/herd.

For accurate DD detection, sensor systems must provide specific, accurate, and regular data on disease status. Results of previous studies show that more research is need to obtain accurate data on disease status. These type of data on DD status will make estimating breeding values for R_0 on a large scale possible.

6.4 Estimating susceptibility versus estimating infectivity

In chapter 2 we showed that the power to detect SNP effects for susceptibility was high, but the power to detect SNP effects for infectivity was lower, especially when the difference in the allele effect size was small. Furthermore, we showed in chapter 4 that there is more variation in the relative susceptibility estimates compared to the relative infectivity estimates (Figure 6.4).

Part of the difference in susceptibility and infectivity estimates is because genetic differences in infectivity have to be estimated indirectly from the number of susceptible group mates that become infected and from the genotype fractions among the infected individuals at different points in time. This is especially an issue in populations that consist of large groups where there are multiple infected individuals at any given point in time. Furthermore, for estimating infectivity, the environment plays an important role. We assumed that that the total infectious pressure was proportional to the number of infectious cows plus the infectious pressure from the environment; saturation of the environment did not occur. In the following, I will discuss the implications for estimating infectivity when this assumption is not correct.

In this thesis we assumed that the total infectious pressure is proportional to the number of infectious cows. This is in accordance with Kermack and McKendrick (1927) who proposed that the chance of an infection in a closed population is proportional to the number of susceptible and infectious individuals. The model parameters determine the rates with which individuals move through the states,

$$(S, I) \rightarrow (S-1, I+1) = \beta SI/N$$

$$(S, I) \rightarrow (S+1, I-1) = \alpha I,$$

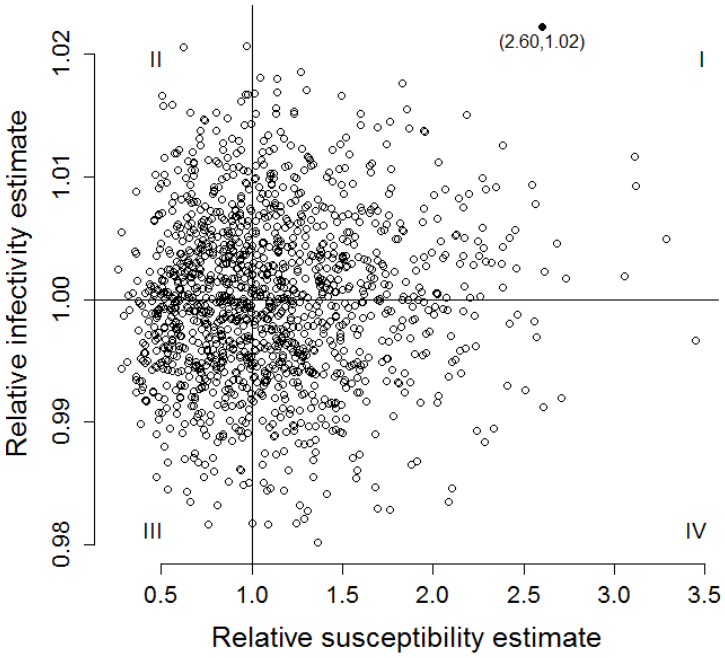


Figure 6.4. Relative susceptibility and infectivity for Digital Dermatitis from Model 2 in Chapter 4 for all genotyped cows. Compared to a cow with average susceptibility and infectivity for DD (*i.e.*, susceptibility = infectivity = 1): cows in quadrant I are more susceptible and more infectious, cows in quadrant II are less susceptible and more infectious, cows in quadrant III are less susceptible and less infectious, and cows in quadrant IV are more susceptible and less infectious. One cow is annotated, the relative susceptibility is 2.6, thus the cow is 2.6 times more susceptible to DD compared to the average susceptible cow, and the relative infectivity is 1.02, thus the cow is 1.02 more infectious compared to the average infectious cow.

where S and I both denote the disease status and number of individuals with that disease status; S + I = N; β is the transmission rate parameter; and α is the recovery rate parameter.

Around the same time, Greenwood (1931) proposed a stochastic model in which the probability that a susceptible individuals will become infected is a constant

(*incidence rate*), not depending on the number of infectious individuals. With the Greenwood assumption (adjusted from Stegeman, 2002),

$$(S, I) \rightarrow (S - 1, I + 1) = incidence\ rate * S$$

$$(S, I) \rightarrow (S + 1, I - 1) = \alpha I,$$

with 0 < *incidence rate* ≤ 1 for I > 0, and *incidence rate* = 0 for I = 0.

