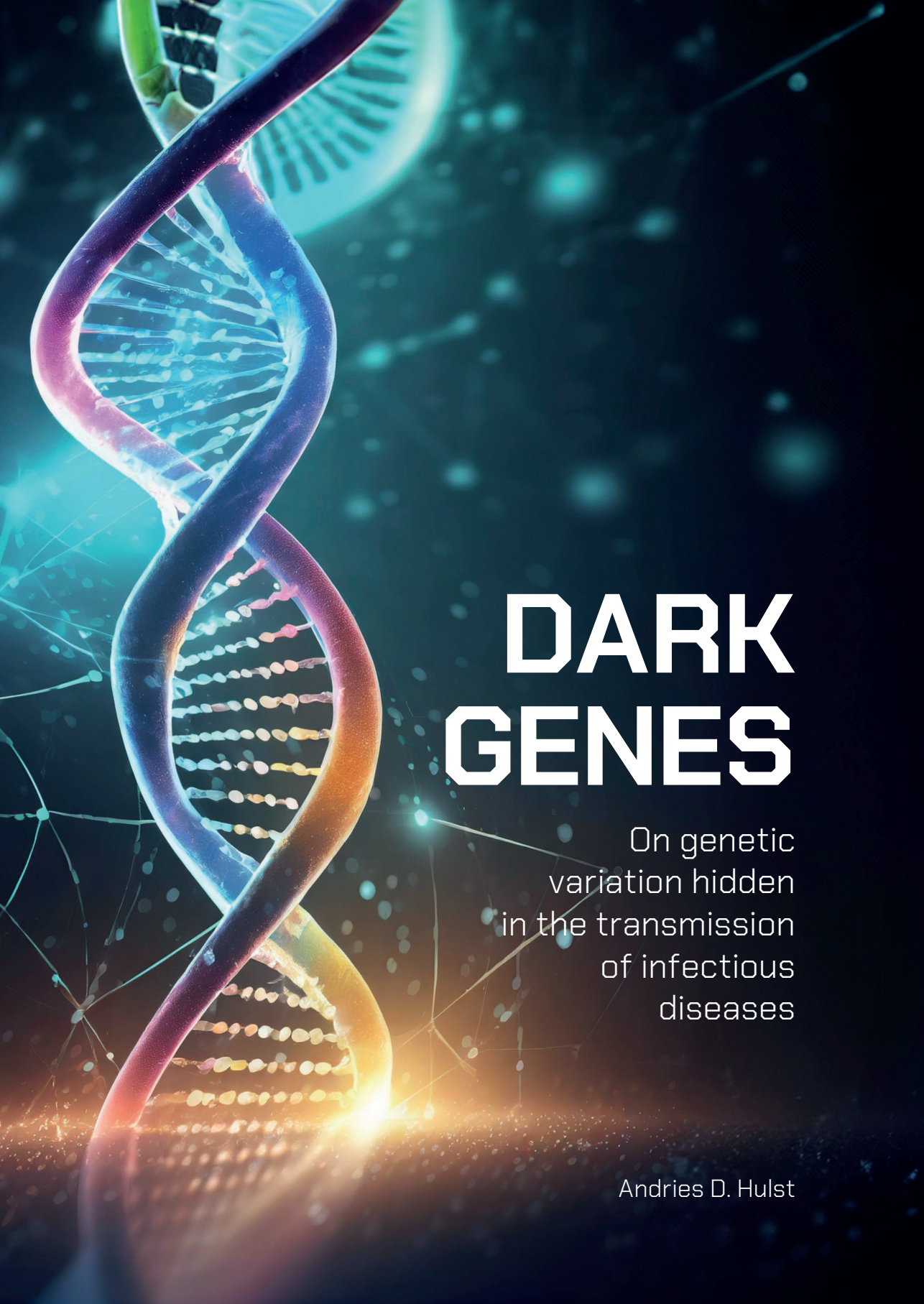


DARK GENES

On genetic variation hidden in the transmission of infectious diseases

Andries D. Hulst

Propositions

1. Gradual multi-trait selection will not achieve breeding goals for infectious disease traits in livestock.
   (this thesis)

2. Getting vaccinated against an infection is typically a mutualistic act.
   (this thesis)

3. Replacing grass with maize as feed for dairy cows to reduce methane emissions is unsustainable.

4. Reducing the number of animals to make the livestock system more extensive will increase the risk of transmission of zoonoses.

5. The Dutch vaccination strategy against human influenza wastes vaccines by applying them to the wrong group of people.

6. By opposing nuclear energy, 'green' NGOs secure their own future relevance.

7. Positive discrimination maintains inequality.

Propositions belonging to the thesis, entitled

Dark Genes: On genetic variation hidden in the transmission of infectious diseases

Andries D. Hulst
Wageningen, 1st of March 2024

Dark Genes:
On genetic variation hidden in the transmission of infectious diseases

Andries D. Hulst

Thesis committee

Promotors

Prof. Dr M.C.M. De Jong

Professor of Quantitative Veterinary Epidemiology

Wageningen University & Research

Dr P. Bijma

Associate professor Animal Breeding and Genomics

Wageningen University & Research

Other members

Prof. Dr D.K. Aanen, Wageningen University & Research

Prof. Dr R.J. de Boer, Utrecht University, The Netherlands

Prof. Dr J.A. Woolliams, University of Edinburgh, Edinburgh, Scotland

Dr E.F. Knol, Topigs Norsvin, 's-Hertogenbosch, The Netherlands

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

Dark Genes:
On genetic variation hidden in the transmission of infectious diseases

Andries D. Hulst

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 Friday 1 March 2024
at 4 p.m. in the Omnia Auditorium.

Andries D. Hulst
Dark Genes: On genetic variation hidden in the transmission of infectious diseases,
216 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summaries in English and Dutch

ISBN 978-94-6447-978-2
DOI: https://doi.org/10.18174/641830

Abstract

Hulst, A.D. (2024). Dark Genes: On genetic variation hidden in the transmission of infectious diseases. PhD thesis, Wageningen University, The Netherlands.

Infectious diseases form a continuous threat to life on earth. Various interventions have been developed to combat infectious diseases in humans, animals, and plants, like vaccination, and culling of infected animals. Genetic selection has attracted attention as intervention against infections and has been quite successfully applied in plants. In animals, genetic selection against infections is also applied, but low heritability of resistance traits typically restricts the rate of genetic improvement. In this approach, however, the transmission of infections, i.e., the fact that animals infect each other, is fully ignored. In this thesis, I show that transmission of infections causes indirect genetic effects, which drastically changes response to selection in the prevalence of the infection, and even makes eradication of an infection through genetic selection possible. The origin of the indirect genetic effects is very comparable to the mechanism that causes herd immunity in a vaccination program. If animals are genetically more resistant to infection, this not only reduces their own probability of becoming infected, but also the average number of infectious individuals in the population, which then reduces the probability to become infected for all animals in the population. The size of this indirect effect increases with decreasing prevalence of the infection. Thus, the lower the prevalence, the larger the indirect effect, and the larger the response to genetic selection. Importantly, this rule does not only apply to genetic selection, but to any intervention that reduces the probability of infection. A general problem with infectious disease interventions is that they create selection for adaptations in the pathogen that make the intervention less or not effective. This thesis shows that to prevent such adaptations, the aim of genetic selection should be to eradicate an infection as fast as possible from a local livestock population. In that way, the pathogen gets limited chance to adapt. Thus, combining genetic selection with other interventions is advised to achieve local eradication. In conclusion, this thesis shows that genetic selection of animals against infectious diseases is very promising, but the typical approach in animal breeding of gradual reduction in prevalence should be reconsidered and possibly replaced by an approach that aims for stepwise local eradication of the infection.

Contents

Chapter 1

General Introduction

In May 1796, Edward Jenner, an English doctor, infected an eight-year-old boy with pus obtained from cowpox blisters on the hands of a milkmaid. He expected and tested that by doing so, he made this boy resistant to small pox, one of the most lethal human diseases (JENNER 1800). He probably did not expect that with this discovery, he had also discovered one of the most important interventions to combat infectious diseases for the following centuries. Today, Jenner's seminal work has resulted in the eradication of small pox from the human population (HENDERSON 1976; HENDERSON 2011), and in available vaccines and running vaccination programmes for numerous human and animal diseases (WORLD HEALTH ORGANIZATION 2023).

Half a century after Jenner's discovery, biologists travelled the world searching for, collecting, and describing newly discovered species from everywhere, of which a lot had never been described before. Two of them, Charles Darwin and Alfred Russel Wallace, independently arrived at a new theory for the development of the large variety of species (DARWIN 1864; WALLACE 1869; QUAMMEN 1996). With their theory of natural selection and evolution, an important foundation for the discovery of the role of genetic variation and inheritance was laid. Around the same time, Gregor Mendel did his famous discoveries on crosses between pea plants (MENDEL 1866), laying the basis for what was later called Mendellian inheritance. More than a century later, this scientific basis has led to major human-directed genetic change on animal and plant populations, both aesthetically and in production-related traits.

Of more recent time is the realisation that the two fields can be brought together: genetic selection can also be used as an intervention against infectious diseases. This has been widely applied in plant breeding, where a great variety of crops has been bred resistant to various pathogens (e.g. VANDERPLANK 1984). In animal breeding examples are limited to a few cases, most likely because of higher genetic complexity compared to plants. Therefore, there is a need to improve quantitative genetic understanding of traits related to infectious disease resistance, in particular of animals. This thesis aims to do so by further enhancing the integration of infectious disease epidemiology and quantitative genetics. It shows that the potential to reduce the prevalence of an infectious disease is much larger than is predicted by common quantitative genetic models. It explores the

risk of pathogen adaptation to genetically selected hosts. It evaluates methods to estimate genetic parameters for infectious disease traits, and, lastly, elaborates on strategies to incorporate selection against infectious diseases in animal breeding programs.

1.1 Infectious disease and its impact on livestock

Infectious diseases have a profound effect on life on earth. Populations of plants and animals are generally affected by a variety of pathogens, which are often also a shaping force in the evolution of these populations (KARLSSON *et al.* 2014; EBERT AND FIELDS 2020). More visible is the direct impact of infectious diseases on the health of plants, and health and welfare of animals and humans and the economic costs that come with that impact (BENNETT 2003; KNAP AND DOESCHL-WILSON 2020). The costs associated with infectious diseases of livestock reach far beyond the direct costs of treatment and control of the infection, which are in itself already substantial (RUSHTON 2009). The indirect costs concern production losses due to dying of the animals, reduced growth or delivery bans of products from diseased animals. Furthermore, outbreaks or endemics of certain infectious diseases might lead to trade restrictions, affecting the sales and market prices of animal products (BENNETT 2003). In case of zoonotic infections, additional, considerable indirect costs arise because of human infection and additional measures taken in the prevention thereof.

For each livestock species, several infectious diseases can be identified that have ongoing large impact on health, welfare and economy (DAVIES *et al.* 2009), while measures to control or prevent the infection are not effective enough to considerably reduce the impact. In the following, I will discuss three of such infections. In dairy cattle, mastitis is an example of such a disease. It has a complex aetiology, different strains of bacteria can cause the infection and contamination of the environment plays a role (e.g. WHITE *et al.* 2001; WHITE *et al.* 2006). Despite several available preventive and control measures, mastitis continues to be one of the most important infectious diseases of dairy cattle, with large impact on cow welfare and farm economics, where the economic impact concerns costs of treatment and preventive measures, reduced milk production, and a ban on the supply of milk from affected and treated cows.

In pigs, porcine reproductive and respiratory syndrome (PRRS), caused by Porcine reproductive and respiratory syndrome virus (PRRSV) is one of the most impactful and costly infectious diseases. Infection with PRRSV causes reproduction failure in sows and suppresses the immune system of pigs, making them more susceptible to other infections (NIEDERWERDER *et al.* 2016). Among the associated problems are reduced growth and stillbirths, that have large impact on the economic benefit of the pig industry (NEUMANN *et al.* 2005). Control measures are mainly directed at preventing transmission of the infection, e.g., within farm by test and removal of infected pigs or between farms by closing of herds. Furthermore, several vaccines are available to immunize pigs (VU *et al.* 2017; CHAE 2021). The effectivity of these measures is however variable, partly due to a large genetic diversity of the virus (VU *et al.* 2017).

In poultry, coccidiosis is an example of an infectious disease with large impact on health, welfare and economic benefit. It is caused by *Eimeria* protozoans, of which oocytes are ingested and cause infection in the intestine of chickens (DAVIES *et al.* 1963). Infection with *Eimeria* results in impaired growth, but also suppresses the immune system, causing high mortality of chicks. Control of coccidiosis occurs through vaccination, the use of anticoccidial drugs, and strict management measures (FATOBA AND ADELEKE 2018). However, due to the emergence of resistant strains of the parasite, these measures are not very effective anymore. Consequently, coccidiosis continues to seriously impair the welfare and health of poultry.

The above are just a few examples of infectious diseases with large impact on the health and welfare of livestock, to give an impression of the diversity of pathogens causing infection, the associated symptoms, and welfare implications. Furthermore, they show that costs of infectious diseases reach beyond the direct costs of treatment, through impaired production, trade restrictions, and potential costs of human infection.

1.2 Principles of genetic selection

In genetic selection of livestock, the main goal is to improve the next generation such that it outperforms the current generation for a breeding goal that usually consists of a large number of traits with their desired direction of improvement. A (simplified) breeding goal for a dairy cattle population could be to increase milk yield, to increase the percentage of fat and protein in the milk, while reducing the incidence of mastitis in the population. In commercial breeding programs, breeding goals consist of a large number of traits in different categories, such as production, health, and fertility. Traditionally, there has been a high interest in improving production and efficiency traits, to increase yield while reducing costs (e.g., feed) of production. In recent decades, breeding goals became wider with the inclusion of traits related to welfare, health, and fertility (NEETESON-VAN NIEUWENHOVEN *et al.* 2013).

An important parameter in animal breeding is the heritability. The heritability expresses the proportion of the total (phenotypic) variation in a trait that is additive genetic, i.e., that is inherited by the next generation. A large value of heritability (i.e., >0.5) potentially makes it easier to achieve a high rate of genetic improvement in a trait. However, limited available genetic variation in a trait or inaccurate breeding value estimation will still restrict response to selection. Similarly, a low heritability does not necessarily lead to a low rate of genetic improvement. If genetic variation is large and breeding values are accurately estimated, a high rate of genetic improvement can still be achieved.

Complicating factor is that change in a certain trait can lead to a correlated change in another trait. For example, an increase in milk yield might lead to a (correlated) decrease in milk fat percentage, because genes that have an effect on milk yield also influence the fat percentage of the milk. Correlations can be favourable and unfavourable, dependent on the breeding goal. In the example with milk yield and fat percentage the correlation is unfavourable, because the breeding goal was to increase both milk yield and fat percentage. This unfavourable correlation limits the achievable rate of genetic improvement in the breeding goal. On the other hand, a favourable correlation between traits in a breeding goal will increase the rate of genetic improvement.

The most speaking examples of the large potential of genetic selection are the improvements that were achieved in production traits in the previous century. Milk production of dairy cows in the Netherlands, for example, has more than doubled since the 1950s (from ~4000 to ~9000 kg/year). Even more dramatically, body weight of broilers at 42 days of age increased by about a factor 5 between 1957 and 2001 (from ~0.6 to ~3 kg) (HAVENSTEIN *et al.* 2003). Although this phenotypic change could in principle also be attributed to improvements in feeding and management, an experiment by HAVENSTEIN *et al.* (2003), where broilers from the 1957 population were raised under 2001 conditions and vice versa, showed that about 80 percent of the improvement could be attributed to genetic change. Important to note here is that production traits typically have a high heritability, whereas traits related to welfare, health, and fertility often have a lower heritability, and are therefore expected to be more difficult to improve.

1.3 Genetic selection as intervention against infectious disease

Although health traits are typically not as easy to improve as production traits, there has been interest to select livestock for increased resistance to infectious diseases for some decades. This interest in genetic selection was mainly driven by the failure of other methods, such as vaccination, drug treatment, and management, to control certain infectious diseases. Furthermore, increasing emergence of antibiotic and anthelmintic resistant pathogens increased the need to find additional or alternative strategies against infectious diseases.

Interest in breeding for resistance to infectious disease is not limited to animals. In plant breeding, it is much more common to breed disease-resistant varieties. This is most likely caused by less complexity of the resistance traits compared to animals. Furthermore, in most plant species it is easier to spread a specific gene throughout the whole population, e.g., through inbreeding or cloning. In plants, often genes exist that provide resistance to a specific infectious disease. These genes can relatively easily be introgressed to make a resistant production variety (VANDERPLANK 1984). In animals, on the other hand, resistance to most infectious disease is likely a quantitative trait, with numerous genes each contributing little to the trait value (e.g. TIEZZI *et al.* 2015; BIEMANS *et al.* 2019b). In combination

with less efficient reproduction compared to plants, this makes it more difficult to breed a resistant population.

Typically, low heritability estimates are found for infectious disease traits from classical quantitative genetic approaches, indicating a limited genetic basis for such traits (e.g. PÉREZ-CABAL *et al.* 2009; WELDERUFAEL *et al.* 2017; MARTIN *et al.* 2018). However, the interest, or one could also say the need, to breed disease resistant livestock led to efforts to explore and improve the prospects of genetic selection to reduce the impact of infectious diseases on livestock populations. An important contribution to this was made by Steve Bishop and co-workers. His work on infectious diseases started off in the end of the 90s with a range of studies on selection against gastro-intestinal nematodes in sheep (BISHOP AND STEAR 1997; BISHOP AND STEAR 1999; BISHOP AND STEAR 2003). From this early work, it became clear that the classical quantitative genetic models underestimated potential response in infectious disease traits. Which in turn led to the realisation and demonstration that integration of quantitative genetics and infectious disease epidemiology was needed to enhance genetic selection against infectious diseases (MACKENZIE AND BISHOP 2001; BISHOP AND MACKENZIE 2003; NIEUWHOF *et al.* 2009).

Later studies intensified integration of both fields and showed the existence of indirect genetic effects (IGEs) in the transmission of infectious diseases, which could considerably increase response to selection (LIPSCHUTZ-POWELL *et al.* 2012; ANCHE *et al.* 2014; TSAIRIDOU *et al.* 2019). An IGE is the effect of the genotype of an individual on the trait values of other individuals (GRIFFING 1967; MOORE *et al.* 1997; MUIR 2005; BIJMA 2014). IGEs on infection status were shown to arise both from an individual's infectivity, the propensity of an individual to infect other individuals, and susceptibility, the propensity of an individual to get infected by other, infectious, individuals. Infectivity is thus a full indirect effect, i.e., it has no effect on the infection status of the individual itself, only on that of other individuals. For susceptibility, the indirect effect arises because individuals with lower susceptibility are less often infected and thereby also infect fewer other individuals (ANCHE *et al.* 2014).

Further studies focussed on methods to estimate genetic effects for infectious disease traits, which resulted in the development of two different tools for estimation. First, the generalised-linear-mixed-model approach, which is tailored

to the transmission process and able to estimate individual genetic effects for both susceptibility and infectivity (ANCHE *et al.* 2015; BIEMANS *et al.* 2017; BIEMANS *et al.* 2019a). Second, the so-called SIRE software tool, that uses a Bayesian optimisation algorithm to estimate individual genetic effects for susceptibility, infectivity, and also recoverability (POOLEY *et al.* 2020; POOLEY *et al.* 2022b).

Despite these efforts, there are still only few examples of successful selection against infectious diseases in livestock. The most well-known example is the selection for infectious pancreatic necrosis (IPN) resistance in Atlantic salmon. Using marker-assisted selection on an IPN-resistance QTL with large effect, infection-induced mortality was reduced by 70% in two years (AQUAGEN 2012). Unfortunately, the existence of QTLs with such a large effect for infectious disease resistance seems uncommon in animals, so that the success in selection against IPN might be a very particular case. Furthermore, as will be explored in the next paragraph, dependence on a single resistance QTL might impose serious risk for the emergence of pathogen strains that can escape the genetic resistance of hosts.

1.4 Pathogen adaptation to infectious disease interventions

A general problem with control measures targeted at infectious diseases is that they exert selection pressure on the pathogen population causing the infection. Because these pathogen populations consist of living organisms that often evolve rapidly, due to high mutation rates and short generation times, such selection pressure might lead to genetic adaptations in the pathogen population that make the interventions partly or fully ineffective. The most prominent example of this is the widespread antibiotic resistance (DAVIES AND DAVIES 2010), but resistance to other interventions such as vaccination, antiviral drugs, and disinfectants also occurs.

The high adaptive potential of pathogen populations leads to the expectation that genetic selection of animals for increased resistance to infectious diseases will evoke such adaptations in the pathogen. In plant breeding, where selection for resistance is more common, counteracting adaptations of the pathogen are widespread. Introgressed resistance genes are often relatively easily

circumvented by mutations in the pathogen (McDonald and Linde 2002). Likewise, the pathogen that cause IPN in Atlantic salmon seems to have started evolving in response to the example from previous paragraph. Recently, a new variant of the IPN-virus was found that causes higher mortality in genetically resistant salmon (Hillestad *et al.* 2021).

In response to previous examples, however, strategies have been developed to diminish the risk of pathogen evolution. An option is to combine multiple resistance mechanisms in one treatment, to make it harder for the pathogen to adapt. It then needs a combination of mutations to circumvent the different working mechanisms instead of just one, which drastically reduces the chance that a resistant pathogen evolves. An example is therapy against HIV where different antiviral drugs are combined in one therapy, such that pathogen strains that have evolved resistance to one of the drugs are killed by one of the others (Kennedy and Read 2017). The same principle is used in plant breeding, instead of using one resistance gene, multiple resistance are now jointly introgressed into a variety to prevent pathogens from adapting to the single resistance genes one by one (Mundt 2018). A different strategy is applied to prevent antibiotic resistance development, which is especially relevant to new antibiotics. The use of those is restricted to cases where there are no other options available. If they are used, the dose applied and duration of the treatment are high and long enough to make sure that all target bacteria are killed rapidly (Walsh 2003). In that way, there is a reduced chance that a resistant mutant of the bacteria will emerge, before the entire population has been killed.

1.5 Research problem and relevance

The possibility to combat animal infections by using genetic selection has been considered an interesting option for many years. Despite this interest, there are very few examples of successful selection against infectious diseases. Incidence of clinical mastitis, for example, remains high, despite great effort to improve cow genetics and herd management in the past decades (Martin *et al.* 2018). Furthermore, quantitative genetic research has usually treated infectious disease

resistance as an individual trait and thereby ignored the epidemiological knowledge on the transmission of infectious diseases.

A complicating factor is that infectious diseases are caused by living organisms, which have the potential to evolve adaptation strategies to applied interventions. Pathogen adaptation following genetic selection for improved infectious disease resistance has received little attention in livestock, although there seems to be quite some awareness of the risk (many papers write one or two sentences).

Because of the estimated small potential and little knowledge about pathogen adaptation, strategies to effectively incorporate infectious disease traits in breeding programs are currently lacking. Nevertheless, it is also because of these two factors that optimum strategies for genetic selection might differ from those used for other traits. Thus, improved understanding of the role of genetics in transmission and of pathogen adaptation is needed to come to strategies for breeding disease resistant livestock.

These strategies are needed to correctly determine the potential of genetic selection to reduce the impact of important infectious diseases in livestock populations, especially for those that are difficult to control through other measures (e.g., PRRS, mastitis). As a result, health and welfare of the animals might improve, treatment costs and costs due to production losses reduce, as might the amount of labour that is needed to run a farm. On top of those effects on-farm, there might be favourable effects on human health, due to reduced risk of transmission of zoonotic infections, and on societal perception of livestock farming. In general, improved understanding of infection transmission and pathogen adaptation is of wider relevance than just to the livestock sector. For example, for plant breeders the here developed insights and strategies might be of particular interest.

1.6 Aims and outline of this thesis

The overall aim of this thesis is to enhance insight into breeding of livestock against infectious diseases by integration of theory from animal breeding and quantitative epidemiology.

In Chapter 2, I explore the theoretical potential of genetic selection to reduce the prevalence of an infectious disease using simulations of an established epidemiological model. I simulated an endemic infectious disease in a population consisting of animals with genetic variation in their susceptibility to that disease. The size of genetic variation was tuned to commonly found heritability estimates for disease status. Then I performed selection of the best sires of this population to determine the potential response to selection. The results show that after only one round of selection prevalence of the infection was considerably reduced in the population, with indirect genetic effects as the main driver of this large response.

Chapter 3 deals with finding the theory behind the findings in chapter 2, by identifying expressions for response to selection in prevalence and for the size of the indirect genetic effect. The most important finding in this chapter is that the total breeding value for prevalence is a factor 1/prevalence greater than the breeding value for binary disease status, as usually defined by animal breeders. Thus, with prevalence lower than a half, the total response in prevalence is more than double the response expected based on the breeding values for binary disease status, which further increases for infections with lower prevalence.

In Chapter 4, I translate the results found in Chapter 3 beyond the field of animal breeding, and further explore the mechanisms behind the large response. It turns out that the ratio 1/prevalence holds more general for interventions against infectious diseases. Thus also, for example, for vaccination. Thereby, the ratio provides an easy way to predict the total effect of an intervention once the efficacy, as commonly estimated in trials, is known.

In Chapter 5, I develop a model for transmission of evolving pathogens following genetic selection of animals for increased resistance to an infection. In this model, host animals are selected for resistance to the infection, which reduces the prevalence of the infection, eventually even to zero, implying eradication of the infection. As long as infection is present, escape mutants of the pathogen might

emerge and spread through the host population. Using this model, an 'invasion window' is identified, which is the range in frequency of resistant hosts, in which escape mutants of the pathogen can emerge and invade the host population. The window is affected by the strength of host resistance, the benefit of the escape mutation in resistant hosts and the costs of the mutation in non-resistant hosts. I propose several strategies to make sure that the population stays out of the invasion window. Importantly, the results show that the change in genetic composition of the population should happen relatively fast and thus that gradual genetic change, as is for instance achieved with multi-trait selection, will not be the best strategy if resistance is part of the breeding goal.

The findings in the previous chapters show the need for proper estimates of available genetic variation in infectious disease traits. In Chapter 6, I evaluate the generalised linear mixed model (GLMM), as developed in VELTHUIS *et al.* (2003); LIPSCHUTZ-POWELL *et al.* (2014); ANCHE *et al.* (2015); BIEMANS *et al.* (2019a), on its ability to estimate genetic parameters and breeding values for infectious disease susceptibility. Furthermore, I investigate data requirements for accurate estimation with the GLMM. The results show that the GLMM is able to provide unbiased estimates and accurate breeding values for susceptibility, if the length of the observation interval is short enough, given the transmission speed of the infection.

Lastly, in the General Discussion I explore how the findings in this thesis can be used in animal breeding practice to improve the results of breeding against infectious diseases. There, attention is paid to feasibility of strategies, expected results, and to societal acceptance of the proposed strategies.

Chapter 2

Why genetic selection to reduce the prevalence of infectious diseases is way more promising than currently believed

Andries D. Hulst[1,2], Mart C.M. De Jong[1], Piter Bijma[2]

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

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

Genetics 217(4)

Abstract

Genetic selection for improved disease resistance is an important part of strategies to combat infectious diseases in agriculture. Quantitative genetic analyses of binary disease status, however, indicate low heritability for most diseases, which restricts the rate of genetic reduction in disease prevalence. Moreover, the common liability threshold model suggests that eradication of an infectious disease via genetic selection is impossible because the observed-scale heritability goes to zero when the prevalence approaches zero. From infectious disease epidemiology, however, we know that eradication of infectious diseases is possible, both in theory and practice, because of positive feedback mechanisms leading to the phenomenon known as herd immunity. The common quantitative genetic models, however, ignore these feedback mechanisms. Here we integrate quantitative genetic analysis of binary disease status with epidemiological models of transmission, aiming to identify the potential response to selection for reducing the prevalence of endemic infectious diseases. The results show that typical heritability values of binary disease status correspond to a very substantial genetic variation in disease susceptibility among individuals. Moreover, our results show that eradication of infectious diseases by genetic selection is possible in principle. These findings strongly disagree with predictions based on common quantitative genetic models, which ignore the positive feedback effects that occur when reducing the transmission of infectious diseases. Those feedback effects are a specific kind of Indirect Genetic Effects; they contribute substantially to the response to selection and the development of herd immunity (i.e., an effective reproduction ratio less than one).

2.1 Introduction

Infectious diseases are of great concern in agriculture, because they cause considerable economic damage in terms of production losses and costs of treatment. In livestock, moreover, infectious diseases harm animal health and welfare, and cause indirect cost through trade restrictions and impact on human health in the case of zoonoses (BENNETT 2003). Although combatting infectious diseases has always been a challenge, the increase in antibiotic resistance, followed by restrictions on the use of antibiotics has intensified the need for additional solutions (SPEKSNIJDER *et al.* 2015).

Genetic selection for improved disease resistance is a common strategy to combat, especially endemic, infectious diseases, as an alternative or additional intervention to classical control measures such as antibiotic treatment (BISHOP *et al.* 2010). In livestock, genetic variation in disease resistance seems to be present in virtually all species, and for a wide range of diseases (BISHOP *et al.* 2010). While this observation implies that genetic improvement of infectious disease traits is possible *in principle*, the rate of genetic improvement is restricted by the limited heritability of the disease traits.

Infectious disease traits are often measured as the binary disease status of the individual (healthy/diseased = 0/1), and heritability estimates for disease status are typically below 10 percent. For mastitis in dairy cattle, for instance, observed-scale heritability estimates range from 0.01 to 0.09, with corresponding prevalence ranging from 0.09 to 0.39 (MARTIN *et al.* 2018). Although the heritability of binary traits is sensitive to prevalence, with highest values at intermediate prevalence, the above estimates are fairly low. Given the population average value for binary disease status (p), the additive genetic variance in binary disease status is fully determined by heritability; $\sigma^2_{A_{obs}} = h^2_{obs} p(1-p)$. Thus these low heritabilities also indicate limited genetic variation in binary disease status. Because the mean value of individual disease status is equal to the prevalence of the disease, the formula for the additive genetic variance also seems to imply that the potential response to selection for reduction of disease prevalence is limited.

Higher heritabilities are found with threshold models, which assume a continuous liability underlying the observed binary disease status. (DEMPSTER AND LERNER 1950; ROBERTSON 1950). However, while threshold models are statistically much more appropriate than linear models (GIANOLA 1982), a high heritability on the liability scale does not imply a large response to selection, because response in binary disease status depends on the heritability on the observed binary scale (ROBERTSON 1950).

An important expectation following from classical quantitative genetic models is the impossibility to eradicate a disease via selection. Eradication seems impossible because additive genetic variance on the observed binary scale goes to zero when the prevalence approaches zero (ROBERTSON 1950). Hence, for breeders this implies that genetic reduction of disease prevalence becomes increasingly difficult and ultimately impossible at lower prevalence. In epidemiology however, successful eradication of several infectious diseases has been achieved using vaccination, with vaccines being neither 100% effective nor successfully applied to everyone (HALLORAN *et al.* 1992). A well-known example in livestock is the worldwide eradication of rinderpest, which has been successful without the need to vaccinate all the animals. In Somalia, for example, only about 50% of the animals had vaccination immunity for rinderpest at eradication (MARINER *et al.* 2012).

The eradication of infectious diseases using vaccination relates to a phenomenon known as "herd immunity" (FINE 1993). When a sufficient proportion of the population is immune to the disease, the disease can no longer spread in the population and thus dies out. Hence, the mechanisms determining the prevalence of an infectious disease are fundamentally different from those determining the prevalence of a non-communicable disease, such as, say, heart failure. This difference is relevant not only for the case of vaccination, but also for the results of selection for disease resistance; a reduction in disease susceptibility due to genetic selection would have the same effect on disease prevalence as a corresponding reduction in susceptibility achieved with vaccination. However, despite this fundamental difference, mainstream theory and methods in

quantitative genetics and livestock genetic improvement completely disregard the key role of transmission between individuals.

Whereas the previous paragraph stresses the importance of transmission dynamics, other explanations for the low heritability of infectious diseases have also been proposed. The low heritability of individual disease status, even at intermediate prevalence, seems to disagree with the moderate to high heritability estimates for immune response traits (KNAP AND BISHOP 2000; HENRYON *et al.* 2006; THOMPSON-CRISPI *et al.* 2012). To explain this discrepancy, BISHOP AND WOOLLIAMS (2010) argue that incomplete exposure of individuals to the infectious agent, low test specificity and sensitivity, and incomplete data recording can cause underestimation of the heritability of binary disease status. In addition, LIPSCHUTZ-POWELL *et al.* (2012) point to the potential importance of genetic variation in infectivity, defined as the propensity of infected individuals to infect others, which is not captured by the current models. Hence, these mechanisms might also partly explain why not all genetic variation is captured with the current breeding methods.

Nevertheless, while more accurate disease data and accounting for genetic variation in infectivity might indeed improve the response to selection (TSAIRIDOU *et al.* 2019), these factors cannot explain why epidemiological models that account for transmission dynamics often suggest substantially greater response to selection than genetic models, even in the absence of measurement errors and when genetic variation is in disease resistance only (*i.e.*, in the absence of genetic variation in infectivity). Using an epidemiological model tailored to foot rot in sheep, for example, NIEUWHOF *et al.* (2009) have shown that selection for resistance reduces prevalence faster than predicted by common quantitative genetic models of disease status. This suggests that the ordinary breeding values for individual disease status do not correctly predict prevalence in the offspring generation. Empirical examples of a relatively large reduction in disease prevalence resulting from genetic selection were found for infectious pancreatic necrosis (IPN) in Atlantic salmon and for clinical mastitis in dairy cows. The salmon industry managed to decrease IPN mortality with over 70% in two years, using marker assisted selection for an IPN-resistance QTL (AQUAGEN 2012).In a selection

experiment, HERINGSTAD *et al.* (2007) observed a phenotypic decrease in the prevalence of clinical mastitis of 15% after 5 generations of single trait selection against clinical mastitis, while the reduction predicted by estimated breeding values was only 8%. Although this phenotypic response could be explained by environmental changes (e.g. improvements in management), a correlated response in clinical mastitis from single trait selection on higher protein yield observed in the same experiment contradicts this explanation. Selection for higher protein yield resulted in a phenotypic increase in clinical mastitis prevalence from 10% to 25%, while the increase predicted by EBVs was only 2%. Hence, an environmental decreasing effect on clinical mastitis is highly unlikely, because this disagrees with the much higher phenotypic increase observed after selection for higher protein yield. Results of HERINGSTAD *et al.* (2007), therefore, indicate greater genetic response than predicted by EBVs from classical quantitative genetic models.

The discrepancy between predictions based on epidemiological *vs.* genetic models illustrates that the relationship between the heritability of binary disease status and the amount of genetic variation that can actually be used to reduce disease prevalence is still unclear, even for the most basic and well-established epidemiological models. As a consequence, prospects of genetic selection to reduce infectious disease prevalence are not correctly predicted by current quantitative genetic models.

Here we show that the low heritabilities of binary disease status are fully consistent with a large genetic variation in disease resistance. Heritability estimates from common quantitative genetic models by no means represent the variation available to breeders for genetic improvement, even for the simplest epidemiological model and with genetic variation in disease resistance only. Consequently, the potential for genetic selection to reduce the prevalence of infectious diseases will be considerably larger than currently believed, even when the conventional breeding values for disease status are used as selection criterion. This occurs because the effect of genetic selection on the disease status of individuals acts in two ways: Both directly via a reduced susceptibility of the

individuals, and indirectly via reduced exposure of individuals to infected herd mates, because of a lower susceptibility of those herd mates.

2.2 Material and Methods

To quantify the relationship between the heritability of binary disease status and the potential response to genetic selection for lower prevalence, we will first find the genetic variation in disease susceptibility required to reproduce the common heritability values of disease status. For this purpose, we will use simulation of an endemic infectious disease with an epidemiological model that is appropriate for many endemic infectious diseases at the farm level, while keeping the mathematics as simple as possible. We aim to find the epidemiological parameters and genetic variances that result in observed-scale heritabilities for binary disease status of 0.02, 0.05 and 0.10, using a common prevalence of endemic infectious diseases in livestock. After identifying the epidemiological parameters and genetic variances corresponding to those heritabilities, we will estimate the potential response to selection using those parameters and variances as input.

Because the common linear models capture genetic variation in disease resistance only, we simulated individuals to differ only in disease resistance. For reasons of mathematical convenience and consistency with epidemiological models, we in fact modelled disease susceptibility instead of resistance. Susceptibility is the (relative) probability an individual becomes infected given exposure (DOESCHL-WILSON AND KYRIAZAKIS 2012), and is the opposite of disease resistance; individuals with a high susceptibility have a low resistance, and *vice versa.*

This section starts with a description of the epidemiological model we used for simulation of the endemic disease. This description initially ignores variation in susceptibility, to make the model more easily understandable. Next we describe how individual genetic variation in susceptibility can be introduced into the epidemiological model. Then we describe how we simulated a population with variation in susceptibility, followed by a description of the implementation and the

scenarios used to quantify response to selection. Table 2.1 provides a notation key.

Table 2.1 Notation key and overview of input values.

SYMBOL	MEANING	SIMULATED VALUE(S)
β	Transmission rate parameter (t^{-1})	0.03 day^{-1}
α	Recovery rate parameter (t^{-1})	0.02 day^{-1}
S	Susceptible state of an individual	-
I	Infected state of an individual	-
S (in italics)	Number of susceptible individuals (#)	-
I (in italics)	Number of infected individuals (#)	-
N	Herd size (#)	102
$\frac{I}{N}$	Prevalence	0.33
β_i	Transmission rate parameter for individual *i* (t^{-1})	$\beta\gamma_i$ day^{-1}
γ_i	Individual susceptibility	$e^{A_{\gamma_i}+E_{\gamma_i}}$
A_{γ_i}	Individual additive genetic effect (breeding value) for the logarithm of susceptibility	$\sim N(0, \sigma^2_{A,\gamma})$
E_{γ_i}	Individual permanent environmental effect for the logarithm of susceptibility	$\sim N(0, \sigma^2_{E,\gamma})$
$\sigma^2_{A_\gamma}$	Simulated additive genetic variance for the logarithm of susceptibility	-
$\sigma^2_{E_\gamma}$	Simulated permanent environmental variance for the logarithm of susceptibility	-
a^2_P	Chosen additive genetic proportion of the full permanent variance in log-susceptibility among individuals.	0.25; 0.5; 0.75; 1
$h^2_{\gamma,liab}$	Liability-scale heritability for the logarithm of susceptibility	-
h^2_{Obs}	Target observed-scale heritability of binary disease status	0.02; 0.05; 0.1

2.2.1 Susceptible-Infectious-Susceptible model without genetic variation

To illustrate our point with a minimum of mathematical detail, we used one of the simplest epidemiological models for endemic infectious diseases, the so called Susceptible-Infectious-Susceptible (SIS) model (HETHCOTE 1989). Although the model is simple, it provides a realistic representation of the transmission of several endemic infectious diseases in livestock populations and is well-established in veterinary epidemiology for that reason. Note that in the context of a SIS-model, "susceptible" merely means that an individual is not infected and can in principle become infected; it does not indicate high susceptibility or low disease resistance.

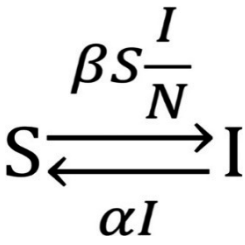


Figure 2.1 Susceptible-Infected-Susceptible-Model.

In the SIS-model, individuals can be in one of two states: susceptible (S) or infected and then also infectious (I; Figure 2.1). In epidemiology, the symbols S and I are generally used to indicate both the state of an individual and the total number of individuals in that state. To prevent confusion, we will use *S* and I (*in italics*) to indicate the number of susceptible and infected individuals in a herd, and S and I for the state of an individual. Thus, *S* indicates the number of individuals with disease state S in a herd. The infection is endemic within the separate herds, and we assume that transmission can occur only between herd mates. *N* is the total number of individuals in a herd, which is equal to the sum of the number of susceptible and infected individuals in the SIS-model ($S + I = N$).

Transitions in individual state, so-called events, are possible in both directions (from S to I and from I to S). Susceptible individuals can become infected through contacts with infected herd mates, while infected individuals can recover and thus become susceptible again. The types of events (*i.e.,* transmission or recovery) and the bookkeeping of the individuals involved in these events define our

stochastic model. The average number of susceptible individuals that becomes infected per unit of time (*e.g.*, per day) depends on the total number of susceptible individuals in the herd at that moment, the fraction infected among their herd mates ($\frac{I}{N}$, *i.e.* the prevalence of the disease), and the transmission rate parameter of the disease, β (DIEKMANN *et al.* 2012):

$$Rate\ \mathrm{S} \rightarrow \mathrm{I} = \beta S \frac{I}{N} \tag{2.1}$$

For a population of N = 100 individuals in the endemic state with, for example, a prevalence of 0.1 ($I = 10,\ S = 90$), and a β of 0.03, on average $0.03 \cdot 90 \cdot 0.1 = 0.27$ susceptible individuals will become infected each day.

The number of recovering individuals per unit of time depends on the number of infected individuals in the herd at that moment (I) and on the recovery rate parameter, α:

$$Rate\ \mathrm{I} \rightarrow \mathrm{S} = \alpha I \tag{2.2}$$

The recovery rate parameter α also determines the average duration of the infectious period, which is equal to $\frac{1}{\alpha}$. Continuing the previous example of a population of 100 individuals and a prevalence of 0.1 ($I = 10$), with an α of 0.02, each day on average $0.02 \cdot 10 = 0.20$ infected individuals will recover and thus become susceptible again.