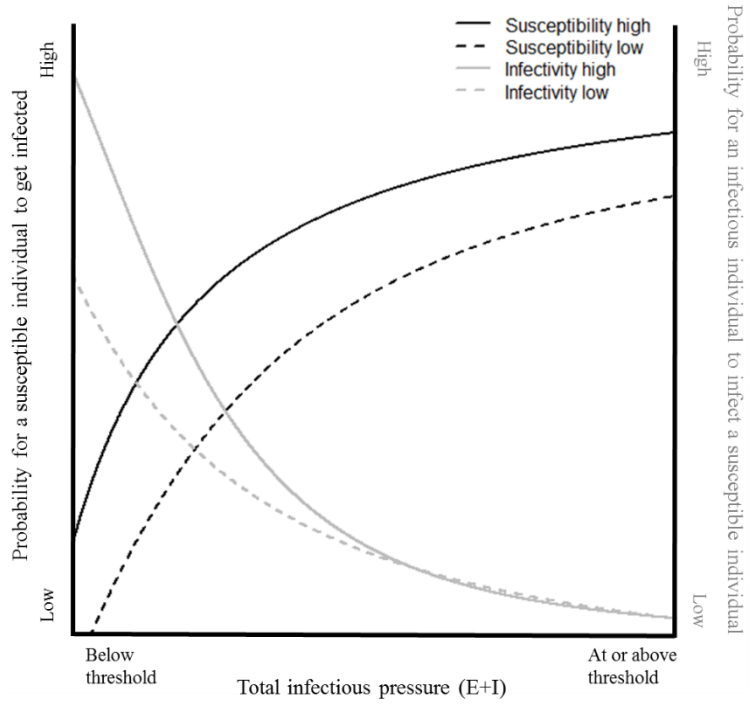


Figure 6.5. Pathogen load against the probability for two susceptible individuals (one with a high susceptibility and one with a low susceptibility) to get infected, and the probability for two infectious individuals (one with a high infectivity and one with a low infectivity) to infect a typical susceptible individual.

Whether one of these two models is correct for modelling DD-transmission is unknown. However, the existence of a threshold value for the infectious pressure (*i.e.*, pathogen load in the "environment") could be an additional explanation for the

lack of power to detect infectivity effects. Figure 6.5 shows a possible relation between the pathogen load in the environment and the probability to get infected, or infect an individual in a heterogeneous population. The probability for a susceptible individual to get infected increases with increasing pathogen load in the environment.

Differences in susceptibility can exist and be estimated independent of the pathogen load. For example, the morphology of a claw can make a cow more or less susceptible compared to another cow (Van der Waaij et al., 2005). However, in the presence of a threshold, difference in infectivity only be estimated as long as the number of infectious individuals is below the threshold, thereafter, the "environment" is saturated, the infection probability has reached its asymptote, and differences in infectivity are both indistinguishable and irrelevant. Whether such a threshold exist for DD is unknown. However, if this is the case, observing transmission in small groups of animals could lead to better infectivity estimates, provided that the number of infectious animals is small enough for the pathogen load to remain below the threshold. In that case, the infection probability has not reached its asymptote, so differences in infectivity can exist and thus estimated.

6.5 Closing remarks

In this thesis we focused on developing methods to estimate genetic variation and breeding values from time-series data on an endemic disease. Furthermore, we explored the transmission dynamics and genetic variation for DD. We showed that phenotypic differences exist between cows (Chapter 3), and that there is substantial genetic variation in host susceptibility and infectivity for DD that can be estimated (chapter 4). Genomic breeding values for R_0 had an accuracy of ~60%. Even though the reference population in this study was relatively small, ~1400 genotyped individuals, results shows that lowering DD transmission with selective breeding for R_0 is very promising. Future work should focus on obtaining specific, accurate, and regular phenotype data on disease status. These type of data will increase the accuracy of the susceptibility and infectivity estimates, and therefore the accuracy of R_0. The development of specific and accurate sensor system provides an opportunity to obtain such data. Furthermore, they make estimating breeding values for R_0 on a large scale possible.

6.6 References

Alsaaod, M., Syring, C., Dietrich, J., Doherr, M.G., Gujan, T., Steiner, A., 2014. A field trial of infrared thermography as a non-invasive diagnostic tool for early detection of digital dermatitis in dairy cows. The Veterinary Journal 199, 281-285.

Amory, J., Barker, Z., Wright, J.L., Mason, S., Blowey, R., Green, L.E., 2008. Associations between sole ulcer, white line disease and digital dermatitis and the milk yield of 1824 dairy cows on 30 dairy cow farms in England and Wales from February 2003–November 2004. Preventive veterinary medicine 83, 381-391.

Anche, M., de Jong, M., Bijma, P., 2014. On the definition and utilization of heritable variation among hosts in reproduction ratio R0 for infectious diseases. Heredity 113, 364-374.

Anche, M.T., Bijma, P., De Jong, M.C., 2015. Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47, 1-15.

Argáez-Rodríguez, F.d.J., Hird, D.W., de Anda, J.H., Read, D.H., Rodríguez-Lainz, A., 1997. Papillomatous digital dermatitis on a commercial dairy farm in Mexicali, Mexico: Incidence and effect on reproduction and milk production. Preventive Veterinary Medicine 32, 275-286.

Bishop, S.C., Woolliams, J.A., 2010. On the genetic interpretation of disease data. PLoS One 5, e8940.

Bisset, S., Morris, C., 1996. Feasibility and implications of breeding sheep for resilience to nematode challenge. International journal for parasitology 26, 857-868.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. John Wiley & Sons.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A., 1990. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of mathematical biology 28, 365-382.

Doeschl-Wilson, A.B., Villanueva, B., Kyriazakis, I., 2012. The first step toward genetic selection for host tolerance to infectious pathogens: obtaining the tolerance phenotype through group estimates. Frontiers in genetics 3.

Frössling, J., Rosander, A., Björkman, C., Näslund, K., Pringle, M., 2017. Detection of Treponema phagedenis–like antibodies in serum and bulk milk from cows with and

without digital dermatitis. Journal of Veterinary Diagnostic Investigation, 1040638717733778.

Gibson, J., Bishop, S., 2005. Use of molecular markers to enhance resistance of livestock to disease: a global approach. Revue Scientifique Et Technique-Office International Des Epizooties 24, 343.

Gilmour, A., Gogel, B., Cullis, B., Welham, S., Thompson, R., 2015. ASReml user guide release 4.1 structural specification. Hemel Hempstead: VSN International Ltd.

Greenwood, M., 1931. On the statistical measure of infectiousness. Epidemiology & Infection 31, 336-351.

Holzhauer, M., Bartels, C., van den Borne, B., Van Schaik, G., 2006. Intra-class correlation attributable to claw trimmers scoring common hind-claw disorders in Dutch dairy herds. Preventive veterinary medicine 75, 47-55.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. The Royal Society, pp. 700-721.

Knap, P.W., Bishop, S.C., 2000. Relationships between genetic change and infectious disease in domestic livestock. BSAS occasional publication, 65-80.

Lalkhen, A.G., McCluskey, A., 2008. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 8, 221-223.

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

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

Rausher, M.D., 2001. Co-evolution and plant resistance to natural enemies. Nature 411, 857-864.

Relun, A., Guatteo, R., Roussel, P., Bareille, N., 2011. A simple method to score digital dermatitis in dairy cows in the milking parlor. J Dairy Sci 94, 5424-5434.

Roberts, M., Heesterbeek, H., 1993. Bluff your way in epidemic models. Trends in microbiology 1, 343-348.

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

Solano, L., Barkema, H., Jacobs, C., Orsel, K., 2017. Validation of the M-stage scoring system for digital dermatitis on dairy cows in the milking parlor. Journal of dairy science 100, 1592-1603.

Springbett, A., MacKenzie, K., Woolliams, J., Bishop, S., 2003. The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165, 1465-1474.

Stear, M., Bishop, S., Mallard, B., Raadsma, H., 2001. The sustainability, feasibility and desirability of breeding livestock for disease resistance. Research in veterinary science 71, 1-7.

Steeneveld, W., Hogeveen, H., 2015. Characterization of Dutch dairy farms using sensor systems for cow management. Journal of dairy science 98, 709-717.

Steeneveld, W., Hogeveen, H., Lansink, A.O., 2015. Economic consequences of investing in sensor systems on dairy farms. Computers and Electronics in Agriculture 119, 33-39.

Stegeman, A., 2002. Population dynamics of infectious diseases.

Van der Waaij, E., Holzhauer, M., Ellen, E., Kamphuis, C., De Jong, G., 2005. Genetic parameters for claw disorders in Dutch dairy cattle and correlations with conformation traits. Journal of dairy science 88, 3672-3678.

Van Hertem, T., Maltz, E., Antler, A., Romanini, C., Viazzi, S., Bahr, C., Schlageter-Tello, A., Lokhorst, C., Berckmans, D., Halachmi, I., 2013. Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity. Journal of dairy science 96, 4286-4298.

Whay, H., 2002. Locomotion scoring and lameness detection in dairy cattle. In Practice, 6.

Winckler, C., Willen, S., 2001. The reliability and repeatability of a lameness scoring system for use as an indicator of welfare in dairy cattle. Acta Agriculturae Scandinavica, Section A-Animal Science 51, 103-107.

Woolhouse, M., Stringer, S., Matthews, L., Hunter, N., Anderson, R., 1998. Epidemiology and control of scrapie within a sheep flock. Proceedings of the Royal Society of London. Series B: Biological Sciences 265, 1205-1210.

Summary

Summary

For an infectious disease, the probability that a susceptible individual gets infected depends on its own susceptibility, on the number of infected group mates, and on the infectivity of those group mates. Susceptibility is the relative risk of a host to get infected when exposed to a typical infectious individual, or to the infectious material of a typical infected individual in the environment. Susceptibility therefore affects the disease status of the focal individual itself. Infectivity is the relative propensity of an infected individual to infect a typical susceptible individual. Infectivity, therefore, affects the disease status of other individuals rather than of the focal individual itself.

Susceptibility, infectivity, the contact rate, and the duration of the infectious period together determine the basic reproduction ratio (R_0). The R_0 is the average number of secondary cases caused by a typical infectious individual in a fully susceptible population. It determines the ability of an infection to establish itself in a population. The threshold value is one; if $R_0 < 1$ a typical infectious individual will infect on average less than one susceptible individual and the disease will die out with certainty. If $R_0 > 1$ a major outbreak is possible, and sometimes such a disease may persist in a population. For endemic diseases in homogeneous populations, the prevalence in the equilibrium follows from R_0 as $1 - \frac{1}{R_0}$. Breeding strategies that aim to reduce the prevalence of endemic diseases should thus aim to reduce R_0. Because R_0 depends on both susceptibility and infectivity of the host population, genetic variation in both those traits should be taken into account.

This thesis focusses on Digital Dermatitis (DD) in dairy cattle. DD is an endemic infectious claw disease associated with lameness. Lesions form mainly on the hind feet above the interdigital space next to the heel bulbs. Because DD is endemic on most dairy cattle farms, time-series data on individual disease can be collected, which may facilitate genetic selection against DD. In this thesis, we investigated transmission dynamics for DD and estimated genetic effects for both host susceptibility and host infectivity with different models.