Combining transmission and recovery in the above example shows that in this situation the average number of infected individuals increases, and the number of susceptible individuals decreases (*i.e.,* 0.27 > 0.20). This continues until the prevalence reaches 0.33; after which the average number of infected and susceptible individuals remains constant ($\beta S \frac{I}{N} = 0.03 \cdot 67 \cdot 0.33 = 0.66$, $\alpha I = 0.02 \cdot 33 = 0.66$). The endemic disease is then said to be in equilibrium, the so-called pseudo-steady state in stochastic models. The SIS-model tends to such a dynamic equilibrium, in which the average number of infections is equal to the average number of recoveries. The actual number of susceptible and infective individuals fluctuates around this equilibrium because we have a stochastic model in a finite

population. The expected prevalence of an endemic disease is equal to the average proportion of infected individuals in the equilibrium.

Thus, at equilibrium, the transmission rate equals the recovery rate:

$$\beta S \frac{I}{N} = \alpha I \tag{2.3}$$

The equilibrium prevalence, therefore, follows from solving Equation 2.3 for $\frac{I}{N}$.Using the fact that $S = N - I$, yields:

$$P_{equilibrium} = \frac{I}{N} = 1 - \frac{\alpha}{\beta} = 1 - \frac{1}{R_0} \tag{2.4}$$

where $R_0 = \beta/\alpha$, which is the so-called basic reproduction ratio (DIEKMANN *et al.* 1990). The prevalence is thus only a function of the ratio of α and β and can therefore also be expressed as a function of the basic reproduction ratio. With variation among individuals in susceptibility (introduced in next paragraph), Equation 2.4 is no longer exactly equal to the average prevalence (BIEMANS *et al.* 2017).

R_0 is a key parameter in epidemiology. It represents the average number of susceptible individuals that get infected by a single average infectious individual in a totally susceptible population. Hence, R_0 has a threshold value of 1: if $R_0 > 1$ a single infected individual will infect initially more than one new individual, so that the infection can spread in the infection-free population after introduction. On the other hand, if $R_0 < 1$, an infectious disease will die out. Hence, also for an endemic disease to persist, R_0 must be greater than one. Moreover, any measure that reduces R_0 to a value smaller than 1 results in eradication of the infectious disease. Obviously, Equation 2.4 does not apply when $R_0 < 1$. In that case, the disease is absent and prevalence at equilibrium is zero.

2.2.2 Introducing variation in susceptibility into the SIS-model

Here we describe how variation in disease susceptibility can be introduced among individuals. With equal exposure to infected herd mates, individuals with higher susceptibility are more likely to become infected. Because we will only simulate variation in susceptibility, we define an individual transmission rate parameter for the recipient animal, following (ANCHE *et al.* 2014):

$$\beta_i = \beta\gamma_i \tag{2.5}$$

where β is the transmission rate parameter for the average susceptible individual (infectiousness being the same for all individuals), and γ_i the (relative) susceptibility of individual *i*. For the average individual, $\gamma_i = 1$, so that β_i is equal to β. Individuals with above-average susceptibility have $\gamma_i > 1$, such that their β_i is greater than β, which means that they have a higher than average probability to become infected (given equal exposure). Accordingly, individuals with below average susceptibility have $\gamma_i < 1$, $\beta_i < \beta$, and a lower than average probability to become infected. Note that γ_i cannot be lower than 0, since this would result in a negative transmission rate.

Replacing β by β_i in Equation 2.1, and using $S = 1$ representing the single individual *i*, yields the transmission rate for individual *i* exposed to all individuals in the population, of which a fraction $\frac{I}{N}$ is infectious:

$$Rate\ \mathrm{S}_i \rightarrow \mathrm{I}_i = \beta_i \frac{I}{N} = \beta\gamma_i \frac{I}{N}. \tag{2.6}$$

We know from the previous paragraph that individual susceptibility (γ_i) must be nonnegative and have a value of 1 for the average individual. For this reason, we simulated γ_i from a log-normal distribution. The log-normal distribution is nonnegative and has positive-skewness, which is often observed for disease traits (LLOYD-SMITH *et al.* 2005). Thus, following ANACLETO *et al.* (2015), we simulated an additive model with normally distributed random effects for the logarithm of susceptibility, so that susceptibility of individual *i* is:

$$\gamma_i = e^{A_{\gamma_i} + E_{\gamma_i}} \tag{2.7}$$

where A_{γ_i} denotes the individual additive genetic effect, or breeding value for the logarithm of susceptibility, and E_{γ_i} the permanent individual environmental effect for the logarithm of susceptibility, with $A_\gamma \sim N(0, \sigma^2_{A_\gamma})$ and $E_\gamma \sim N(0, \sigma^2_{E_\gamma})$. Thus $\overline{A_\gamma} = \overline{E_\gamma} = 0$, as common in quantitative genetics. Consequently, $\gamma_i = 1$ for an individual with the mean A_γ and E_γ (but $\bar{\gamma}$ is not exactly 1, because of the positive skewness).

For brevity, we will refer to A_{γ_i} as the 'breeding value for susceptibility' or just 'breeding value' and to E_{γ_i} as the 'individual permanent environmental value for susceptibility' or just 'permanent environmental value'. Beware though that A_{γ_i} and E_{γ_i} refer to the logarithm of individual susceptibility.

We included a permanent environmental effect for susceptibility, E_γ, to allow for a similarity between repeated records on the same individual that is not purely genetic, but also environmental (*e.g.* because of developmental effects). This relates to repeatability models as discussed in *e.g.* FALCONER AND MACKAY (1996). We use the symbol a^2_P to denote the additive genetic proportion of the full permanent variance in log-susceptibility among individuals,

$$a^2_P = \frac{\sigma^2_{A_\gamma}}{\sigma^2_{A_\gamma} + \sigma^2_{E_\gamma}} \tag{2.8}$$

2.2.3 Simulation of a population with variation in susceptibility

We simulated a population with individual variation in susceptibility. The simulated population consisted of two generations: parents and offspring. We simulated offspring breeding values according to the infinitesimal model (FISHER 1918):

$$A_{\gamma_i} = \frac{1}{2} A_{\gamma_{sire_i}} + \frac{1}{2} A_{\gamma_{dam_i}} + MS_{\gamma_i} \tag{2.9}$$

in which $A_{\gamma_{sire}}$ and $A_{\gamma_{dam}}$ were simulated from Normal distributions as shown above, and MS_{γ_i} denotes the Mendelian sampling term, sampled from a normal distribution $MS_{\gamma_i} \sim N(0, \frac{1}{2}\sigma^2_{A_\gamma})$. The additive genetic variance for susceptibility was

obtained using an iterative procedure tuned to the desired value of observed heritability (see subsection Implementation). The permanent environmental variance in susceptibility was calculated from the additive genetic variance and the chosen additive proportion (a_P^2; Equation 2.8), $\sigma_{E_\gamma}^2 = \sigma_{A_\gamma}^2(1 - a_P^2)/a_P^2$.

2.2.4 Implementation

Our simulations consisted of the iteration of four consecutive steps, aiming to find the additive genetic standard deviation in susceptibility (σ_{A_γ}) that is required to arrive at the desired heritability value of disease status at the observed scale:

1) Stochastic simulation of 20 replicates of a population with a certain genetic standard deviation (σ_{A_γ}) and relative magnitude of genetic permanent effects ($\mathrm{a_P^2}$) for the logarithm of susceptibility.

2) Stochastic dynamic simulation of an endemic infectious disease in herds of the population and recording of binary disease status data for all animals in all replicates.

3) Estimation of the heritability of disease status ($\mathrm{h_{Obs}^2}$) from the recorded disease data with a linear mixed model for all replicates.

4) Change σ_{A_γ} by 0.05 and rerun the entire simulation starting at step 1 if the mean $\mathrm{h_{Obs}^2}$ of the 20 replicates is not within two standard errors from either 0.02, 0.05 or 0.1.

Figure S2 provides a schematic overview of the above steps.

R (R CORE TEAM 2020) was used for simulation of the population and the infectious disease (steps 1 and 2). The code is available in supplemental files S3 (population simulation) and S4 (infectious disease simulation). The endemic disease was simulated using the Gillespie algorithm (GILLESPIE 1976), which is the standard procedure for stochastic simulation of infectious diseases (KEELING AND ROSS 2008). The algorithm simulates the successive infection and recovery events (see section SIS-model without genetic variation) in two steps:

a) Monte Carlo sampling of the time to the next event,

b) Monte Carlo sampling of the type of event (either infection or recovery) and the individual involved.

The probability of sampling an event and the individual involved depends on the individual transmission rates (β_i) and the population recovery rate (α) (see section Introducing variation into the SIS-model) . More details are provided in Appendix 2.7.2.

We started the simulations by infection of a fraction of randomly chosen individuals in each herd. To speed up the simulations and to prevent early dying out of the disease by chance (this occurs when no infected individuals are left in a herd, $I = 0$), we started the simulations near the equilibrium prevalence. To prevent an effect of the initially infected individuals on the parameter estimation, we ran the simulation with a burn-in period of 500 days without recording of data for parameter estimation. The recording period was 1000 days, such that the total simulated period was 1500 days. We chose this relatively long period to allow for repeated records on disease status, which are essential for the estimation of genetic parameters, given the highly dynamic nature of infectious diseases (DOESCHL-WILSON *et al.* 2014; BIEMANS *et al.* 2017). For the parameter estimation, the disease status of all individuals in each herd was recorded monthly, resulting in 33 records per individual.

In step 3, we estimated the heritability of binary disease status with a linear mixed model including a random animal effect with a pedigree-relationship matrix (ID_i), fitted with ASReml 4.1 (GILMOUR *et al.* 2015). Next to the random genetic effect, the model contained a random herd effect ($Herd_k$) and a random non-genetic animal effect (a so-called permanent effect) to account for repeated records on the same individual (IDR_i),

$$\mathrm{y_{ikt}} = \mu + Herd_k + ID_i + IDR_i + e_{ikt}, \qquad (2.10)$$

where y_{ikt} is the t^{th} binary disease record (0/1) of individual *i* present in herd *k*, μ is the overall mean prevalence of the disease in the population, e_{ikt} is the residual variance at the observed scale.

2.2.5 Input values

In our simulations, the prevalence ($\frac{I}{N}$) and the mean duration of the infectious period ($\frac{1}{\alpha}$) were set to 0.02 day^{-1}, so that the average duration of the infectious period was 50 days, consistent with BIEMANS *et al.* (2018) This was done to reflect a common endemic disease in livestock: digital dermatitis (DD), a bacterial infectious claw disorder in cattle. The prevalence was set to 0.33, representing herds in the Netherlands that suffer considerably from DD (HOLZHAUER *et al.* 2006). To obtain a prevalence of 0.33, β was set to 0.03, based on Equation 2.4. Note that R_0 was thus 1.5, as can be calculated from the prevalence or from $\frac{\beta}{\alpha}$.

In part of the scenarios with a high desired value for $\mathrm{h}^2_{\mathrm{Obs}}$, a large variance in susceptibility was needed. In these scenarios, the positive skewness of the lognormal distribution caused the population mean susceptibility ($\overline{\gamma_i}$) to be (much) larger than 1. Since γ acts as a scaling factor on β, $\overline{\gamma_i} > 1$ resulted in $\overline{\beta_i} > \beta$ (Equation 2.5), which in turn resulted in a prevalence higher than the desired value of 0.33 (Equations 2.4 and 2.6). To avoid inconsistencies in prevalence between scenarios, the β was iteratively corrected for each combination of $\mathrm{h}^2_{\mathrm{Obs}}$ and $\mathrm{a}^2_{\mathrm{P}}$, such that the prevalence was 0.33 in all scenarios. This issue is further addressed in the discussion.

The simulated population consisted of a parental generation of 102 sires, each mated to 102 dams, with one offspring per mating, resulting in a total of 10404 offspring with a half-sib family structure. The offspring generation was kept in 102 herds of size 102, each consisting of offspring of six randomly sampled sires, each sire contributing 17 offspring. This structure is somewhat similar to a dairy cattle population, but more balanced. The disease was simulated in the offspring generation only.

The relative magnitude of genetic permanent effects ($\mathrm{a}^2_{\mathrm{P}}$) was set to 0.25, 0.50, 0.75 or 1. Since practically nothing is known about the relative magnitude of genetic and non-genetic permanent effects on susceptibility, we simulated values of a^2_P in the range of 0.25 to 1. Repeatabilities thereby range from equal to the heritability to the 4-fold of the heritability.

2.2.6 Estimation of response to selection

The above procedure results in a genetic standard deviation in susceptibility (σ_{A_γ}) that corresponds to heritabilities of disease status of 0.02, 0.05 or 0.10. Next, we used this σ_{A_γ} and the same epidemiological model to investigate the potential response to selection in disease prevalence for this range of observed heritabilities. To do so, we performed additional simulations in which we selected the six best out of the 102 sires, based on their true (*i.e.,* simulated) breeding value (TBV) for susceptibility. Each of these six sires was mated to 1734 randomly chosen dams, resulting in a total of 10404 offspring, which were allocated randomly to 102 herds. Thus, there was selection in sires only. To better illustrate the effect of selection, the starting prevalence in the offspring generation was kept at 0.33, even though this was higher than the expected prevalence in that generation (since $\overline{\gamma_i} < 1$ after selection). To show trends in the number of infected individuals over time, we used daily records of the entire 1500 days. Hence, the simulations will illustrate the gradual decrease of prevalence from the initial 0.33 to the new equilibrium value.

2.2.7 Heritability on the underlying scale

The heritability of binary disease status can be defined on two different scales: the observed scale of the binary record, and the underlying scale of the linear additive model for the genetics of susceptibility. The latter is the so-called liability scale (DEMPSTER AND LERNER 1950; ROBERTSON 1950). In our simulations, we estimated the observed-scale heritability for binary disease status, but the heritability of the underlying susceptibility was unknown (note that we did not simulate residual variance in susceptibility (Equation 2.7). To provide a heritability value for susceptibility, we calculate heritability on the liability scale in this section. Because our model is additive for log-susceptibility, the liability scale refers to the logarithm of susceptibility. Note, we do not describe new simulations or analyses in this section, but merely motivate the calculation of a liability-scale heritability for our model.

Binary data (y) can be generated from an underlying linear model in two different ways: using a threshold model and using a generalized linear (mixed) model (GLMM). First, in the threshold model, a linear model is used to specify an individual liability, including a residual. Individuals with a liability value greater than a pre-defined threshold have $y = 1$; the others have $y = 0$. Hence, given the threshold, an individual's liability fully determines its observed binary record. Second, in the GLMM, a linear model without a residual is combined with a link function to specify the probability that $y = 1$. Subsequently, the binary records are sampled from a Bernoulli distribution with this probability. Hence, in this approach, the linear model specifies the expectation of the binary record, $E(y) = P(y = 1)$; not the actual record. Both models are fully equivalent when the link function of the GLMM is the cumulative density function (cdf) that corresponds to the probability density function (pdf) of the residual of the threshold model (De Villemereuil *et al.* 2016). In the classical threshold model, for example, the residual on the liability scale follows a standard normal pdf, while the probit link of a GLMM corresponds to the standard normal cdf. Thus the classical threshold model is equivalent to a GLMM with a probit link (See e.g. Supplementary Information A in De Villemereuil *et al.* (2016)). The link function in a GLMM "replaces" the liability residual in the threshold model. For this reason, the residual liability variance is also known as the "link variance" (De Villemereuil *et al.* 2016). However, a liability-scale heritability can be defined only based on the threshold model, because the underlying linear model in a GLMM has no residual. (In other words, a GLMM has no "phenotypic" variance on the underlying scale, so the denominator of heritability cannot be determined).

We simulated binary disease status data by specifying the probability of events (infection *vs.* recovery; see step 2 and Appendix 2.7.2) using the rates for these events as the rates of a Poisson process (for the infection event this is $1 - e^{-\beta\gamma_i\frac{I}{N}}$), which is the GLMM approach. Thus, to find the heritability on the liability scale, we have to translate our model into the corresponding threshold model. GLMM come with different link functions, each relating to a different pdf of the residual of a corresponding threshold model. As our binary disease data are generated by a Poisson process, the appropriate link function is the complementary log-log. (This is a standard result for GLMM (McCullagh 2018), and motivated in detail for binary disease data in Anche *et al.* (2015)). Thus, we have to find the residual

liability variance of the threshold model that corresponds to a GLMM with a clog-log link function. The link variance of a GLMM with a cloglog link follows from the Gumbel distribution and equals $\frac{\pi^2}{6}$ (Appendix 2.7.1, NAKAGAWA *et al.* (2017)).

Therefore, we define the liability for the j^{th} record of individual *i* as

$$l_{ij} = A_{\gamma_i} + E_{\gamma_i} + e_{l_{i,j}}$$

where $var\left(e_{l_{i,j}}\right) = \frac{\pi^2}{6} \approx 1.64$, and heritability on the underlying liability scale of log-susceptibility equals:

$$h^2_{\gamma,liab} = \frac{\sigma^2_{A_\gamma}}{\sigma^2_{A_\gamma} + \sigma^2_{E_\gamma} + \frac{\pi^2}{6}} \tag{2.11}$$

2.3 Results

This section starts with the genetic variances and corresponding liability heritabilities for the logarithm of susceptibility that correspond to observed heritabilities of 0.02, 0.05 and 0.10 for disease status. Subsequently, we show to what extent individual breeding values for susceptibility are reflected in individual disease status. Next, we show the response to selection for each value of the observed heritability. Finally, we illustrate the mechanisms underlying the observed response to selection, using simulations of herds consisting of individuals with similar susceptibility.

2.3.1 Genetic variation in susceptibility

Table 2.2 shows the genetic standard deviations in the logarithm of susceptibility (σ_{A_γ}) and liability heritabilities ($h^2_{\gamma.liab}$) that correspond to observed scale heritabilities of binary disease status (h^2_{obs}) of 0.02, 0.05 and 0.1. Results are shown for additive genetic proportions of the permanent variance (a^2_P) of 1, 0.75, 0.5 and 0.25. To illustrate the magnitude of the genetic variation in susceptibility, Table 2.2 also shows the genetic values for R_0 for the 10% of individuals with highest breeding values and the 10% individuals with the lowest breeding values

for susceptibility. These R_0-values are relevant for the prevalence of the infectious disease (see Equation 2.4) and give an indication of the potential response to selection. In supplementary Table S1 actual h^2_{obs} estimates for the different σ_{A_γ} are given with corresponding standard errors.

Table 2.2 shows that common observed scale heritability values for binary disease status correspond to a substantial genetic variation in the logarithm of susceptibility (σ_{A_γ}). An observed scale heritability of 0.02, for example, corresponds to a genetic standard deviation in susceptibility of 0.3. With a σ_{A_γ} of 0.3, the mean breeding value of the 10% individuals with lowest breeding values for susceptibility (A_γ^-) is –0.53, accordingly, the mean breeding value of the 10% individuals with highest breeding values (A_γ^+) is 0.53. Corresponding genetic susceptibility values ($\gamma^- = e^{A_\gamma^-}, \gamma^+ = e^{A_\gamma^+}$) for these individuals are 0.59 and 1.70, which results in values for R_0 of 0.59*1.5=0.88 and 1.7*1.5=2.54 respectively. There is thus an almost 3-fold genetic difference in R_0.between the highest and lowest 10% of individuals. This difference is substantial in itself, but more importantly, R_0 for the 10% individuals with lowest susceptibility is below 1, which means that an individual with such susceptibility, on average, infects less than 1 herd mate. Implying that an infectious disease will die out. For higher values of observed heritability, the corresponding genetic standard deviation increased considerably, resulting in values for R_0 even further below 1 for a h^2_{obs} of 0.05 and 0.10.

The relative magnitude of genetic vs. environmental permanent effects on susceptibility (a_p^2) only had a minor effect on the observed heritability, as illustrated by the similar values of the genetic standard deviation in susceptibility needed to obtain a certain observed heritability. The effect of a_p^2 was substantial only for an observed scale heritability of 0.10. In this case, we needed a considerably higher genetic standard deviation in susceptibility for lower values of a_p^2, especially for a $a_p^2 = 0.25$. Similarly, the liability heritability is quite consistently, approximately twice the value of h^2_{obs}, across the different values of a_p^2.

In conclusion, a remarkably large genetic variation in susceptibility is needed to reproduce common observed scale heritabilities for disease status. As a

consequence, a considerable proportion of the population had genetic R_0 values smaller than one.

2.3.2 The relationship between susceptibility and individual disease status

Figure 2.2 shows the relationship between breeding value for susceptibility and disease prevalence (over all herds) when herds are composed of offspring of random sires. The figure shows the prevalence over time (mean of 20 replicates) in the 10% most and 10% least susceptible individuals in the population, for $a_p^2 = 1$. Results were very similar for other values of a_p^2, as expected given the minor effect shown above.

The distance between two lines increases with h_{obs}^2, because a higher h_{obs}^2 corresponds to a greater genetic variation in susceptibility. For an observed heritability of 0.02, the prevalence in the 10% most susceptible individuals is about a factor of 2 higher than in the 10% least susceptible individuals (25% vs. 50%). This difference of about 25 percentage points is considerable given the very low heritability. For an observed heritability of 0.10, the distance increases up to a factor of 4 (15% vs. 60%). Hence, though a heritability of 0.10 is only a moderate value, the corresponding difference in prevalence is large.