In chapter 2, we proposed a generalized linear mixed model to estimate SNP effects on both host susceptibility and host infectivity from time-series data on individual disease status. The model accounted for variation in exposure of susceptible individuals to infectious group mates, and for the infectivity genotypes of those group mates. We used simulated phenotype data to quantify the quality of susceptibility and infectivity estimates for different SNP-effect sizes and recording intervals. SNP effects were on average underestimated and thus conservative. The power to detect SNP effects was high for susceptibility but lower for infectivity. When the total number of records was limited to eleven, the optimal recording

interval with respect to statistical power was similar for susceptibility and infectivity, and around 25 to 50% of the length of the average infectious period.

In chapter 3, we applied a generalized linear mixed model to field data on DD to estimate the infectivity of the different disease classes. Furthermore, we estimated the duration of the different classes and the distribution of the first observed classes after infection. We used the infectivity, duration, and distribution to determine the contribution of each class to R_0. The estimated R_0 for DD in our dataset was 2.36. The class characterized by irregular skin with dyskeratotis or surface proliferation contributed 88.5% to R_0. This class contributed so much to R_0 mainly because it had a high occurrence and a long duration. The infectivity of this class was similar to that of the other classes.

In chapter 4, we combined phenotype data on DD with genotype data of the same cows to estimate the heritable variation in host susceptibility and infectivity, and individual breeding values for R_0. We used four different generalized linear mixed models. The model that included only a genetic effect for susceptibility and no interaction between farm and period had the best fit. This model did not include genetic effects for infectivity, but variation in exposure of susceptible individuals to infectious herd mates was accounted for. Genomic estimated breeding values (GEBV) for susceptibility ranged from four times less susceptible to about four times more susceptible than an average cow. GEBV for infectivity ranged from ten percent less infectious to ten percent more infectious than an average cow. GEBV for R_0 ranged from 0.62 to 6.68 while the mean R_0 was 2.36. After correcting for bias, the GEBV for R_0 showed large variation, six cows had a GEBV smaller than one, and the approximate accuracy of the GEBV was ~0.6. These results show that genetic selection against DD is very promising; there is substantial heritable variation and a meaningful accuracy can be obtained from a limited amount of data.

In chapter 5, we investigated the association between genetic markers (SNPs) and different traits related to susceptibility and infectivity for DD. We used linear models to identify SNPs that are associated with different host susceptibility traits, and a generalized linear mixed model to identify SNPs that are associated with host susceptibility and host infectivity. We did not find any significant associations, but identified 135 suggestive associations on twenty different chromosomes for several host susceptibility traits with the linear model. We did not find any significant or suggestive associations for host susceptibility and infectivity with the generalized linear model. The SNP effects for the susceptibility trait disease status in the linear model had a substantial correlation (0.70) with the trait susceptibility in the GLMM. Heritability estimates for different susceptibility traits ranged from 0.09 to 0.37.

In the general discussion (chapter 6), I first discussed disease traits that were not considered in this thesis, such as the duration of the infectious period, infectivity via the environment, tolerance, and resilience. Second, I addressed breeding against infectious diseases in practice, such as the correlation between the estimated breeding value for R_0 in this thesis with the estimated breeding value for milk production and DD that are currently used. Furthermore, I addressed the role that sensor systems can play in the collection of phenotype data. Finally, I proposed saturation of the environment as an additional explanation for the lack of power to estimate infectivity effects. I concluded that (genetic) variation in the duration of the infectious period should be taken into account when calculating the breeding value for R_0, to make optimal use of all heritable variation that exists. I also concluded that sensor systems provide an opportunity for specific, accurate, and regular data collection on disease status. These data can be used to accurately estimate breeding values for R_0. Finally, I concluded that infectivity estimates can be improved when the role of the environment is taken into account.

Samenvatting

Samenvatting

Wanneer een infectieziekte een groep dieren treft, is de kans dat een individu in die groep geïnfecteerd raakt afhankelijk van de vatbaarheid van dat dier, van het aantal groepsgenoten dat geïnfecteerd is en van de infectiviteit van die groepsgenoten. Vatbaarheid is de kans dat een gezond dier geïnfecteerd raakt wanneer het wordt blootgesteld aan een infectieuze groepsgenoot, of aan infectieus materiaal dat in de omgeving is uitgescheiden door een infectieuze groepsgenoot. Vatbaarheid is dus gerelateerd aan de gezondheid van het dier zelf. Infectiviteit daarentegen is de kans dat een geïnfecteerd dier een vatbare groepsgenoot infecteert via direct contact of via de omgeving. Infectiviteit beïnvloed dus de gezondheid van andere dieren.

Vatbaarheid en infectiviteit bepalen met onder andere de duur van de infectieuze periode de standaardreproductiefactor (R_0). De R_0 is gedefinieerd als het gemiddelde aantal infecties dat wordt veroorzaakt door een dier met een gemiddelde infectiviteit in een populatie waarin alle andere dieren vatbaar zijn. De waarde van R_0 bepaalt of een infectie kan blijven bestaan in een populatie. De drempelwaarde is één: als $R_0 < 1$ dan infecteert een infectieus dier gemiddeld minder dan één ander dier, waardoor de ziekte zich niet kan verspreiden en zal uitsterven. Als $R_0 > 1$ dan infecteert een infectieus dier gemiddeld meer dan één ander dier, waardoor de ziekte zich kan verspreiden en grote uitbraken kunnen ontstaan. Maar nog steeds blijft de kans bestaan dat de ziekte door toeval uitsterft.

Een ziekte die over een langere periode in een gelijkblijvend aantal dieren van een groep voorkomt noemen we een endemische ziekte. Het gemiddelde aantal geïnfecteerde dieren in een groep noemen we de prevalentie. Hoe hoog de prevalentie is hangt af van de R_0. De prevalentie kan berekend worden als $1 - \frac{1}{R_0}$. Om een ziekte in een groep te verminderen moeten we de prevalentie verlagen, en daarvoor moeten we focussen op het verlagen van R_0. Omdat R_0 afhangt van de vatbaarheid en infectiviteit van de dieren in een groep zijn beide kenmerken belangrijk.

In dit proefschrift onderzoeken we de mogelijkheid om de prevalentie te verlagen door middel van fokkerij. Niet alle dieren hebben dezelfde vatbaarheid, er is variatie. Sommige dieren raken voortdurend geïnfecteerd en andere dieren vrijwel nooit. Het is mogelijk dat deze verschillen tussen dieren deels door de genetica (het DNA) van de dieren worden veroorzaakt. Hetzelfde geldt voor infectiviteit: sommige dieren scheiden meer pathogenen uit dan andere, wat deels door de genetica van het dier komt. Door dieren te selecteren met een lage vatbaarheid en infectiviteit kan de ziekte worden teruggedrongen als met deze dieren gefokt wordt. Bijvoorbeeld

doordat de dieren met goede genen niet gemakkelijk geïnfecteerd raken of, zodra ze geïnfecteerd raken, de ziekte niet gemakkelijk verspreiden.

In dit proefschrift kijken we naar Mortellaro in melkvee. Mortellaro is een endemische infectieziekte die op veel melkveebedrijven voorkomt. Koeien die geïnfecteerd zijn met Mortellaro hebben laesies op (voornamelijk) de achterpoten. Deze laesies zitten op de overgang tussen de klauw en de huid en kunnen pijnlijk zijn waardoor de koeien kreupel gaan lopen. Omdat Mortellaro endemisch is, is de ziekte voor een langere periode op een bedrijf aanwezig. Daardoor is het mogelijk om over een lange tijd informatie over de gezondheid van een koe te verzamelen. Die informatie kan gebruikt worden om de vatbaarheid en infectiviteit van een koe te schatten; belangrijk voor selectie en fokkerij. In dit proefschrift onderzoeken we hoe Mortellaro zich verspreidt en schatten we de genetische effecten voor vatbaarheid en infectiviteit met verschillende modellen.

In hoofdstuk 2 presenteren we een model waarmee SNP-effecten voor vatbaarheid en infectiviteit kunnen worden geschat. SNPs zijn plekken op het genoom waar variatie voorkomt. SNPs kunnen worden gebruikt als genetische merkers waarmee bijvoorbeeld plekken op het genoom kunnen worden opgespoord die invloed hebben op de vatbaarheid van een dier. In hoofdstuk 2 hebben we infecties in groepen dieren gesimuleerd. Aan de hand van het aantal dieren dat geïnfecteerd raakte en de SNPs van deze dieren hebben we de vatbaarheid en infectiviteit geschat. Dit hebben we gedaan voor verschillende effectgroottes en met observatie-intervallen van verschillende lengte. De effecten van de SNPs werden over het algemeen onderschat en zijn dus conservatief; de ware effecten zijn nog groter. De optimale duur van het observatie-interval bleek 25% tot 50% van de totale duur van de infectieuze periode te zijn.

In hoofdstuk 3 hebben we het model toegepast op de echte Mortellaro waarnemingen. Mortellaro heeft zes verschillende ziekteklassen en met het model konden we de infectiviteit van elke ziekteklasse schatten. Daarnaast hebben we de duur van elke klasse geschat en bepaald welke klassen direct na infectie geobserveerd worden. Met deze informatie konden we de bijdrage van iedere klasse aan R_0 bepalen. De geschatte R_0 in onze dataset was 2.36, een dier met een gemiddelde infectiviteit infecteert dus gemiddeld 2.36 andere dieren. De ziekteklasse waarin een verstoorde huid wordt waargenomen droeg 88.5% bij aan R_0. De hoge bijdrage van deze klasse werd voornamelijk veroorzaakt door de lange duur; de infectiviteit van deze klasse kwam overeen met die van de andere klassen.

In hoofdstuk 4 hebben we de Mortellaro-waarnemingen gecombineerd met genetische informatie van dezelfde koeien, met als doel de fokwaardes voor R_0 en de genetische variatie in vatbaarheid en infectiviteit te schatten. Daarvoor hebben

we vier verschillende modellen gebruikt. Het beste model was het model met een genetisch effect voor vatbaarheid maar zonder een genetisch effect voor infectiviteit. In de modellen corrigeerden we voor de variatie in het aantal infectieuze dieren waaraan een vatbaar dier werd blootgesteld. De fokwaarden voor de vatbaarheid van koeien varieerden van vier keer minder vatbaar tot vier keer meer vatbaar dan een gemiddelde koe. De fokwaarden voor de infectiviteit van koeien varieerden van tien procent minder infectieus tot tien procent meer infectieus dan een gemiddelde koe. De fokwaarde voor R_0 varieerde van 0.62 tot 6.68 met een gemiddelde van 2.36. Deze resultaten laten zien dat genetische selectie tegen Mortellaro veelbelovend is; er is voldoende genetische variatie, en zelfs met een beperkte dataset hebben de waarden een beduidende nauwkeurigheid.