A comparison of Figure 2.2 and Table 2.2 shows that the actual prevalence in the 10% most and least susceptible individuals (Figure 2.2) is considerably different from the prevalence that is expected based on the values for R_0 (Table 2.2). For a heritability of 0.10, for example, the actual prevalence is about 0.6 in the 10% most susceptible and about 0.15 in the 10% least susceptible individuals, while the expected prevalence based on R_0 is 0.82 and 0 respectively (see Table 2.2 and Equation 2.4). This indicates that the genetic susceptibility values are not completely reflected in the disease status of individuals. This phenomenon occurs because individual disease status depends not only on individual susceptibility, but also on exposure to infectious herd mates. With randomly composed herds, individuals experience similar exposure to infectious herd mates, irrespective of their own susceptibility. For this reason, differences in prevalence between subgroups in the population were smaller than expected based on their genetic value for R_o.

Table 2.2 Genetic standard deviation in susceptibility required for observed scale heritability of disease status of 0.02, 0.05 and 0.10.

$\mathbf{h^2_{obs}}$	0.02				0.05				0.10			
$\mathbf{a^2_P}$	σ_{A_γ}	$h^2_{\gamma,liab}$	$R_{0_{-10\%}}$	$R_{0_{+10\%}}$	σ_{A_γ}	$h^2_{\gamma,liab}$	$R_{0_{-10\%}}$	$R_{0_{+10\%}}$	σ_{A_γ}	$h^2_{\gamma,liab}$	$R_{0_{-10\%}}$	$R_{0_{+10\%}}$
1.0	0.30	0.05	0.88	2.54	0.50	0.13	0.62	3.62	0.75	0.25	0.40	5.62
0.75	0.30	0.05	0.88	2.54	0.50	0.13	0.62	3.62	0.75	0.23	0.40	5.62
0.50	0.30	0.05	0.88	2.54	0.50	0.12	0.62	3.62	0.85	0.23	0.34	6.70
0.25	0.30	0.04	0.88	2.54	0.55	0.11	0.57	3.95	1.10	0.19	0.22	10.4

Required genetic standard deviation in the logarithm of susceptibility (σ_{A_γ}), liability heritability for susceptibility $h^2_{\gamma,liab}$ and R_0 for the 10% of individuals with highest ($R_{0_{+10\%}}$) and 10% with lowest ($R_{0_{-10\%}}$) values for susceptibility for each combination of a^2_P(rows) and h^2_{obs} (columns).

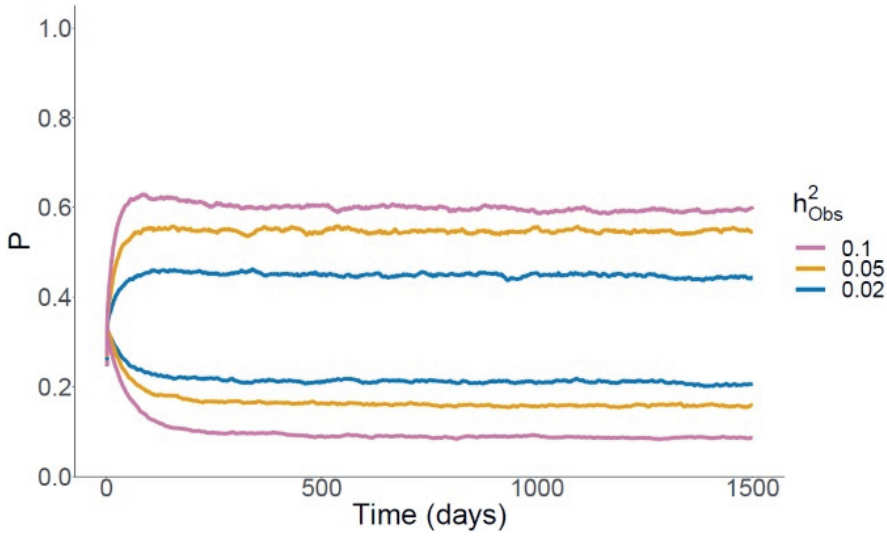


Figure 2.2 Population prevalence with random herds. Prevalence (P) of the infectious disease in the total population (over all herds) for a_p^2 of 1 and different h_{obs}^2. Lines indicate the 10-percent of individuals with highest and the 10-percent with lowest breeding value for susceptibility for the respective h_{obs}^2, herds formed by random allocation of sires.

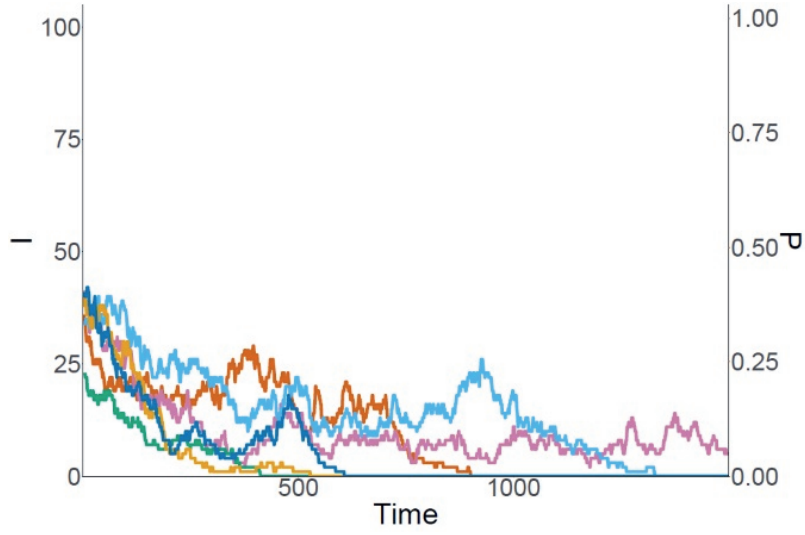


Figure 2.3 Prevalence in six herds after sire selection. Number of infected individuals (I) and prevalence (P) over time in six representative herds, represented by the different colors, consisting of offspring from six sires selected for low susceptibility. $h_{obs}^2$of 0.05, a_p^2 of 1. Each herd consists of 102 individuals.

2.3.3 Response to selection:

To investigate the effects of selection for lower susceptibility on the prevalence of the disease, the six sires with the lowest true breeding values for susceptibility were selected to produce all offspring. Figure 2.3 illustrates the effect of this single generation of sire selection on the prevalence in six herds for the scenario with a $a_p^2 = 1, (h_{\gamma,liab}^2 = 0.13)$ and a $h_{obs}^2 = 0.05$. Herds were selected such that they represent the different prevalence patterns we observed for this scenario in the simulations. The figure shows that the prevalence in all herds fluctuates around a decreasing trend from the starting prevalence of 0.33. Eventually, the infection 'dies out' in some herds, which means that there are no infectious individuals anymore to sustain infection.

For a more extensive investigation of response to selection, Figure 2.4 shows the proportion of infection-free herds over time (mean of 20 replicates), for all 12 scenarios. For each scenario, we selected the six sires with the lowest true breeding value for susceptibility to produce all offspring. The figure clearly shows that the disease disappeared from a substantial proportion of herds. Even when the observed heritability was only 0.02, 20 to 50 of the 102 herds became infection-free after a single generation of sire selection.

Response was larger with higher h_{obs}^2; for the highest value ($h_{obs}^2 = 0.10, a_p^2 \neq 0.25$), the infection was eliminated from practically all herds. Again, the relative magnitude of genetic permanent effects (a_P^2) only had a considerable effect when it was 0.25. The lines for the other values of a_P^2 are very close to each other for a given h_{obs}^2. The response for $a_P^2 = 0.25$ and $h_{obs}^2 = 0.10$ seems not in accordance with this general pattern. For example, it is lower than the response for $a_P^2 = 0.25$ and $h_{obs}^2 = 0.05$. This is most likely a consequence of the very large genetic variance in susceptibility required to reproduce a heritability of disease status of 0.10 for that scenario, together with the lognormal distribution of γ (see Discussion).

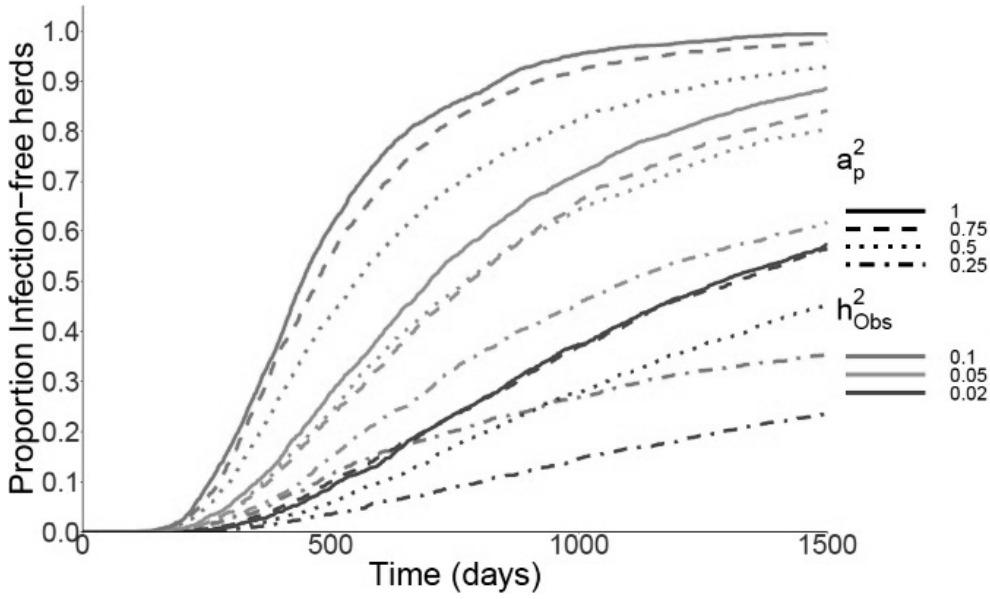


Figure 2.4 Infection free herds after sire selection. Proportion of infection-free herds over time for all combinations of h^2_{obs} and a^2_p, herds consisting of offspring from six sires selected for low susceptibility. Different colors represent different h^2_{obs}, different line types represent different a^2_p. The initial prevalence was 0.33 in each herd.

2.3.4 The mechanism underlying response to selection:

Results in Figure 2.4 show that the responses to selection are considerably greater than the differences in disease status between the 10% most and 10% least susceptible individuals within a generation (Figure 2.2). We hypothesize that this difference originates from positive feed-back effects in the transmission dynamics, resulting in some degree of herd immunity in the results in Figure 2.4. To investigate this hypothesis and clarify the mechanism underlying these unexpectedly large responses to selection, we simulated a population where individuals were grouped into herds based on their breeding value for susceptibility. The first herd consisted of the 102 least susceptible individuals, the second herd of the 102 individuals with second lowest breeding values for susceptibility, and so forth, until the last herd, consisting of the 102 most susceptible individuals. The herds consisting of the least susceptible individuals resemble the herds after selection, since herds consist entirely of individuals with

low susceptibility in both cases. Hence, similar to the case with selection, individuals with low susceptibility are accompanied by herd mates with low susceptibility, which reduces their exposure to the infectious agent.

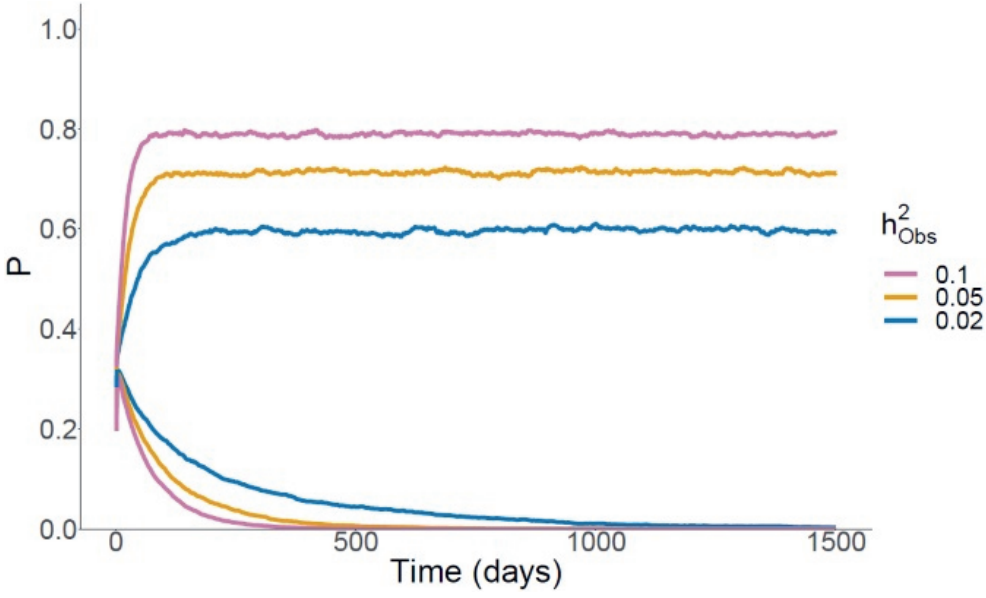


Figure 2.5 Population prevalence with non-random herds. Prevalence (P) of the infectious disease in the total population (over all herds) for a_p^2 of 1 and different h_{obs}^2 Lines indicate the 10-percent of individuals with highest and the 10-percent with lowest breeding value for susceptibility for the respective h_{obs}^2, herds formed by allocation of individuals based on breeding value for susceptibility.

Figure 2.5 shows the prevalence in herds composed of the 10% most and 10% least susceptible individuals based on breeding value for susceptibility. Comparison to Figure 2.2 shows that this formation of herds substantially increases the difference in prevalence between the most and least susceptible individuals. For an $\mathrm{h}^2_{\mathrm{obs}}$ of 0.02, for example, the difference increased from 25% vs. 50% in random herds to 5% vs. 60% in sorted herds. For an $\mathrm{h}^2_{\mathrm{obs}}$ of 0.10, it increased from 15% vs. 60% to approx. 0% vs. 80%. The prevalence in top and bottom individuals in Figure 2.5 is more in accordance with the expected prevalence based on the values for R_0 (Table 2.2). The bottom lines in Figure 2.5, which represent the herds consisting of the least susceptible individuals, show a

clear decreasing prevalence over time, and reach zero for all three observed heritabilities. This fully agrees with the R_0 values below 1 for these individuals.

This observation clearly shows that the susceptibility of herd mates has a considerable effect on the disease status of an individual. An individual is more often infected when its herd mates have a higher-than-average susceptibility, and less often when its herd mates have a lower-than-average susceptibility, on top of the effect of the susceptibility of the individual itself. This mechanism explains the larger than expected response to selection; when all individuals in a herd descend from superior parents, not only the individual itself will be less susceptible, but it will also be accompanied by herd mates that are less often infected. This reduces the exposure of the individual to the infectious agent. These results show that susceptibility not only has a direct effect on the disease status of the individual itself, but also an Indirect Genetic Effect (IGE) on the disease status of its herd mates (see also ANCHE *et al.* (2014)).

2.4 Discussion

Here we showed that low heritability estimates for infectious disease status (0/1=healthy/diseased) are fully consistent with a large amount of genetic variation in disease susceptibility. The genetic variation needed to reproduce an observed-scale heritability of only 0.02 roughly corresponds to R_0-values of 2.5 and 0.9 in the 10 % top and 10% bottom-ranking individuals. This large difference in R_0 corresponded to a large reduction in prevalence of the disease after selection, and even eradication of the disease occurred after a single generation of sire selection in our simulations.

The possibility to arrive at a prevalence of zero using selection, i.e., herd level eradication, is an important result. It contradicts predictions based on classical quantitative genetic models for binary traits that do not account for the transmission of an infection between individuals, such as the classical threshold model by DEMPSTER AND LERNER (1950). For binary traits, the classical model shows that the observed-scale heritability approaches zero when prevalence approaches zero or one (ROBERTSON 1950). Hence, continued selection against an infectious disease will reduce the heritability in the threshold model, so that response to

selection will approach zero as well, and it is impossible to reach a prevalence of zero.

The difference between our findings and predictions based on classical models originates from positive feedback effects occurring in the transmission of infectious diseases. These feedback effects entail that individuals with a low susceptibility are not only less likely to become infected themselves, but they also infect fewer herd mates, just because they are less likely to be in the infectious state. This shows that genetic variation in disease susceptibility leads to so-called Indirect Genetic Effects (IGE). In general, IGE are effects of the genotype of an individual on the trait values of other individuals (GRIFFING 1967; MOORE *et al.* 1997; MUIR 2005; BIJMA 2014). The fact that genetic variation in susceptibility inevitably leads to IGE was also shown by (ANCHE *et al.* 2014) and (BIJMA 2020). The larger difference in prevalence between the least and most susceptible individuals when allocation to herds was based on breeding value for susceptibility, compared to allocation at random, illustrates this IGE of susceptibility.

The response to selection we found can also be placed in the more general framework of the Price equation (PRICE 1970). The Price equation states that response to selection in a trait is the sum of two components. The first component represents the contribution directly attributable to selection. The second term represents the effect due to "incomplete fidelity of transmission of the trait value to the next generation" (GARDNER 2020). This transmission term defines non-selective effects, for example a difference in non-additive effects between the parent and offspring generation, due to a change in allele frequency. The response to selection in prevalence can be described in two ways using the Price equation, depending on whether the IGE of susceptibility is incorporated into the selection term or into the transmission term (BIJMA 2020). The common breeding values for disease status from the linear animal model do not capture the IGE of susceptibility and therefore represent the latter case, where the IGE is considered a non-selective effect arising from a change in the environment (less exposure to infectious individuals in the population). However, one can also define a so-called total breeding value for prevalence, including both the direct and indirect genetic effect of susceptibility (BIJMA 2011; BIJMA 2020). In this approach, the indirect

effect of susceptibility is shifted into the selection term. The latter approach makes sense, because the IGE due to genetic variation in susceptibility is a special kind, which arises entirely via the direct genetic effect on the disease status of the individual itself. Hence, the direct and indirect genetic effect due to susceptibility are fully correlated, such that selection on susceptibility is automatically on the indirect effect as well. This complete correlation also increases the total genetic variation in disease status (BIJMA 2010). Consequently, the breeding value for prevalence predicts a larger response in prevalence because of selection on susceptibility than the ordinary breeding values for disease status. Especially at lower prevalence, the selection differential in susceptibility required to eliminate a disease is much smaller than expected based on the classical breeding value for disease status (BIJMA 2020).

The positive feedback mechanism described above is well known in epidemiology, with herd immunity ($R_0 < 1$) as the most prominent example (FINE 1993). As illustrated by the eradication of rinderpest in cattle, it is not necessary that all individuals are fully resistant to infection to reach herd immunity. If a sufficient fraction of the population is vaccinated, a disease will have no possibility to transmit, because there are not enough sufficiently susceptible individuals in the population to sustain transmission. For herd immunity (*i.e.,* for R_0), it is in principle irrelevant whether a certain reduction in susceptibility is obtained by genetic selection or by vaccination causing incomplete immunity, because the positive feedback mechanisms underlying herd immunity result from the reduction in susceptibility, independent of how the reduction is obtained, and are therefore equally present in both cases. Thus, as with vaccination, a sufficient fraction of the population should have a sufficiently low susceptibility to reach herd immunity using genetic selection.

In our model, we simulated genetic variation in susceptibility only. Next to susceptibility, the transmission and prevalence of infectious diseases is affected by two other host traits: infectivity and the duration of the infectious period (DOESCHL-WILSON *et al.* 2011). Infectivity is the propensity of an infected individual to transmit the disease to a susceptible individual per unit of time. It has impact only on the disease status of other individuals, not on the disease status of the individual itself. In other words, infectivity only has an indirect genetic effect, no

direct genetic effect. Consequently, the common genetic analysis of binary disease status of the individual does not capture genetic variation in infectivity (LIPSCHUTZ-POWELL *et al.* 2012). Variation in infectivity thereby does not contribute to the established values of the heritability of disease status, which were the starting point of our analysis. Incorporating variation in infectivity in our simulations would therefore not change the genetic variation in susceptibility needed to arrive at the target observed heritabilities. However, if host infectivity shows genetic variation and it is incorporated in selection, this would increase the potential of the population to respond to selection for lower disease prevalence (e.g. TSAIRIDOU *et al.* 2019).

The second trait, the duration of the infectious period, relates to the ability of an individual to recover from infection, and determines the time an individual stays in the infectious state. In contrast to infectivity, variation in duration of the infectious period is directly reflected in an individual's disease status. In fact, it is the reverse of susceptibility, since it determines the rate at which individuals change from the infected to the susceptible state. In an earlier analysis with the same SIS-model, we found that the effect of variation in the duration of the infectious period on the heritability of disease status is comparable to that of susceptibility. Moreover, just like susceptibility, the duration of the infectious period clearly has an indirect effect as well. If individuals have a short infectious period, they are less likely to infect others, just because they are in the infectious state for a shorter period of time. Because the effects of the duration of the infectious period are similar to those of susceptibility, we therefore chose to simulate the genetic variation in susceptibility only.

In some scenarios, we needed to simulate a large variance in susceptibility to reproduce the desired heritability of disease status. A large variance in susceptibility had two counteracting effects on the prevalence of the disease in our simulations: (1) an increase in prevalence because $\bar{\gamma}$ is much larger than 1 leading to a higher R_0; (2) a decrease in prevalence because increasing heterogeneity among individuals reduces the prevalence (SPRINGBETT *et al.* 2003). With a small variance these two effects balanced each other, but with increasing variation the first effect became dominant, such that the actual prevalence in some of the simulations was considerably higher than the desired value of 0.33.