In hoofdstuk 5 hebben we de relatie tussen SNPs en vatbaarheid en infectiviteit onderzocht. We hebben 135 suggestieve associaties op twintig verschillende chromosomen gevonden met een lineair model. De effecten van het kenmerk infectiestatus en het kenmerk vatbaarheid waren substantieel gecorreleerd (0.70). De erfelijkheidsgraden van verschillende vatbaarheidskenmerken varieerden van 0.09 tot 0.37.

In de algemene discussie (hoofdstuk 6) bediscussieer ik allereerst ziektekenmerken die in dit proefschrift niet aan bod gekomen zijn, zoals de duur van de infectieuze periode, infectiviteit via de omgeving, tolerantie etc. Daarna bespreek ik de correlatie tussen de geschatte fokwaarden voor R_0 uit dit proefschrift en de geschatte fokwaarden voor melkproductie en Mortellaro die in de praktijk worden gebruikt. Verder bespreek ik de mogelijkheden die sensorsystemen bieden voor het doen van waarnemingen. Ten slotte opper ik dat het moeilijk kunnen schatten van infectiviteit wellicht komt door verzadiging van de omgeving. Ik concludeer dat de (genetische) variatie in de duur van de infectieuze periode moet worden meegenomen in de fokwaardeschatting van R_0, dat met sensorsystemen regelmatig specifieke en accurate gegevens over de gezondheid van een dier verzamelend kunnen worden en dat de schattingen voor infectiviteit verbeterd kunnen worden wanneer de rol die de omgeving speelt wordt meegenomen in de berekening.

Curriculum Vitae

About the author

Floor Biemans was born on the 7th of February 1989 in Helmond, the Netherlands. In 2007 she obtained her high school degree from the Dr.-Knippenbergcollege in Helmond. In the same year she enrolled in the Animal Science bachelor program at Wageningen University & Research. In her bachelor she did the minor Diseases in Humans and Animals. At the Animal Breeding and Genomics group she wrote her bachelor thesis entitled "Differences in fearfulness in chicken: a comparison between a low mortality line and a control line". She obtained her bachelor's degree in 2010 and in the same year enrolled in the Animal Science master program at Wageningen University & Research. During her master she did the Research Master Track. She wrote two theses. Her major thesis at the Animal breeding and Genomics group was entitled "The basic reproduction ratio in genetically heterogeneous populations". Her minor thesis at the Adaptation Physiology group was in collaboration with Linköping University in Sweden where she resided half a year. This thesis was entitled "Energy expenditure in different chicken species: energy expenditure on different events of red jungle fowl and broilers determined by oxygen consumption at different light and temperature regimes". Floor obtained her master's degree in 2013. That year she also started her PhD research at the Quantitative Veterinary Epidemiology group and the Animal Breeding and Genomics group at Wageningen University & Research. In this project she investigated the transmission dynamics and genetic background of digital dermatitis in dairy cattle. She focussed on estimation of host genetic effects on susceptibility and infectivity for this disease. The results of this research were presented at international conferences and published in peer reviewed journals. In January 2018 Floor started working as a postdoctoral researcher at the Quantitative Veterinary Epidemiology group.

Peer reviewed papers

Biemans, F., Jong, M.C.M. de, Bijma, P. 2017. *A model to estimate effects of SNPs on host susceptibility and infectivity for an endemic infectious disease*. Genetics Selection Evolution 49:53.

Biemans, F., Bijma, P., Boots, N.M., Jong, M.C.M. de. 2017. *Digital Dermatitis in dairy cattle: The contribution of different disease classes to transmission*. Epidemics. In press.

Manuscripts in preparation

Biemans, F., Jong, M.C.M. de, Bijma, P. *Genetic variance components of host susceptibility, infectivity and R0 for Digital Dermatitis in dairy cattle*.

Biemans, F., Jong, M.C.M. de, Bijma, P. *A genome-wide association study for susceptibility and infectivity of dairy cattle to Digital Dermatitis*.

Contributions to conferences

Biemans, F., Bijma, P., Jong, M.C.M. de. 2014. *Estimating genetic effects on susceptibility and infectivity for infectious diseases*. 23rd International Genetic Epidemiology Society. Vienna, Austria.

Biemans, F., Bijma, P., Jong, M.C.M. de. 2015. *Estimating genetic differences in hosts' susceptibility and infectivity for infectious diseases*. 14th International Symposium on Veterinary Epidemiology and Economics. Mérida, Yucatán, México.

Biemans, F., Bijma, P., Jong, M.C.M. de. 2016. *Digital Dermatitis in dairy cattle: infectivity of the different stages*. 16th International Conference on Production Diseases in Farm Animals. Wageningen, the Netherlands.

Biemans, F., Bijma, P., Jong, M.C.M. de. 2017. *Estimation of genetic effects on susceptibility and infectivity from data on Digital Dermatitis*. 6th International Conference on Infectious Disease Dynamics: Epidemics. Sitges, Spain.

Contributions to seminars

Biemans, F., Bijma, P., Jong, M.C.M. de. 2015. *Estimating genetic differences in hosts' susceptibility and infectivity for infectious diseases*. 27th Veterinary Epidemiology and Economics symposium. Deventer, the Netherlands.