To prevent inconsistencies in the estimation of heritability and response to selection, we needed to correct for this higher prevalence, to obtain a prevalence of 0.33 in all scenarios. A prevalence much lower or higher than 0.33, close to 0 or 1, would have the effect that much higher genetic variances are needed to reach our target observed heritabilities. On first sight, the correction should ensure that the property $\beta\overline{\gamma_i}$ is equal to 0.03, either by decreasing β to $\frac{0.03}{\overline{\gamma_i}}$ or by setting $\overline{\gamma_i}$ to 1 (via the introduction of an extra term in Equation 2.7). However, such a correction of either β or $\overline{\gamma_i}$ resulted in a prevalence (much) lower than 0.33 because of the decreasing effects of heterogeneity on prevalence. We therefore chose to iteratively correct β until actual prevalence was 0.33 in all scenarios.

We observed a considerably lower response to selection for $\mathrm{a_p^2} = 0.25$ and $\mathrm{h_{obs}^2} = 0.10$ than for the other scenarios. The total (genetic + environmental) permanent susceptibility variance needed to reproduce $\mathrm{h_{obs}^2} = 0.10$ was exceptionally large for this scenario. It could even be considered unrealistically high, since it corresponds to a coefficient of variation of 111% which is above maximum values observed in literature (e.g HOULE 1992). Mean population susceptibility ($\overline{\gamma_i}$) was much larger than 1 for this variance, because of the skewness of the lognormal distribution of γ. We prevented effects of this higher mean susceptibility on the prevalence by the correction of β described in the previous paragraph. In case of selection, however, (extremely) high individual dam and environmental effects, resulting from the highly skewed distribution of susceptibility, cause a much higher average susceptibility in the offspring than would be expected from the mean breeding value of the selected sires. This higher susceptibility in the offspring resulted in a larger than expected $\beta\overline{\gamma_i}$, and consequently in a lower proportion of infection-free herds compared to the other scenarios. Because the combination of $a_p^2 = 0.25$ and $h_{obs}^2 = 0.10$ leads to unrealistically high variation in susceptibility, we feel this issue is not very relevant.

To illustrate the potential response to selection, we selected sires on their true (simulated) breeding values for susceptibility. We took this approach to reveal the additive genetic variance available for genetic improvement and to clarify the mechanisms underlying response to selection, without interference of the accuracy of breeding value estimation. However, the correlations between the estimated breeding values for disease status from the linear model and the true

breeding values for log-susceptibility of sires were between 0.75 (h^2_{Obs} of 0.02) and 0.90 (h^2_{Obs} of 0.10). This indicates that differences due to selecting sires based on their EBV for disease status instead of selecting them on their TBV for susceptibility would be relatively small. These high accuracies likely result from the large number of offspring per sire in our simulations and the intermediate prevalence of the disease, since individual differences in susceptibility are best visible at intermediate prevalence. If the prevalence is close to zero, the accuracy of EBVs will be smaller.

Even though our results clearly show that selection against infectious diseases should be much more promising than the common low heritability for disease status suggests, questions may arise as to whether our conclusions are achievable in practice and are not too optimistic. These questions might be motivated by the limited availability of empirical examples of large response to selection in disease prevalence. Even though the common quantitative models ignore indirect genetic effects, the previous paragraph shows that they still relatively accurately ranked the individuals on their susceptibility, which suggests that more examples of large response might be available than the two we found and mentioned in the introduction. In the next paragraphs we will identify important aspects related to the visibility and validity of our results in real data examples.

A first important point is that the feedback effects underlying our response are only effective if the entire herd is selected for low susceptibility. If only part of a herd is selected for low susceptibility, the offspring of selected parents are still exposed to (the offspring of) non selected individuals with higher susceptibility. In practice, this might for instance be the case in dairy cattle herds, where individuals usually belong to different generations. Another complicating factor is that herd-year effects in models for breeding value estimation may hide the feedback effects. When the model contains the direct genetic effect only, the IGE-component of genetic differences in the prevalence of the disease between herds, or between consecutive years of the same herd, will end up in the herd-year effect, and thus be attributed to differences in herd management. Hence, response due to positive feedback (*i.e.*, due to IGE) is difficult to see in classical quantitative genetic analysis.

A key assumption underlying our results is that the pathogen can replicate only in the host individual, meaning that a reduction in individual host susceptibility fully translates in reduced exposure of the host population to the pathogen. When a pathogen is able to replicate outside the host population, for instance in the environment or in another species, the impact of the feedback effects on transmission will be smaller. An example of such a case is Bovine Tuberculosis, where badgers are an external reservoir in which the pathogen can replicate (BOEHM *et al.* 2009). Because the key mechanism underlying our results is the positive feedback of selection for lower susceptibility, we expect that as long as no external replication occurs, our conclusions remain valid. For instance, in case where the transmission process is delayed because of a latency period between the moment of infection and the time an individual becomes infectious to others, or when the pathogen can survive but not replicate in the environment. The feedback effects in transmission are still present in those cases; only the total duration of the infection-cycle is prolonged (MA AND EARN 2006).

A general problem in relation to management of infectious diseases is the evolutionary response of pathogens to the applied interventions, such that these interventions become less effective. Pathogens are known to be able to adapt to every type of intervention, with the widespread antibiotic resistance as probably the best-known example (DAVIES AND DAVIES 2010; KENNEDY AND READ 2017). Here we discuss briefly what breeders could do to minimize the risk of pathogen adaptation. Importantly, evolution of pathogen resistance can only occur when there is transmission. Consequently, the most promising interventions with respect to prevention of resistance development are those that prevent transmission to occur ($R_0 < 1$) as soon as possible. The higher the selection pressure put on the pathogen population, especially when targeting the pathogen in multiple ways, the lower the probability of development of resistance (REX CONSORTIUM 2013). A well-known example is combination therapy for HIV, which is relatively successful in preventing pathogen resistance development by targeting the pathogen in multiple ways (KENNEDY AND READ 2017).

With respect to genetic selection, the above paragraph implies that strong selection for polygenic resistance should minimize the risk of pathogen adaptation. Plant breeders, for example, recognize that breeding for broad-spectrum

resistance is the most sustainable way to manage infectious diseases such as potato blight (VLEESHOUWERS *et al.* 2011). In livestock populations, results of Genome Wide Association Studies show that important endemic infectious diseases, such as mastitis and digital dermatitis are highly polygenic traits (TIEZZI *et al.* 2015; BIEMANS *et al.* 2019b). This would argue for a few generations of strong selection for lower susceptibility of the host population. However, such selection is currently uncommon in practice because of undesired correlated responses in other traits in the breeding goal, such as yield traits. For mastitis, for instance, the correlation of disease status with milk yield is positive, with most estimates between 0.20 and 0.55 (RUPP AND FOUCRAS 2010). Selection solely for a lower prevalence of mastitis would thus result in a negative (correlated) response in yield. Nevertheless, many questions about the adaption of pathogens to artificial selection in livestock are still unanswered.

The results of our simulations provide a new perspective on genetic selection for lower infectious disease prevalence and show that this might lead to much better results than current quantitative genetic models predict. Because of the complicating factors described in the above paragraphs, carefully designed selection experiments are needed to confirm our results with real data examples. Next to confirmation of our results with real data, work into the genetic background of the disease traits is needed as well, particularly on the presence of genetic variation in infectivity (LIPSCHUTZ-POWELL *et al.* 2012). The use of a method for statistical genetic analysis that accounts for the dynamic nature of transmission and for differences in infection exposure between individuals is essential here and might also increase the response to selection. Examples of such methods are models using Bayesian inference (POOLEY *et al.* 2020) and Generalized linear mixed models (BIEMANS *et al.* 2019b)).

In this paper we used simulations of an established epidemiological model to demonstrate that genetic selection against infectious diseases is much more promising than expected based on commonly used quantitative genetic models. As most models of biological systems, it will be needed to tailor our model to fit certain specific cases. However, by incorporating the positive feedback effects that have been demonstrated time and again in the field of epidemiology, our model

provides a fundamentally better description of infectious disease transmission than the classical quantitative genetic models for binary traits, which do not account for feedback effects. The main implication for breeders of this work is that a low ordinary additive genetic variance in binary disease status, as often observed in practice, should not be interpreted as a limiting factor for potential response to selection. Instead, we showed that these low values are fully consistent with a large amount of available genetic variation in disease susceptibility, which in turn translates into a larger than expected response to selection in the prevalence of the disease. Feedback effects occurring in transmission are of crucial importance for this response, and make it possible to eradicate an infectious disease, at least in theory.

2.5 Data Availability

Supplemental Material available at figshare: https://doi.org/10.25386/genetics.13090076. Table S1 contains heritability estimates for different genetic standard deviations in the logarithm of susceptibility. Figure S2 provides a schematic overview of our methodology. File S3 contains the R-code for simulation of a population with genetic variation in the logarithm of susceptibility. File S4 contains the R-code for simulation of the endemic disease.

2.6 Acknowledgements

We thank Pierre de Villemereuil for his detailed, very helpful comments which led to considerable improvements of the manuscript.

2.7 Appendix

2.7.1 Derivation of link variance for the complementary log-log link function

The link variance for the liability model as presented in Equation 2.11 in the main text, can be derived by calculation of the variance belonging to the pdf of the inverse of the cloglog link function. This pdf is obtained by differentiation of the inverse of the complementary log-log function.

The formula for the complementary log-log link function is:

$$x(p) = log(-log(1-p))$$

The inverse of this function is:

$$p(x) = 1 - e^{-e^x}$$

Differentiation of the inverse results in the formula for the pdf:

$$f(x) = e^{x-e^x}$$

The variance of this probability density function can be obtained by using the general formula for the variance of a function:

$$VAR(X) = E(X^2) - E(X)^2$$

in which *E* denotes the expected value. The expected values are calculated by integration of $f(x)$:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \gamma^2 + \frac{\pi^2}{6}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = -\gamma$$

in which γ denotes the Euler-Mascheroni constant. $VAR(X)$ becomes:

$$VAR(X) = \gamma^2 + \frac{\pi^2}{6} - (-\gamma)^2 = \frac{\pi^2}{6}$$

In fact, the inverse of the cloglog and its corresponding pdf are (mirrored) standard Gumbel distributions, the variance of the standard Gumbel distribution is indeed equal to $\frac{\pi^2}{6}$.

2.7.2 Gillespie-algorithm

In the simulation of the infectious disease, the time to the next event is sampled from an exponential distribution, which requires calculation of the total rate parameter r, defined as the sum of all individual transmission and recovery rates:

$$r = \sum_i \beta_i \frac{I}{N} + \alpha I$$

After sampling a time to the next event based on r, the type of event and the individual involved in the event is sampled based on the contribution of each individual to r, such that the probability of infection for a certain susceptible individual i is $\frac{\beta_i \frac{I}{N}}{r}$ and the probability of recovery for a certain infected individual j is $\frac{\alpha}{r}$ (this α is the same for each individual in our simulations). The effect of individual susceptibility on the probability of infection is thus multiplicative. An individual with a susceptibility of two has a two times higher β_i than an average individual with a susceptibility of one. Consequently, given same exposure to infectious individuals, its share in r, and thus its probability of infection, is two times higher as well.

Chapter 3

A quantitative genetic theory for infectious diseases

Piter Bijma[1], Andries D. Hulst[1,2], Mart C.M. De Jong[2]

[1] Animal Breeding and Genomics, Wageningen University & Research, 6700AH, Wageningen, The Netherlands

[2] Quantitative Veterinary Epidemiology, Wageningen University & Research, 6700AH, Wageningen, The Netherlands

Proceedings of 12th World Congress on Genetics Applied to Livestock Production: 163

Abstract

We integrated quantitative genetics and epidemiology to develop a quantitative genetic theory of the prevalence of endemic infectious diseases. Results show that infectious diseases respond very differently to selection than common non-communicable traits, and strongly suggest that the genetic variance determining the potential response of prevalence to selection must be much larger than currently believed. Moreover, heritable variation and response to selection increase significantly when the prevalence of the infection decreases, ultimately leading to local extinction of the infection due to herd immunity. These results change our perspective on the prospects of genetic selection against infectious diseases.

3.1 Introduction

A reduction of the prevalence of infectious diseases in livestock is highly desirable. Individual disease status, however, measured as $y = 0$ or 1 indicating non-infected vs infected, typically shows low heritability (h^2). Moreover, the classical threshold model predicts that h^2 goes to zero when a disease becomes rare. Hence, genetic selection against infectious diseases seems difficult. This perspective, however, ignores that infectious diseases rely on transmission between individuals. Here we present a quantitative genetic theory for response to selection and heritable variation in the prevalence of endemic infectious diseases.

3.2 Methods & results

The endemic prevalence (P) of an infectious disease is the average fraction of the population that is infected, and is equal to the mean of individual disease status (P = avg(y)). Standard epidemiological theory shows that the endemic prevalence follows from the basic reproduction number (R0; WEISS AND DISHON (1971); Figure 3.1),

$$P = 1 - \frac{1}{R_0} \tag{3.1}$$

R_0 is the number of individuals that gets infected by a single typical infected individual in a naive population. For example, with $R_0 = 3$, an infected individual would on average infect three others. However, if 2/3 of the population is infected already, then only a single new infection will occur on average. Hence, prevalence reaches an equilibrium at P = 1 – 1/3 = 2/3. To reduce prevalence, therefore, breeders should lower R_0. At low R_0, prevalence is very sensitive to changes in R_0 (Figure 3.1), which has significant implications for response to selection.

3.2.1 Breeding values for R_0, prevalence and disease status

R_0 is the product of the contact rate between individuals, the *sus*ceptibility of individuals to become infected, the propensity of individuals to infect others (*inf*ectivity) and the duration of the infectious period (t). Because R_0 is a product and because it takes strictly positive values, we define an additive model with normally distributed genetic effects for the logarithm of R_0,

$$A_{\ln(R_0)} = A_{\ln(sus)} + A_{\ln(inf)} + A_{\ln(t)} \tag{3.2}$$

We assume that additive genetic (co)variances are constant on this log scale, i.e., are independent of the level. (This implies that additive genetic variance in R_0 decreases as R_0 decreases). Connecting Equation 3.1 and 3.2 leads to a breeding value for individual disease status (A_y) and a breeding value for prevalence (A_P; BIJMA *et al.* (2022); here assuming no genetic variation in infectivity),

$$A_y = P(1-P)A_{\ln(R_0)} \tag{3.3a}$$

$$A_P = (1-P)A_{\ln(R_0)} \tag{3.3b}$$

A_y is the ordinary ('direct') breeding value for the disease status of the individual itself, while A_P is a total breeding value that measures the total effect of an individual's genes on the prevalence in the population. Hence, response of prevalence to selection equals the per generation change in mean A_P. Note that A_P is a factor 1/P greater than A_y. Moreover, for a constant $A_{\ln(R_0)}$, A_y has a maximum for P=0.5, while A_P increases when P decreases. Thus, the common h^2 of individual disease status (y=0,1) has a maximum at P=0.5, similar to the threshold model (ROBERTSON 1950), but the additive genetic variance that breeders can use for response to selection increases significantly as prevalence falls (Figure 3.2). This result is a consequence of the increasing slope of the relationship between P and R_0 at lower values of R_0 (Figure 3.1).

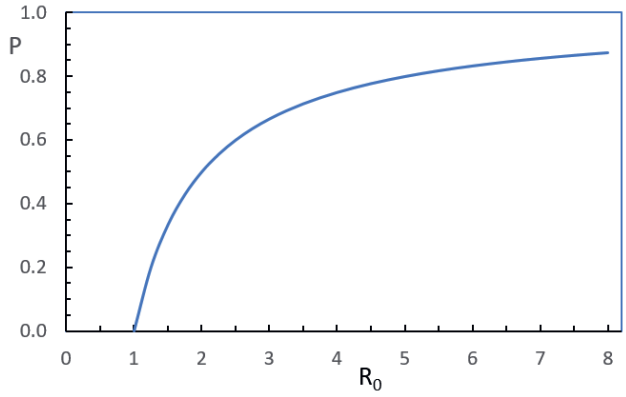


Figure 3.1 Prevalence as a function of R_0.

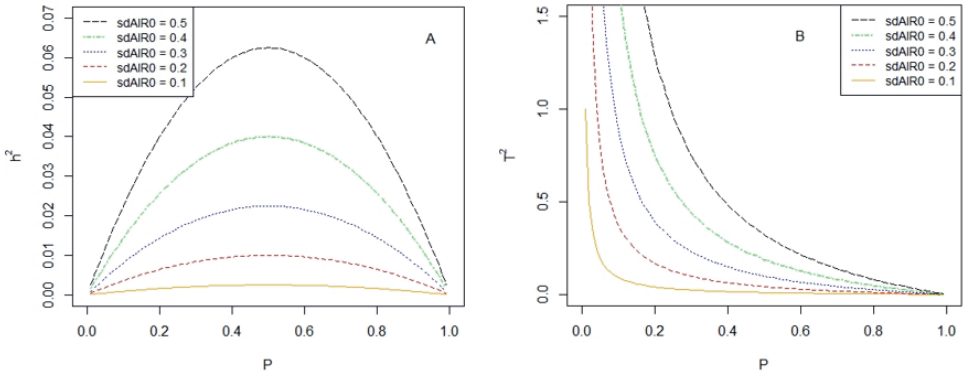


Figure 3.2 Heritability of individual disease status (h^2, panel A), and additive genetic variance in the endemic prevalence as a fraction of phenotypic variance (T^2, panel B), as a function of prevalence (P). For different additive genetic standard deviations in ln(R_0), and without genetic variation in infectivity.

Figure 3.2 shows that common heritabilities of disease status correspond to a large additive genetic variance in prevalence, particularly when prevalence is small. For example, for P=0.2 and sd($A_{\ln(R_0)}$)=0.4, the common observed-scale heritability of binary disease status is only 0.026, while the additive genetic variance that can be used for response to selection is 64% of phenotypic variance (T^2=0.64). Hence, for this scenario, the additive genetic standard deviation in prevalence is $\sqrt{0.64 \times 0.2 \times (1-0.2)} = 0.32$, which is very large.

3.2.2 Response to selection

Figure 3.2B shows a strong increase in additive genetic variance as prevalence falls, suggesting that response to selection for lower prevalence should accelerate over generations. Figure 3.3 shows response to selection observed in simulations following standard methods in epidemiology, not using any of the above theory. Figure 3.3 also shows the mean breeding values for prevalence (A_P) and for individual disease status (A_Y), expressed as deviations from prevalence in the first generation. As expected, based on Figure 3.2, the response accelerates as prevalence decreases. In the final generations (t=16, and t=70) the infection disappears due to herd immunity in the local population, not because individuals are fully resistant to infection. The mean breeding value for prevalence (Equation 3.3b) closely matches the observed response, while the mean breeding value for individual disease (Equation 3.3a) deviates more and more from the observations as prevalence decreases. Hence, response in Figure 3.3 is significantly greater than the common breeding values for individual disease status suggest.

3.3 Discussion

We developed a quantitative genetic theory of infectious diseases, focussing on R_0 and on the endemic prevalence of the infection. The breeding value for the logarithm of R_0 is central to the theory and is the sum of the breeding values for the logarithms of susceptibility, infectivity and duration of the infectious period. From the breeding value for $\ln(R_0)$ we derived a breeding value for individual binary disease status and a breeding value for the prevalence of the infection in the population. The first relates to ordinary h^2 of individual disease status, while the second relates to response to selection. Without genetic variation in infectivity, the breeding value for prevalence is a factor $1/P$ greater than the breeding value for individual disease status. This is a general result and does not rely on the assumption of an additive model on the log scale. For this reason, both the additive genetic variance for prevalence and the response of prevalence to selection increase when prevalence goes down (Figure 3.2B, 3.3).

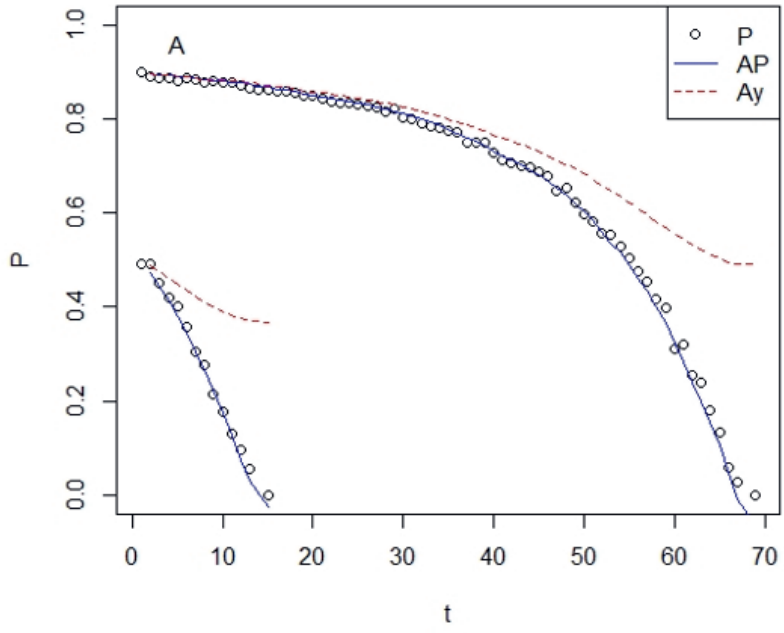


Figure 3.3 Response to mass selection against an infectious disease; observed response (small circles), mean breeding value for prevalence (A_P, solid blue line) and mean breeding value for individual diseases status (A_y, dashed red line). For two populations, one starting at a prevalence of 50%, and the other at 90%. With genetic variation in susceptibility only, with var($A_{\ln(R_0)}$) = var($A_{\ln(sus)}$) = 0.32, and a selected proportion of 0.5. The common observed-scale heritability for individual disease status in any generation can be read from Figure 3.2A, using an x-axis value corresponding to the prevalence in that generation, and does not exceed 0.022.

The large difference between the h^2 of individual disease status and the total additive genetic variance in prevalence as a fraction of phenotypic variance (T^2) shown in Figure 3.2 indicates the presence of considerable indirect genetic effects (IGE). The h^2 reflects the variance of only the direct genetic effect of individuals on their own disease status. The T^2 reflects the variance of the full effect, i.e., of the sum of the effects on the disease status of the individual itself and on the disease status of its herd mates. Diseases that rely on transmission generate considerable IGE because an individual who does not become infected itself also cannot infect other individuals (simply because it is not infected). Hence, the direct effects for susceptibility and for duration of the infectious period are fully

correlated to the corresponding indirect effects, resulting in positive feedback and an increase in the total additive genetic variance. This positive feedback effect increases strongly when prevalence decreases, which causes IGE to make up an increasing proportion of the total effect, and explains the difference between h^2 and T^2 in Figure 3.2A vs B.

The results shown here apply to endemic microparasitic infections, and where the pathogen relies on the host individual (i.e., the livestock) for its replication. Hence, we assume absence of an external reservoir of pathogens, such as a second host species (for example badgers in the case of bovine Tb). Note that temporary survival of the pathogen in the environment does not violate this assumption. This assumption implies that a reduction in, e.g., susceptibility translates fully into reduced exposure in the local population (because of fewer infected individuals). Example traits include mastitis, infectious claw disorders, respiratory infections in young animals (young replacement stock, meat calves, fattening pigs), faecal-oral transmitted infections causing gastro-intestinal diseases, and several zoonosis.

In conclusion, our results show that the quantitative genetics of infectious diseases is very different from the quantitative genetic theory for ordinary, non-communicable, traits, and that prospects for genetic improvement are much better than currently believed, as was also proposed by BISHOP AND STEAR (1997). These findings, coupled with the high importance of infectious diseases for animal health and welfare, and also for human health in the case of zoonoses, suggest that breeders should become familiar with basic concepts in epidemiology, such as R_0. Together with recent advances in sensor technology and AI, which will enable us to collect large-scale high-quality data on individual disease status, these findings facilitate artificial selection against infectious diseases.

Chapter 4

Buy one get some free: the 1/P-fold benefit of interventions against transmission of infections

Andries D. Hulst[1,2], Piter Bijma[2], Mart C.M. De Jong[1]

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

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

Summary

Vaccines against transmission of infections are usually evaluated for their direct efficacy in reducing the risk of infection of those being vaccinated. (ORENSTEIN *et al.* 1985; HALLORAN *et al.* 1991; WEINBERG AND SZILAGYI 2010). Although vaccines can also have considerable effects on the risk that others in the population become infected (DROLET *et al.* 2015; HOES *et al.* 2021), these indirect effects are typically difficult to quantify beforehand or require detailed modelling (e.g. HALLORAN *et al.* 1999; LONGINI AND HALLORAN 2005; KILLEEN *et al.* 2007; TSANG *et al.* 2019; SALO *et al.* 2022). Here we show that the overall efficacy of vaccination against infection is typically at least the inverse of the endemic prevalence of the infection times the direct efficacy. Thus, for an infection that, for example, is present in one out of four individuals on average, the overall efficacy is at least four times the direct efficacy, and indirect effects contribute at least three-quarters of the total. Similarly, in this example, vaccination of a single individual effectively protects a total of four individuals with the direct efficacy of the vaccine. We illustrate this finding with detailed individual-based simulations that accurately mimic the real-life spread of SARS-COV2 in Bogota, Colombia. Our simple rule of thumb demonstrates that indirect effects increase strongly at lower prevalence, often contributing the majority of the total efficacy, and quantifies their minimum magnitude. This informs both public health managers and the general public about the benefits of vaccination against infections beyond the persons vaccinated.

4.1 Main

Vaccines against transmission of infections decrease the infection risk in two ways: (1) through protective effects on the vaccinated individual itself; the direct efficacy, and (2) through reduced exposure of other individuals in the population, as a result of the reduced number of infected individuals due to vaccination; the indirect efficacy (e.g. HALLORAN *et al.* 1999). Together, these lead to a decreased infection risk for all individuals in the population, the overall efficacy.

In vaccine studies, interest is typically in the direct efficacy of the vaccine, measured as one minus the relative infection risk, or attack rate (AR), of vaccinated over unvaccinated individuals, with assumed equal exposure in both groups (GREENWOOD AND YULE 1915; ORENSTEIN *et al.* 1985; ORENSTEIN *et al.* 1988; HALLORAN *et al.* 1992; CLEMENS *et al.* 1996; HALLORAN *et al.* 1999; WEINBERG AND SZILAGYI 2010),

$$Direct\ Efficacy = 1 - \frac{AR_V}{AR_U} \tag{4.1}$$

The attack rate is the proportion of the group that becomes infected over a certain period of time (PORTA *et al.* 2014). The direct efficacy, therefore, measures the relative reduction in infection risk of vaccinated compared to non-vaccinated subjects under equal exposure. While the indirect efficacy is known to be relevant for herd immunity (e.g. FINE 1993), it remains unknown from vaccine efficacy studies.

Here, we use an epidemiological model with Susceptible, Infectious, Resistant, and Susceptible compartments (SIRS) and two types of hosts, unvaccinated and vaccinated, and obtain exact expressions for the direct and overall efficacy. SIRS models represent a very broad class of models in infectious disease epidemiology, as they encompass SIR, SIS and SI models and for our purpose even models including latently infected individuals. The result shows that the overall efficacy of vaccination is typically at least the inverse of the endemic prevalence before vaccination (P_0) times the direct efficacy, until the point where the infection is eradicated (Figure 4.1A). Note that we use vaccination as example, while the results apply to any intervention that protects individuals from becoming infected. The overall efficacy incorporates both the direct effect on the vaccinated

individual(s) and the indirect effect of reduced exposure in the entire population, and equals one minus the ratio of the pre-vaccination attack rate and the attack rate in the post-vaccination endemic equilibrium,

$$Overall\ Efficacy\ = 1 - \frac{AR_{post-vacc}}{AR_{pre-vacc}} \tag{4.2}$$

For example, the overall efficacy of vaccinating an entire population with a pre-vaccination endemic prevalence of one in four, is at least four times the direct efficacy of the vaccine. In other words, suppose a vaccination trial maintaining equal exposure reduces the attack rate from 25% in unvaccinated individuals to 20% in vaccinated individuals, yielding a direct efficacy of 20%. Applying this vaccine to an entire population with a pre-vaccination endemic prevalence of one in four will reduce prevalence by at least four times the direct effect, from 25% pre vaccination to no more than 5% post vaccination (25% minus 4 times 5%, corresponding to an overall efficacy of 4 times 20% is 80%).

For population-wide vaccination, the factor $1/P_0$ is a lower bound obtained for vaccines of very small effect. For larger effects, the overall efficacy exceeds the $1/P_0$ fold of the direct efficacy (Figure 4.1A). Population-wide vaccination with a direct efficacy of 0.4 in a population with an endemic prevalence of a half, for example, yields an overall efficacy of three times the direct efficacy; a value 50% greater than the lower bound $1/P_0$ of two (Figure 4.1A, DE = 0.4, P_0 = 0.5). Thus, the factor $1/P_0$ represents a lower bound for of the relative size of the overall efficacy, and the actual overall efficacy can be larger.

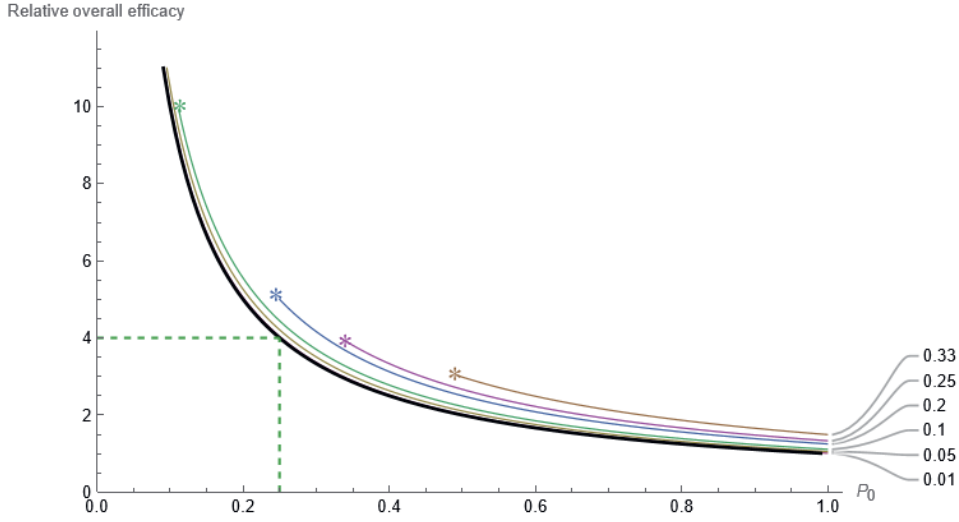
Our result also applies to vaccination of single individuals. Then, vaccination of a single individual in a large unvaccinated population protects $1/P_0$ individuals in total, each with an efficacy equal to the direct efficacy. With an endemic prevalence of one in four, for example, this corresponds to four individuals; three other unvaccinated individuals on top of the vaccinated one (Figure 4.1B). For this case, the ratio of overall over direct efficacy perfectly matches the lower bound of $1/P_0$ irrespective of the effect of the vaccine.

The lower bound holds irrespective of the proportion of the population that receives the vaccine, of the proportion that had already been vaccinated beforehand, and of whether the vaccine provides complete or incomplete

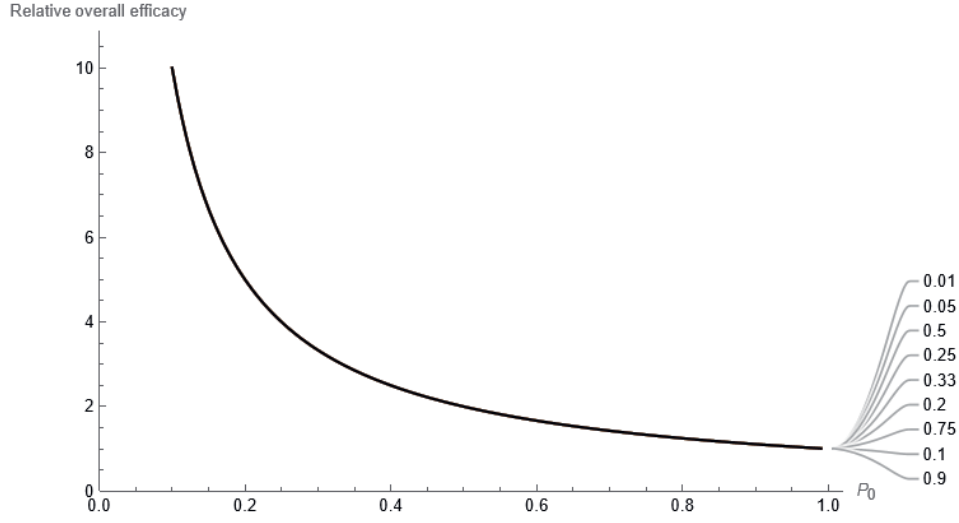
protection (Figure 4.1A-C; Appendix 4.5.1). Moreover, the result applies to both endemic and epidemic infections. For epidemic infections, the unobservable endemic prevalence can be expressed in terms of an estimated value of the basic reproduction number (R_0), so that the lower bound becomes $1/P_0 = R_0/(R_0 - 1)$. Thus, the product of the ordinary direct efficacy and the inverse of the endemic prevalence before vaccination, or its epidemic equivalent in terms of R_0, provides a general lower bound for the overall efficacy of vaccines against infection transmission.

In our derivations, we used homogeneous mixing of the population. To test the generality of our findings, we did an extra analysis on an existing extensive individual-based simulation of SARS-COV2 in 7.14 million people in Bogota, Colombia, including detailed spatial-temporal variation in transmission (ESPAÑA *et al.* 2022) (see Appendix 4.5.3 for details). Given the estimated R_0 of 2.85 before vaccination, we predict a ratio of overall to direct efficacy of $R_0/(R_0 - 1)$ = 1.54. With a fraction vaccinated equal to 0.486, and an average direct efficacy of 0.502 for those vaccinated, and an outbreak size in the simulation of 0.296 without vaccination, this corresponds to a predicted outbreak size of 0.185 with the vaccination. Note that without the indirect effect of vaccination the predicted outbreak size would have been 0.224. The observed outbreak size in the simulation with vaccination exactly matched our prediction. Thus, our rule of thumb gave an accurate prediction of the overall efficacy also for this heterogeneous population.

A


Relative overall efficacy
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0
P_0
0.33
0.25
0.2
0.1
0.05
0.01

B


Relative overall efficacy
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0
P_0
0.01
0.05
0.5
0.25
0.33
0.2
0.75
0.1
0.9

C

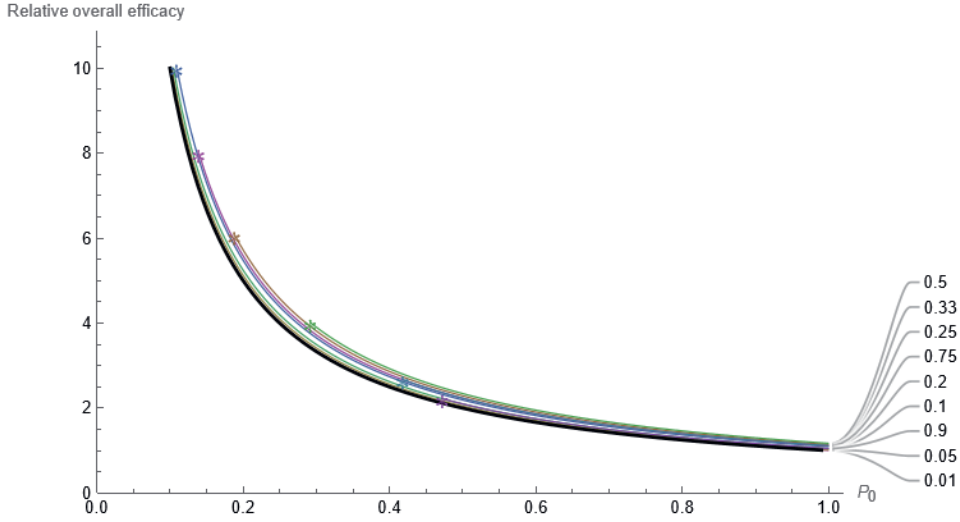


Figure 4.1 Ratio of overall to direct efficacy as a function of endemic prevalence before vaccination (P_0). Vaccination of the entire population (Panel A), of a single individual (Panel B), and of 50% of the population (Panel C). A ratio of 1 indicates equal direct and overall efficacy, a ratio of 2 indicates that the overall efficacy is twice the direct efficacy, *etc*. The vaccine has a direct efficacy of 1% to 90% (except for Panel A, where the maximum efficacy is 33%, because instant eradication occurs otherwise). The solid black line shows $1/P_0$. In Panel B, all 4 lines lie on top of each other. In Panels A and C, the lines stop when eradication is reached, as indicated by the asterisks. The green dashed lines in Panel A show the example for a prevalence of one in four, mentioned in the text.

4.2 Discussion

We showed that the overall efficacy of vaccination against transmission of infections is typically at least the inverse of the endemic prevalence times the ordinary direct efficacy. Thus, reliable estimates for endemic prevalence or R_0 are sufficient to predict overall efficacy from direct efficacy. Estimation of direct efficacy is well established (e.g. ORENSTEIN *et al.* 1985; HALLORAN *et al.* 1991; WEINBERG AND SZILAGYI 2010), and an estimate for the endemic prevalence, or its epidemic equivalent, follows readily from estimates for R_0, as $P_0 = 1 - 1/R_0$ (WEISS AND DISHON 1971). If other interventions, such as detection and isolation, have

already been implemented and have caused a decrease in R_0, then the R_0-value after the implementation of these interventions must be used to predict the overall efficacy of additional vaccination. Thus, the overall efficacy of vaccination will be relatively larger when vaccination is implemented after other interventions, which are not lifted after vaccination, because of a greater value of $1/P_0$.

We offer a very simple rule of thumb for the relative magnitude of the overall efficacy of vaccination. This primarily increases our understanding of the indirect effect of vaccination, which until now has required us to resort to detailed epidemiological models, though a model of influenza (EICHNER *et al.* 2017) and a quantitative genetic model of the host population (BIJMA *et al.* 2022) already hinted at our main finding. Moreover, our simple prediction of the overall effect is of great value for communication about, and acceptance of, vaccination programs.

Nevertheless, it is important to be aware that our prediction represents a lower bound for the overall efficacy of vaccination applied to randomly chosen individuals, and more precise estimates might be desired in specific cases, for instance for cost-effectiveness analyses of vaccines or when determining minimum vaccination coverage needed for eradication. For example, if mainly people with many contacts are vaccinated, then the indirect efficacy will be larger, and our rule-of-thumb will be an underestimation.

We have considered vaccines that reduce the risk of infection in the vaccinated individuals (direct effect). This direct effect creates an accompanying indirect effect, because the reduction in the number of infected individuals in the population translates directly into reduced exposure for other individuals (vaccinated or not vaccinated). An exception are cases where there is transmission of the pathogen within an external pathogen reservoir, such as a secondary host species or a vector. Such cases are outside the scope of this work. Moreover, an additional indirect effect may occur when a vaccine also reduces the propensity of infected individuals to infect others, for example by reducing pathogen shedding or the duration of the infectious period. Such effects will increase the ratio of overall over direct efficacy above the factor $1/P_0$. The factor $1/P_0$ presented here represents the overall efficacy directly attributable to the direct efficacy of vaccination, which can thus be predicted from vaccination trials and an estimate of R_0.

4.3 Methods

This section starts with a description of the epidemiological model (SIRS) that was used as the foundation for the derivations in this paper, then we introduce vaccination into the model and obtain exact expressions of the attack rates for unvaccinated and vaccinated individuals and of the overall attack rate. Finally, we use these attack rates to derive expressions for the direct and overall efficacy of a vaccine, for any fraction of the population vaccinated and any degree of protection.

4.3.1 General SIRS-model without vaccination

Our calculations of the direct, overall, and indirect efficacy of interventions are based on a Susceptible-Infectious-Recovered-Susceptible (SIRS) model. In the SIRS-model, individuals can be in three states: susceptible to the infection, infectious, and recovered (which implies in the model fully immune). Recovered individuals might lose their immunity and thereby become susceptible again. In principle, this loss of immunity can also correspond to the death of the individual and the birth of a new susceptible. Furthermore, the SIRS-model covers both SIS (if loss of immunity occurs immediately after recovery) and SIR-models (if immunity lasts forever), and thereby most of the elementary, well-established, epidemiological models (e.g. HETHCOTE 1989). With this flexibility, the SIRS-model fits our purpose of finding a general, but well-founded quantification for the overall efficacy of interventions. It can be adapted to many endemic and epidemic human and livestock infectious diseases.

In the SIRS-model, the rate at which susceptible individuals become infected per unit of time is equal to $\beta\, I/N$ (Equation 4.3a and 4.3b), where β is the transmission rate parameter, and *S, I,* and *N* are the numbers of susceptible individuals, infected individuals and the total number of individuals respectively. The rate at which infectious individuals per unit of time recover is equal to αI (Equation 4.3b and 4.3c), where α is the recovery rate parameter. The rate at which immune individuals lose their immunity per unit of time and thereby become susceptible again is equal to γR (Equations 4.3a and 4.3c), where γ is the rate parameter that determines the rate at which immunity is lost. Thus, the deterministic version of the SIRS model can be described by three differential equations:

$$\frac{dS}{dt} = \gamma R - \beta S \frac{I}{N} \quad (4.3a)$$

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \alpha I \quad (4.3b)$$

$$\frac{dR}{dt} = \alpha I - \gamma R \quad (4.3c)$$

The rate parameters (β, α, γ) determine the transmission characteristics of the infection. The basic reproduction ratio R_0, the number of secondary infections caused by a single infected individual in a further fully susceptible population is equal to β/α (DIEKMANN *et al.* 1990). R_0 is linked to a threshold phenomenon, if R_0 is above one an infection can spread, if R_0 is below one an infection will, with certainty, die out. When R_0 is above 1, the SIRS-model tends to a dynamic equilibrium, in which the average proportion of individuals in each state does not change. The proportion of individuals in each state in the equilibrium is determined by the rate parameters and can be found by solving the differential equations for 0. The proportion susceptible individuals is $\frac{\alpha}{\beta} (= 1/R_0)$, the proportion infected individuals is $\left(1 - \frac{1}{R_0}\right) * \frac{\gamma}{\gamma+\alpha}$, and the proportion recovered individuals is $\left(1 - \frac{1}{R_0}\right) * \frac{\alpha}{\gamma+\alpha}$. Thus, the total fraction of individuals affected by the infection (I+R) is equal to $1 - \frac{1}{R_0}$, which represents the endemic prevalence before vaccination (P_0).