Biemans, F., Bijma, P., Jong, M.C.M. de. 2016. *Digital Dermatitis in dairy cattle: how infectious is it?* WIAS Science day. Wageningen, the Netherlands.

Biemans, F. 2016. Symposium Dies Natalis Wageningen University & Research. Wageningen, the Netherlands.

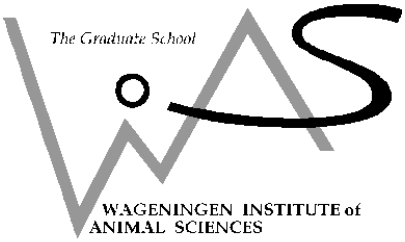
Biemans, F., Bijma, P., Boots, N.M., Jong, M.C.M. de. 2017. *Digital Dermatitis in dairy cattle: Infectivity of the different classes*. WIAS Science day. Wageningen, the Netherlands.

Biemans, F., Jong, M.C.M. de., Bijma, P. 2017. *Met fokkerij infectieziekte tegen gaan*. *Estimating genetic differences in hosts' susceptibility and infectivity for infectious diseases*. 14th Fokkerij en genetica connection. Doorwerth, the Netherlands.

Biemans, F., Bijma, P., Jong, M.C.M. de. 2017. *Digital Dermatitis in dairy cattle: how infectious is it?* NWO symposium Co-Creation? Naturally! Amersfoort, the Netherlands.

Training and education

Training and supervision plan



The basic package (3 ECTS)	
WIAS Introduction Day	2015
WIAS Course on Essential Skills	2015
Ethics and Philosophy in Life Sciences	2014

Scientific Exposure (19.9 ECTS)	
International conferences	
23rd IGES, Vienna, Austria	2014
14th ISVEE, Mérida, Yucatán, México	2015
16th ICPD, Wageningen, the Netherlands	2016
Epidemics[6], Sitges, Spain	2017
Seminars and workshops	
Symposium 'Grip op Klauwen', Bunschoten, the Netherlands	2013
WEES seminars, Wageningen, the Netherlands	2014 -2017
WIAS Science Day, Wageningen, the Netherlands	2014, 2016, 2107
WGS Workshop Carousel, Wageningen, the Netherlands	2014, 2015, 2106
Symposium 'Limits to produce', Wageningen, the Netherlands	2014
'Preventing Lameness in Feedyard Cattle', West Point, Nebraska, USA	2014
Symposium 'Weerbaar Vee', Ede, the Netherlands	2015
F&G Connection days, Doorwerth, the Netherlands	2017
Symposium 'Co-Creation? Naturally!', Amersfoort, the Netherlands	2107
Presentations	
Epidemiological meeting, Lelystad, the Netherlands (oral)	2014
23rd IGES, Vienna, Austria (poster)	2014
27th VEEC, Deventer, the Netherlands (oral)	2015
14th ISVEE, Mérida, Yucatán, México (oral)	2015
WIAS Science Day, Wageningen, the Netherlands (poster)	2016
Symposium Dies Natalis Wageningen University & Research (oral)	2016
16th ICPD, Wageningen, the Netherlands (oral)	2016

WIAS Science Day, Wageningen, the Netherlands (oral)	2017
F&G Connection days, Doorwerth, the Netherlands (oral)	2017
'Co-Creation? Naturally!', Amersfoort, the Netherlands (poster)	2107
Epidemics[6], Sitges, Spain (poster)	2017

In-depth studies (12.6 ECTS)

Disciplinary and interdisciplinary courses	
Social Genetic Effects: Theory and Genetic Analysis	2013
Workshop 'Societal Impact'	2015
Genotype by environment interaction, uniformity and stability	2015
Orientation on mathematical modelling in biology	2016
Genetic Epidemiology of infectious diseases in livestock	2017
Advanced statistics courses	
Generalized Linear Models	2014
PhD students' discussion groups	
Falconer discussion group	2014-2015
Quantitative genetics discussion group	2015-2017
Fortran discussion group	2016-2017

Professional Skills Support Courses (5.2 ECTS)

Information Literacy including Endnote introduction	2013
Data Management	2014
PhD Competence assessment	2014
Interpersonal Communication for PhD students	2014
Techniques for Writing and Presenting a Scientific Paper	2016
Career Orientation	2016
Writing the General Introduction and Discussion	2017
Last Stretch of the PhD Programme	2017

Didactic Skills Training (6.3 ECTS)

Lecturing	
Animal Breeding and Genetics (ABG20306)	2016
Supervising practicals and excursions	
Inleiding Dierwetenschappen	2014, 2015, 2016

Supervising theses	
BSc-thesis (three students)	2015-2016
Preparing course material	
Review RMC proposals	2016, 2017
Management Skills Training (4 ECTS)	
Organisation of seminars and courses	
Seminar 'Breeding against Infectious Diseases in Animals'	2016
Membership of boards and committees	
WIAS Science Day 'Facing the future'	2015
Partycom ABG	2015-2017
Education and training total	**51 ECTS**

Acknowledgements

Acknowledgements

It was impossible to do this PhD alone, so I would like to thank some people.