4.3.2 SIRS-model with vaccination

The application of a vaccine that reduces the chance of an individual to become infected will decrease the value of β for the vaccinated individuals, and thereby the value of the reproduction ratio, reflecting that individuals that received the vaccine are less likely to become infected than unvaccinated individuals. Then, the description of the model becomes:

$$\frac{dS_U}{dt} = \gamma R_U - \beta_U S_U \frac{I}{N} \quad (4.4a)$$

$$\frac{dS_V}{dt} = \gamma R_V - \beta_V S_V \frac{I}{N} \quad (4.4b)$$

$$\frac{dI_U}{dt} = \beta_U S_U \frac{I}{N} - \alpha I_U \quad (4.4c)$$

$$\frac{dI_V}{dt} = \beta_V S_V \frac{I}{N} - \alpha I_V \quad (4.4d)$$

$$\frac{dR_U}{dt} = \alpha I_U - \gamma R_U \quad (4.4e)$$

$$\frac{dR_V}{dt} = \alpha I_V - \gamma R_V \quad (4.4f)$$

Where the subscripts *U* and *V*, denote unvaccinated and vaccinated individuals. Note that the only parameter affected by vaccination is the transmission rate parameter, and that its value only depends on the type of the susceptible individual (U or V) in an infectious contact, not on the type of the infectious individual. Thus, here we assume that vaccination only affects the probability of individuals to become infected (=susceptibility), not their propensity to infect others once infected (=infectivity). Since the main aim of vaccination is to reduce the number of individuals that get infected at some point in time, we will sum the compartments I and R for each type in the following and refer to the equilibrium fractions $\frac{I_U+R_U}{N}$ and $\frac{I_V+R_V}{N}$ as type-specific (unvaccinated resp. vaccinated) attack rates and to the fraction $\frac{I+R}{N}$ as overall attack rate.

Similarly as for the model without vaccination, we can find the equilibrium fractions in each compartment by solving the differential equations for 0, but the resulting equations are way more complex and less informative than for the model without vaccination. The equations for the overall and type-specific attack rates are given in Appendix 4.5.2. Note that the overall attack rate post vaccination depends on the type-specific attack rates (AR_U and AR_V) and the vaccinated fraction f_V:

$$AR_{post-vacc} = f_V AR_V + (1 - f_V) AR_U \quad (4.5)$$

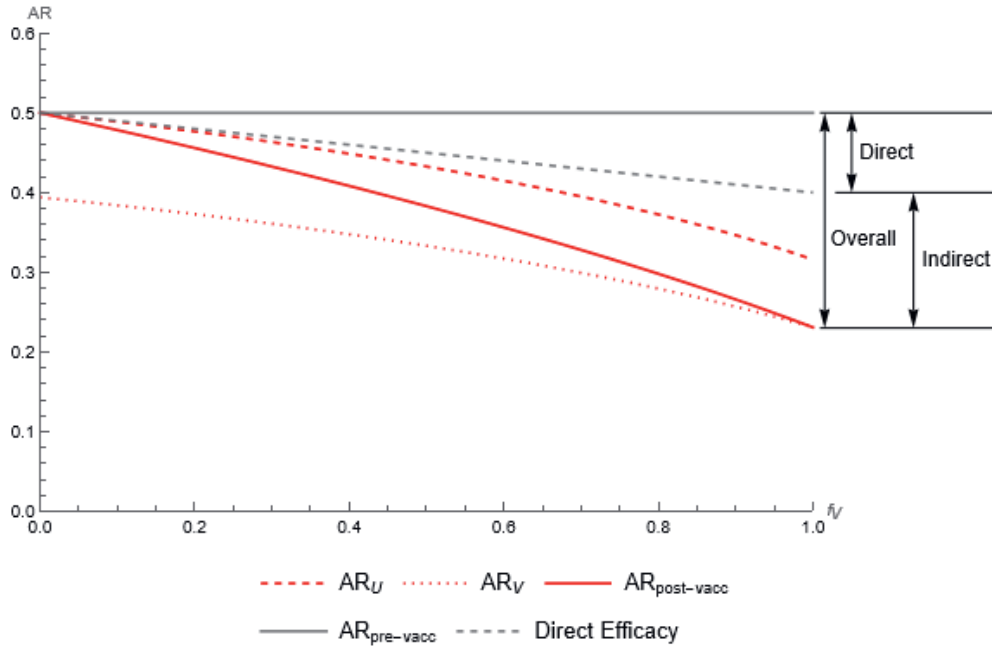


Figure 4.2 Type specific attack rate for unvaccinated (AR_U) and vaccinated (AR_V) individuals, and mean attack rate post- and pre vaccination, as a function of the fraction of the population that received the vaccine (f_V). Overall, direct, and indirect efficacy are indicated in the figure. R_0 is 2, $R_{VACC} = 1.3$, such that the direct efficacy of the vaccine is 0.2 ($1 - \frac{0.4}{0.5}$). The overall efficacy when the full population is vaccinated is $1 - \frac{0.23}{0.5} = 0.54$. The predicted ratio overall *vs.* direct efficacy is 2 in this example, because endemic prevalence is 0.5. The actual ratio overall/direct is larger, 2.7.

4.3.3 Direct and overall efficacy

Now we found exact equations for the attack rates, we can use these to calculate the direct and overall efficacy of vaccination. Consistent with HALLORAN *et al.* (1999), we define the overall efficacy as the total relative reduction in infection risk post vaccination, and the direct efficacy as the relative reduction in infection risk of vaccinated to unvaccinated subjects under equal exposure (Figure 4.2 and Equations 4.1 and 4.2 in the main text). Note that for correct comparison of direct and overall efficacy, the direct efficacy should be multiplied with the fraction of the population that received the vaccine (as is visible in Figure 4.2).

4.4 Acknowledgements

We thank Quirine ten Bosch for her helpful comments on the manuscript and Guido España for providing the results of the SARS-COV2 simulation and his help with the interpretation of the results.

4.5 Appendix

4.5.1 Ratio overall to direct efficacy in different situations

The figures in this section illustrate the ratio overall to direct efficacy for a variety of situations.

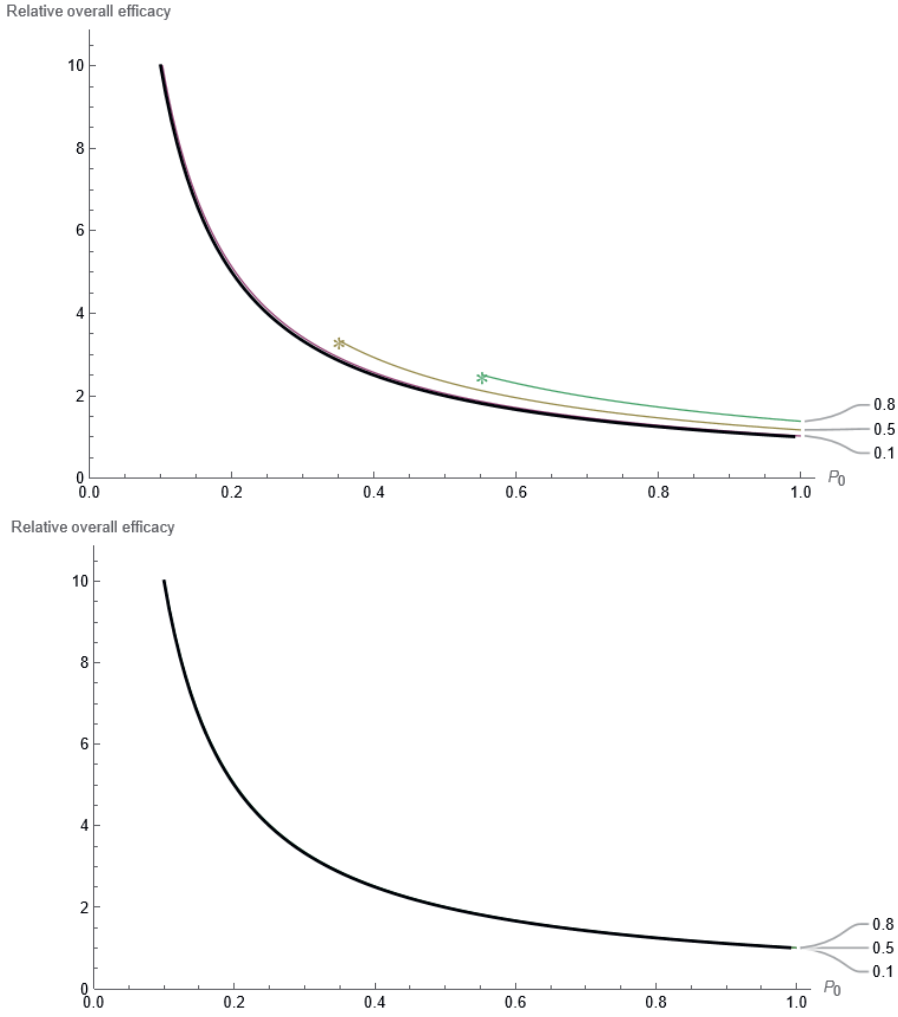


Figure 4.3 Vaccination of different fractions in an unvaccinated population. Direct efficacy of 0.01 (Upper panel) and 0.5 (lower panel). The different lines correspond to different vaccinated fractions (0.1, 0.5, 0.8). The solid black line shows $1/P_0$. Eradication of infection is indicated by asterisks.

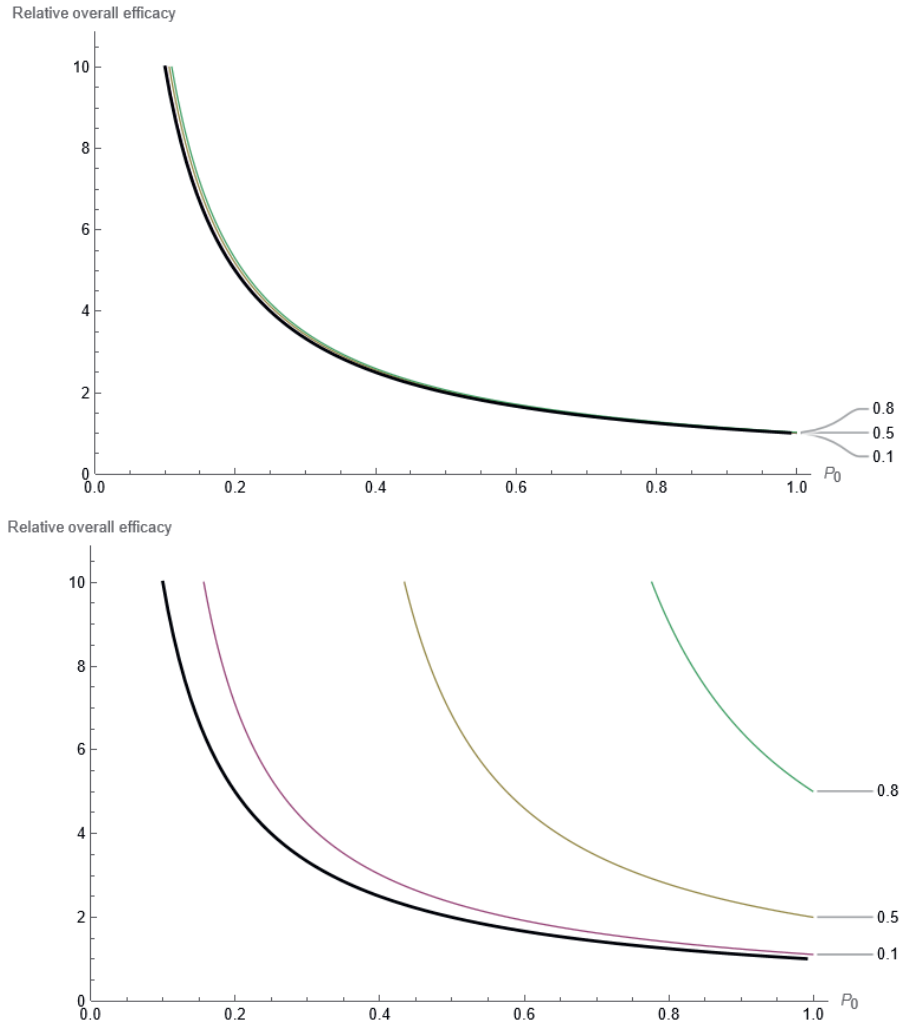


Figure 4.4 Vaccination of a single individual in a population of which different fractions are already vaccinated. Direct efficacy of 0.01 (Upper panel) and 0.5 (lower panel). The different lines correspond to different vaccinated start fractions (0.1, 0.5, 0.8). The solid black line shows $1/P_0$.

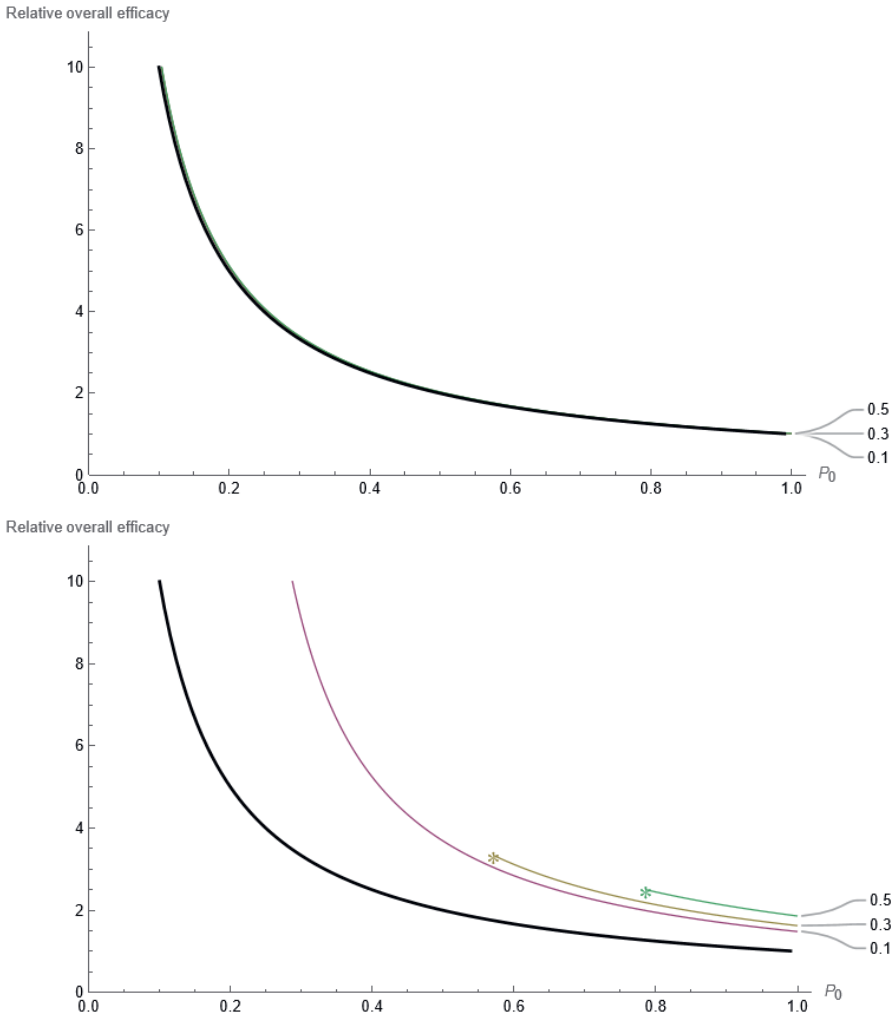


Figure 4.5 Vaccination of different fractions in a population of which 30% is already vaccinated. Direct efficacy of 0.01 (Upper panel) and 0.5 (lower panel). The different lines correspond to different vaccinated fractions (0.1, 0.5, 0.8). The solid black line shows $1/P_0$. Eradication of infection is indicated by asterisks.

4.5.2 Equilibrium equations for overall and type-specific attack rates

f_V denotes the fraction of the population that received the vaccine, AR the overall attack rate ($\frac{I_U+R_U+I_V+R_V}{N}$), AR_V the attack rate in vaccinated individuals ($\frac{I_V+R_V}{N}$), and AR_U the attack rate in unvaccinated individuals ($\frac{I_U+R_U}{N}$). R_V denotes the basic reproduction ratio in vaccinated individuals and is equal to $\frac{\beta_V}{\alpha}$, R_U the basic reproduction ratio in unvaccinated individuals and is equal to $\frac{\beta_U}{\alpha}$.

$$AR = \frac{1}{2}\left(-\frac{1}{R_U} + \frac{-1 + R_V + R_V\sqrt{\frac{1}{R_U^2} + \frac{-2 + 4f_V - \frac{2}{R_V}}{R_U} + \frac{1 - 2R_V - 4f_V R_V + R_V^2}{R_V^2}}}{R_V}\right) \quad (4.6a)$$

$$AR_U = \frac{(1 - f_V)\left(-R_V + R_U(-1 + R_V + R_V\sqrt{\frac{1}{R_U^2} + \frac{-2 + 4f_V - \frac{2}{R_V}}{R_U} + \frac{1 - 2R_V - 4f_V R_V + R_V^2}{R_V^2}})\right)}{R_V + R_U(-1 + R_V + R_V\sqrt{\frac{1}{R_U^2} + \frac{-2 + 4f_V - \frac{2}{R_V}}{R_U} + \frac{1 - 2R_V - 4f_V R_V + R_V^2}{R_V^2}})} \quad (4.6b)$$

$$AR_V = \frac{f_V\left(-R_V + R_U(-1 + R_V + R_V\sqrt{\frac{1}{R_U^2} + \frac{-2 + 4f_V - \frac{2}{R_V}}{R_U} + \frac{1 - 2R_V - 4f_V R_V + R_V^2}{R_V^2}})\right)}{-R_V + R_U\left(1 + R_V + R_V\sqrt{\frac{1}{R_U^2} + \frac{-2 + 4f_V - \frac{2}{R_V}}{R_U} + \frac{1 - 2R_V - 4f_V R_V + R_V^2}{R_V^2}}\right)} \quad (4.6c)$$

4.5.3 Simulation model of SARS-COV2 in Bogota

The goal of this validation is to check if the rule of thumb ($\frac{1}{P_0}$ / $\frac{\mathrm{R0}}{\mathrm{R0}-1}$ is lower bound for the ratio overall to direct efficacy) also holds in a complex model with numerous parameters, and spatial and age-dependent transmission.

Two scenarios were simulated in the model of COVID19 in Bogota by ESPAÑA *et al.* (2022). One follows the actual vaccination strategy, the other one assumes no vaccination at all. There is a clear difference in number of cases between the two scenarios in the period between day 625 and day 725 (Fig S3.1), therefore we choose this period for our analysis. Furthermore, the steep increase in number of cases indicates that something changed in transmissibility of the virus, i.e., loosening of control measures or introduction of a new variant of the virus, which makes it possible to get a reliable estimate of R_0. Our approach is to estimate R_0 from the curve without vaccination (red curve) and use our rule of thumb with this estimate of R_0 to predict the effect of vaccination on the proportion of cases, given vaccine efficacy and proportion vaccinated at the start of the epidemic growth, which were known. We validate the calculated proportion of cases to the curve with vaccination from the simulations (blue curve).

In the default scenario (no vaccination), a fraction 0.296 of the 7.14 million simulated people gets infected (AR_{unv}), in the vaccination scenario, a fraction 0.185 gets infected (AR_{vacc}).

To check if the theory holds, we need an estimate of R_0 for the period at the start of the curve. Complicating factor is that we do not know the fraction susceptible at the start of the new peak (s_0), which is most likely smaller than 1, because part of the population might still have immunity from previous infection.

Therefore, we jointly estimate s_0 and R_0 using a grid search with a GLM. The GLM is fitted as,

$$\frac{C}{S_t} \sim \log(\beta) + \log(\frac{I}{N})$$

Where S_t is the number of susceptibles at timepoint t, and is calculated as $S_0 - \sum_0^t C$, the number of susceptibles at t=0 (S_0, which is $s_0 N$) minus the total number

of cases until t, as counted from the data. $Log(\frac{I}{N})$ is an offset and represents the exposure to infectious individuals at each timepoint. Note that t=0 here corresponds to day 625 in Figure 4.6, which is the starting point of the increase in cases. A range of starting values for s_0 between 0.297 (note that s_0 cannot be smaller than the size of the outbreak) and 1, in steps of 10^{-5}, was tested to determine the optimal fit for the GLM (lowest AIC) for estimation of β. To determine R_0, the estimated value for β from the GLMM with optimal fit is multiplied by the mean duration of the infectious period, 7.35, which was retrieved from the input of the simulations.