First of all, Mart and Piter. Thank you both so much for giving me the opportunity to work on this project! I really learned a lot the last four years. You believed I could do this, even when I did not believe it myself. It was a pleasure to learn from the both of you. Mart, your door was always open when I had a question. You always made time for me to discuss problems and to help me think of solutions. You would not stop until we found an answer. Those instant meetings were often pleasantly accompanied with discussions about people, politics, and life in general, which I really appreciated. Piter, you looked at my manuscripts MANY times. You took each version as opportunity to teach me how to write better so that the content would be accessible for a broader public. The way you can transform a manuscript is like magic. After you rearranged paragraphs and sentences it was sometimes hard to believe that I wrote them myself in the first place. During our joined meetings with Mart you pushed me to think carefully about what I did, and not to take the results for granted. And I am glad you did. During those meetings you, furthermore, took on the role of translator when Mart said something really smart but incomprehensible to me. It is fascinating that both of you only need one word to know what the other one means. Thank you both for everything!!!

Natasja, it is not often that I get the opportunity to work with one of my best friends. You were the best research assistant I could ever imagine, possibly even better. For half a year, you got out of bed when it was still night to accompany me to score cows. Even when your car broke down (multiple times) you were determined to find a solution to be there. Even though it was winter, temperatures were far below zero, and it was freezing cold in those early mornings in the milking parlour, you were there. We did it! And it was great! Thank you so much for joining me for such a large part of my PhD journey; you are the best!

Milou and Pascal, my paranymphs. I am grateful that both of you want to be on stage with me during my defence, I am sure that I can use the support. But I am certain that if anyone can provide this support, it is you. You are both awesome people! We are going to have ourselves a crazy party when it is all over and done with!

Bas, Gerdien, and René, thanks for being part of my supervisory committee. It was really nice to have someone to turn to when I had questions about, amongst others, glmm, statistics, how to set up a field trial, or phenotype and genotype data on cows. Too bad our final meeting was not meant to happen!

Thanks to CRV, and especially to René and Lotte, for contacting the farmers and providing the genotype and phenotype data.

Thanks to all twelve farmers and their families. You let us score your cows many times in the early morning or late evening, and even if we slowed down milking, you always welcomed Natasja and me with open arms. During our conversations, I learned a lot about dairy farming in practice. It was a pleasure to do the field trial on your farms.

Dörte, you let me tag along for one week in the US for a speed course on DD. It was awesome to see the (big!) dairy farms and experience some real American culture. You are a very warm and kind person, I am grateful you took me under your wings. Thanks for teaching and helping me.

Menno, thank you helping me to get some field experience on scoring DD in the Netherlands. During the scoring days in Friesland, I saw many different lesions. You always kept showing interest in my work which I really liked.

Thanks to Elies for showing me claw trimming, Linda and Peer for letting Natasja and me practice scoring in the milking parlour, and the 'de Bruin'-family for letting us score their cows and for their help with setting up the scoring trials. Tessa, thanks for making a beautiful folder.

To all my colleagues at ABG, QVE, and ADP; thanks for all the support, friendships, and happy moments. In the first year of my PhD, I spent most of the coffee breaks with the ADP group and everyone was always very nice and welcoming. Thanks to my office mates at ADP for providing me with a good start of my PhD. After the first year, QVE mixed with the ABG group and I joined the Partycom. This was the best thing I could have done; I got to know a lot of people and experienced many great moments because I had to stay to the end of every event. Yvonne, Kyle, Mathieu, Pascal, and Maria; you are the most awesome party people! I really enjoyed my time in Radix with nice discussions and plenty of cake during the coffee breaks. And of course the foosball matches! I never get bored of those. Yvonne, Kasper, and Ilse, I hope that we will play the final of the tournament someday.

Lisette and Ada, a special thanks to you for all your help and support along the way.

Patricia, my PhD colleague at QVE, we had a great time in México. That week in Mérida is probably the best week of my PhD. It was very nice to experience it with you.

Laurine, Ruth, Kiki, Natalia, and Pranav, thanks for your thesis work on this topic. You explored issues that I could not cover in my thesis; it was a great working with you.

Betty and Olof, thanks for your interest in my work and your continued support. Jesse thanks for the great cover design. It turned out better than I ever could have dreamed.

Pa, moeders, and bro, thanks for always being there! You probably did not always understand what I was doing but you were always there for help. I dedicate this thesis to you. I love all three of you. I hope you are proud.

And last but definitely not least, Roos. Roos you were there from the very beginning. From the moment Piter called me to tell me I could start my PhD, up to the very end when I finished the booklet. I could not have done this without you. You were there to support me when I was down, and you were there to celebrate the highlights. You had to listen to all my worries, troubles, and whining but you never complained. You always listened, helped me think of a solution, and supported me. When a paper was accepted, we celebrated with very tasty cakes that you baked. You kept telling me that everything would work out in the end, and it did. I know that, especially during the last months of my PhD, I was often absent, worrying about work and not paying attention to what you were saying. But you never lost your patience with me; you were just there, my rock in stormy seas. Thanks Roos, for standing by my side every step of the way. I would not know what to do without you. Lieve mafkaas; I can never thank you enough. Together we can conquer anything, let's see what the future brings! I love you!!! :*

Colophon

The research described in this thesis was financially supported by the Netherlands Organisation for Scientific Research (NWO) as part of the research programme “Meer met Minder” (847.13.004). The genotype data and part of the phenotype data used in this thesis were provided by breeding company CRV.

The cover of this thesis was designed by Jesse Michaël Akkerman.

The thesis was printed by Digiforce | Proefschriftmaken.nl.