The optimal AIC value corresponds to an s_0 of 0.52 and a value for beta of 0.388, which leads to an estimate for R0 of 2.85 (Figure 4.7).

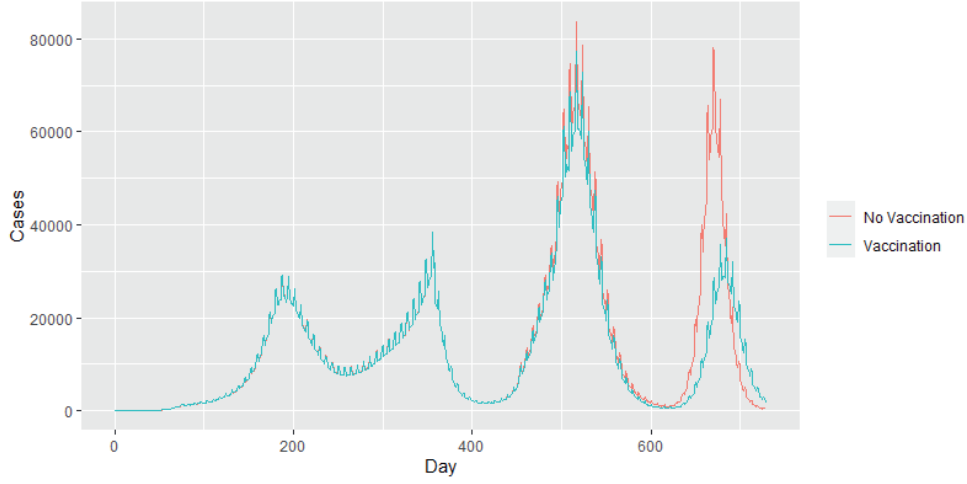


Figure 4.6 Number of cases over time for the COVID19 simulation in Bogota, for the actual scenario with vaccination and an imaginary scenario without vaccination.

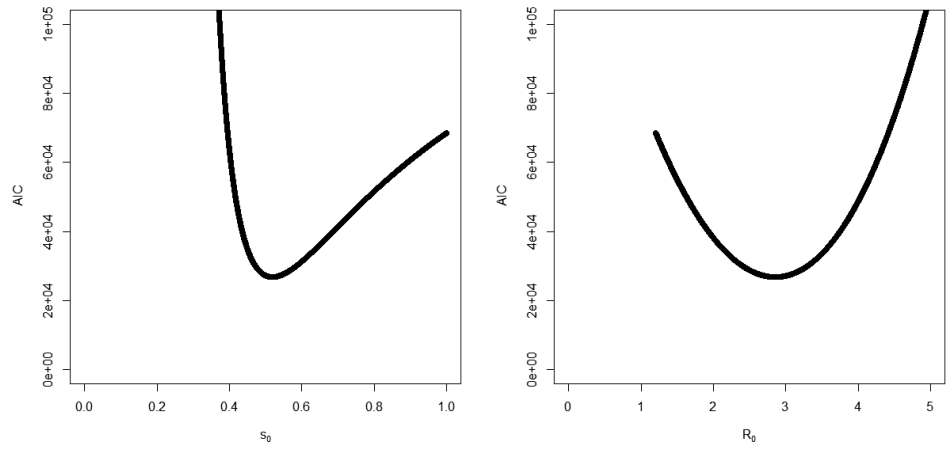


Figure 4.7 AIC plots for s_0 and R_0 from the GLMM optimisation.

Given our theory, an R_0 of 2.85 corresponds to a ratio overall/direct of:

$$\frac{Overall}{Direct} = \frac{R_0}{R_0 - 1}$$

$$\frac{Overall}{Direct} = 1.54$$

Now we need an estimate of the direct efficacy to predict the overall effect of vaccination, and thus the total number of cases when vaccination is applied.

The direct efficacy depends on the efficacy of vaccination. In the simulation model, corresponding to reality, 5 different vaccines are used, each with different efficacy. The applied quantity of each vaccine depends on vaccine deliveries, vaccination capacity, and whether or not (and how many) people get a second dose of the vaccine, which are all known from the simulations. Thus, we calculate the mean direct efficacy of vaccination as a weighed sum of individual vaccine efficacies (note: efficacy against infection), accounting for the fraction singly and doubly vaccinated with a vaccine, since efficacy is different for doubly vaccinated. This yields a mean direct efficacy of 0.502, with corresponding fraction vaccinated $f_{vacc} = 0.486$

The direct efficacy can now be calculated as:

$$Direct\ Efficacy = eff * AR_{unv} * f_{vacc} = 0.072$$

Which can be interpreted as if vaccination would have no indirect effects, the outbreak size would be reduced by 7.2%.

Given the ratio $\frac{Overall}{Direct}$ of 1.54, the overall efficacy will be $0.072 * 1.54 = 0.111$, which means that, based on our rule of thumb, the expected outbreak size with vaccination will be $0.296 - 0.111 = 0.185$. Which exactly corresponds to the fraction that got infected in the vaccination scenario (AR_{vacc}; see above).

Chapter 5

Can breeders prevent pathogen adaptation when selecting for increased resistance to infectious diseases?

Andries D. Hulst[1,2], Piter Bijma[2], Mart C.M. De Jong[1]

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

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

Genetics Selection Evolution (2022) 54:73

Abstract

Recent research shows that genetic selection has high potential to reduce the prevalence of infectious diseases in livestock. However, like all interventions that target infectious diseases, genetic selection of livestock can exert selection pressure on pathogen populations. Such selection on the pathogen may lead to escape strategies and reduce the effect of selection of livestock for disease resistance. Thus, to successfully breed livestock for lower disease prevalence, it is essential to develop strategies that prevent the invasion of pathogen mutants that escape host resistance. Here we investigate the conditions under which such "escape mutants" can replace wild-type pathogens in a closed livestock population using a mathematical model of disease transmission.

Assuming a single gene that confers sufficient resistance, results show that genetic selection for resistance in livestock typically leads to an "invasion window" within which an escape mutant of the pathogen can invade. The bounds of the invasion window are determined by the frequency of resistant hosts in the population. The lower bound occurs when the escape mutant has an advantage over the wild-type pathogen in the population. The upper bound occurs when local eradication of the pathogen is expected. The invasion window is smallest when host resistance is strong and when infection with the wild-type pathogen provides cross immunity to infection with the escape mutant.

To minimise opportunities for pathogens to adapt, under the assumptions of our model, the aim of disease control through genetic selection should be to achieve herd-level eradication of the infection faster than the rate of emergence of escape mutants of the pathogen. Especially for microparasitic infections, this could be achieved by placing animals into herds according to their genetic resistance, such that these herds stay completely out of the invasion window. In contrast to classical breeding theory, our model suggests that multi-trait selection with gradual improvement of each trait of the breeding goal might not be the best strategy when resistance to infectious disease is part of the breeding goal. Temporally, combining genetic selection with other interventions helps to make the invasion window smaller, and thereby reduces the risk of invasion of escape mutants.

5.1 Background

The possibility to use genetic selection as a strategy to combat infectious diseases in livestock has been recognized by animal breeders for a long time (BISHOP AND WOOLLIAMS 2010; BISHOP AND WOOLLIAMS 2014). In 1997, BISHOP AND STEAR (1997) showed that the potential to breed against nematode infections is greater than predicted by common quantitative genetic models. Recent theoretical work confirms the high potential of genetic selection to reduce infectious disease prevalence, that benefits from considerable indirect genetic effects in the transmission of infectious diseases (ANCHE *et al.* 2014; HULST *et al.* 2021; BIJMA *et al.* 2022). An indirect genetic effect (IGE) is an effect of the genes of an individual on the phenotypes of other individuals (GRIFFING 1967; MOORE *et al.* 1997; MUIR 2005). For infectious diseases, these IGE arise because animals that are less likely to get infected are also less likely to infect other animals, simply because they are less often infected themselves. For this reason, genetic selection against infectious disease not only alters prevalence through the effect on an animal itself, but also through reduced exposure of its herd mates.

Apart from genetic selection, a great variety of strategies is and has been used in infectious disease control. Examples of such strategies are vaccination and treatment with antibiotic or antiviral drugs, but also hygienic and biosecurity measures, such as the separation of non-infected populations from infected ones using certification. These strategies have been used to successfully eradicate infections from populations; well-known examples are the eradication of Aujeszky's disease (STEGEMAN *et al.* 1997; ELBERS *et al.* 2000) and bovine tuberculosis (e.g. MORE *et al.* 2015) from many countries, and the global eradication of rinderpest (MARINER *et al.* 2012; EUROPEAN COMMISSION 2020).

However, pathogens are not static, and any intervention strategy may exert selection pressure on the pathogen population. As long as an infection is not eradicated, this leads the pathogens to 'adapt' themselves to the intervention and, eventually, to evolve strategies to escape from it. The widespread antibiotic resistance is probably the most prominent example of this phenomenon, but escape from other interventions also occurs (DAVIES AND DAVIES 2010; READ *et al.* 2015; KENNEDY AND READ 2017). To prevent confusion between resistance of animals to an infection and resistance of pathogens to an intervention, we will

refer to the latter as 'escape'. So, pathogens may escape from the resistance of livestock to disease. When breeding livestock for lower infectious disease prevalence, it seems certainly possible that a pathogen will adapt itself to these selected animals and evolves escape from the genetic resistance of the animals. The recent identification of a new variant of infectious pancreatic necrosis (IPN) virus that causes relatively high mortality in genetically IPN-resistant Atlantic salmon illustrates this expectation (HILLESTAD *et al.* 2021). In plant breeding, such examples of pathogen escape are widespread (e.g. MCDONALD AND LINDE 2002).

Escape of pathogens is a particular concern since genetic change in (micro-)pathogens is usually much faster than the genetic change that artificial selection can create in animals, because of high mutation rates, shorter generation times, and much larger population sizes (e.g. PRICE 1980). As noted by BISHOP AND STEAR (2003), the effects of selection for lower infectious disease prevalence might be diminished when pathogens get the chance to adapt themselves to the selected animals. Thus, to sustainably breed livestock for lower disease prevalence, it is essential to take the evolution of the pathogen into account and to try to prevent pathogen escape from occurring.

For some other interventions, strategies exist to prevent pathogens from developing escape. Therapies for Human Immunodeficiency Virus (HIV) that use a combination of several antiviral medicines at the same time is one example of such a strategy. An individual, mutated, pathogen strain might be able to escape from one or two of the used antivirals, but it is unlikely that it can escape from all, such that the combination of drugs kills all strains of the pathogen that are present in the host (KENNEDY AND READ 2017). Another example is the restricted use of antibiotics. With this strategy, new antibiotics are used as little as possible and, when they are applied, a sufficiently high dose is used, such that all bacteria that are present in the host are killed. In this way the development of bacterial escape is prevented or at least largely delayed (WALSH 2003). These examples illustrate that, to prevent pathogen escape, either no selection pressure should be applied or the selection pressure should be strong enough to fully kill the pathogen population in an individual or a group of individuals.

To our knowledge, strategies to prevent pathogen escape have not been investigated for artificial genetic selection of livestock populations against

infectious diseases. Here, we investigate under which conditions escape mutants of a (micro-)pathogen can develop and spread through a closed livestock population. Specifically, we investigate how invasion of an escape mutant is influenced by the frequency and strength of the resistance of the host and the degree of escape of the pathogen. We develop a mathematical model of infection transmission that accounts for artificial genetic selection on a single resistance gene in the host population and for invasion of escape mutants from the pathogen population. We assume that the resistance gene only affects the propensity of an individual to become infected. Furthermore, for generality, we assume that, at any moment in time, a wide range of pathogen mutants can emerge. Hence, our focus is on the invasion risk of a new mutant that may emerge, rather than on evolutionary change in pathogen virulence, which has been investigated extensively elsewhere (see Discussion). Finally, we aim to identify strategies for genetic selection of livestock for lower infectious disease prevalence that limit or prevent the risk of pathogen escape.

5.2 Methods and results

5.2.1 Outline

In this section, we develop and analyse a mathematical model of infectious disease transmission that allows to investigate the conditions under which escape mutants of a pathogen can invade a host population that is under genetic selection for resistance to infection. The starting point of our analysis is a local, closed, host population (e.g., a herd of cattle without import of animals) that is endemically infected with a pathogen (the 'wild-type' pathogen). Hosts are then genetically selected for a single locus that confers some level of resistance to infection. Here, resistance merely implies that hosts are less likely to get infected with the wild-type pathogen, not necessarily that they cannot get infected at all. The aim of the selection is to reduce the prevalence of the infectious disease and, ultimately, to eradicate it from the local population.

Next, we assume that mutants of the pathogen that can to some degree escape host resistance can arise continuously as long as the wild-type pathogen is present in the host population. In other words, we assume that the escape pathogen is a

mutant of the wild-type strain, so that it can arise only when the wild-type pathogen is (still) present. Our main interest is then to determine whether these escape mutants can invade the endemically infected host population and how the possibility to invade depends on characteristics of the host and of the pathogen, and on the degree of resistance against escape mutants provided by infection with the wild-type pathogen (cross resistance).

We start this section by describing the general epidemiological model that we use as the basis for our study. Next, we introduce genetic variation in the host population to allow for genetic selection for resistance in the host population. Here, we assume that host resistance is determined by a single bi-allelic locus, where resistance is either fully dominant or fully recessive. This model results in two host types: resistant hosts (R) and non-resistant hosts (N). Then we derive expressions for the prevalence of the pathogen in resistant and non-resistant hosts, and for the frequency of resistant hosts needed to eradicate the wild-type pathogen. Next, we further expand the model by allowing pathogen mutants to escape host resistance. Using this model, we assess how the possibility of the mutant to invade the host population depends on the frequency of resistant hosts, the level of resistance provided by the resistance gene, the fitness benefit of the escape mutant in resistant hosts, and the costs of the escape mutation for infection of non-resistant hosts. Finally, we incorporate infection of hosts with both types of pathogen at the same time into the model, to assess the effect of the degree of cross resistance on the possibility of the mutant to invade. To enhance comprehensibility of our complex model, we combined the methods and results sections in a fairly unusual way. The description of each step in the model, as outlined above, is followed by a short section with the results relevant to that step. Table 5.1 shows a notation key.

Table 5.1 Notation key

Symbol	Definition
N	Population size
S_N	Number of non-resistant hosts in the susceptible state
S_R	Number of resistant hosts in the susceptible state
I_{NW}	Number of non-resistant hosts infected with the wild-type pathogen
I_{RW}	Number of resistant hosts infected with the wild-type pathogen
I_{NE}	Number of non-resistant hosts infected with the escape mutant pathogen
I_{RE}	Number of resistant hosts infected with the escape mutant pathogen
I_{NWE}	Number of non-resistant hosts infected with both pathogen types
I_{RWE}	Number of resistant hosts infected with both pathogen types
I_W	Total number of hosts infected with the wild-type pathogen ($I_{NW} + I_{RW} + I_{NWE} + I_{RWE}$)
I_E	Total number of hosts infected with the escape mutant pathogen ($I_{NE} + I_{RE} + I_{NWE} + I_{RWE}$)
f_R	Frequency of resistant hosts
β	Transmission rate parameter
α	Recovery rate parameter ($\alpha = 1$ is assumed throughout)
r	Cross resistance parameter ($r = 1$ for full cross resistance, $r = 0$ for no cross resistance)
P_W^*	Equilibrium prevalence of wild-type pathogen ($(I_{NW}^* + I_{RW}^*)/N$)
P_{NW}^*	Equilibrium prevalence of wild-type pathogen in non-resistant hosts (I_{NW}^*/N)
P_{RW}^*	Equilibrium prevalence of wild-type pathogen in resistant hosts (I_{RW}^*/N)
R_{NW}	Basic reproduction ratio for wild-type pathogen in non-resistant hosts
R_{RW}	Basic reproduction ratio for wild-type pathogen in resistant hosts
R_{NE}	Basic reproduction ratio for escape mutant pathogen in non-resistant hosts
R_{RE}	Basic reproduction ratio for escape mutant pathogen in resistant hosts
R_{INV}	Reproduction ratio of invasion of escape mutant in population endemically infected with wild-type pathogen

5.2.2 SIS-model

Because the purpose of this study was to investigate the basic conditions under which pathogens can escape genetic resistance of hosts, without getting lost in mathematical details of specific cases, we decided to use the relatively simple, but very well-established Susceptible-Infectious-Susceptible (SIS) epidemiological model. This model provides a realistic representation of the transmission of several endemic infectious diseases (HETHCOTE 1989), also in livestock populations.

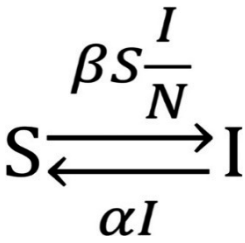


Figure 5.1 Susceptible-Infectious-Susceptible-Model.

In the SIS-model, host individuals can be in the susceptible (S) or the infected (I) state (Figure 5.1). Infected individuals are also infectious, meaning that they can infect susceptible individuals. In the context of the SIS-model, the term "susceptible" merely means that the individual is not infected, so that it can become infected. It does not indicate any degree of susceptibility or resistance of the individual. Individuals 'move' between the two states: susceptible individuals can get infected following contact with an infected individual, while infected individuals can recover from infection and become susceptible again. The transition of individuals between these two states occurs at certain rates. The number of individuals that get infected per unit of time, and thus change from S to I, is known as the transmission rate. The number of individuals per unit of time that recover from infection, and thus change from I to S, is the recovery rate.

The transmission rate depends on the transmission rate parameter (β), on the number of susceptible individuals (S; we will use italics (S and I) to indicate the number of individuals in the corresponding state, and regular font (S and I) to indicate the state of an individual), and on the fraction of the contact individuals

that are infected ($\frac{I}{N}$, with N representing the total size of the local population) (DIEKMANN *et al.* 2012). Thus,

$$Rate[\mathrm{S} \rightarrow \mathrm{I}] = \beta S \frac{I}{N} \tag{5.1}$$

The transmission rate parameter β is the average number of susceptible individuals that become infected per unit of time by one infected individual in an otherwise fully susceptible population (i.e., $S = N$), and reflects the transmissibility of the infection.

The recovery rate depends on the recovery rate parameter (α) and on the number of infected individuals (I), i.e.:

$$Rate[\mathrm{I} \rightarrow \mathrm{S}] = \alpha I \tag{5.2}$$

The recovery rate parameter also determines the average duration of the infectious period, which is equal to $\frac{1}{\alpha}$.

A key parameter in epidemiology is the basic reproduction ratio (R_0), which is defined as the average number of secondary infections caused by a single typical infected individual in an otherwise fully susceptible population (DIEKMANN *et al.* 1990). R_0 has a threshold function, i.e., when it is greater than 1, an infection may persist in the population, and a considerable fraction of individuals may become infected. When R_0 is less than 1, an infection is guaranteed to die out. Thus, to eradicate an infectious disease, it is essential that interventions reduce R_0 to less than 1.

For the SIS-model, a simple expression for R_0 can be derived using the expressions for the transmission and recovery rate (e.g. HETHCOTE 1989). Given that β represents the transmission rate for one infected individual in a fully susceptible population, and that the single infected individual has an average infectious period of $\frac{1}{\alpha}$, the number of secondary infections caused by this individual is equal to β/α. Thus, R_0 is equal to the product of the transmission rate parameter and the duration of the infectious period, i.e.:

$$R_0 = \frac{\beta}{\alpha} \tag{5.3}$$

To simplify the mathematics, we assume a constant value of $\alpha = 1$, such that $R_0 = \beta$. Note that we can use $\alpha = 1$ without loss of generality because one can always choose a time unit such that α is equal to 1 and, therefore, β equals R_0. If R_0 is higher than 1, the SIS-model tends to a situation in which the number of newly infected individuals is equal to the number of recovering individuals, the so-called endemic equilibrium. In that situation, the average numbers of infected and susceptible individuals are stable. The fraction of infected individuals at this equilibrium, i.e., I^*/N, is the endemic prevalence of the infectious disease.

At equilibrium, the transmission rate is equal to the recovery rate:

$$\beta S^* \frac{I^*}{N} = \alpha I^* \underset{R_0 = \frac{\beta}{\alpha}}{\Longleftrightarrow} R_0 S^* \frac{I^*}{N} = I^* \tag{5.4}$$

Solving for I^*/N, using $S^* = N - I^*$ shows that the endemic prevalence ($P^* = I^*/N$) is determined by R_0:

$$P^* = 1 - \frac{1}{R_0} \tag{5.5}$$

Having defined the fundamental epidemiological model and obtained expressions for R_0 and the endemic prevalence, we will expand the model in the next section by including genetic differences in host resistance.

5.2.3 Heterogeneous SIS-model with host resistance to wild-type infection

In this section, we present a model for genetic variation in host resistance for a population that is exposed to the wild-type pathogen and present the results of this model. While the term "resistance" might suggest that resistant individuals cannot get infected at all, in practice resistance is rarely all-or-none. Hence, in the following, resistant merely means being less likely to get infected.

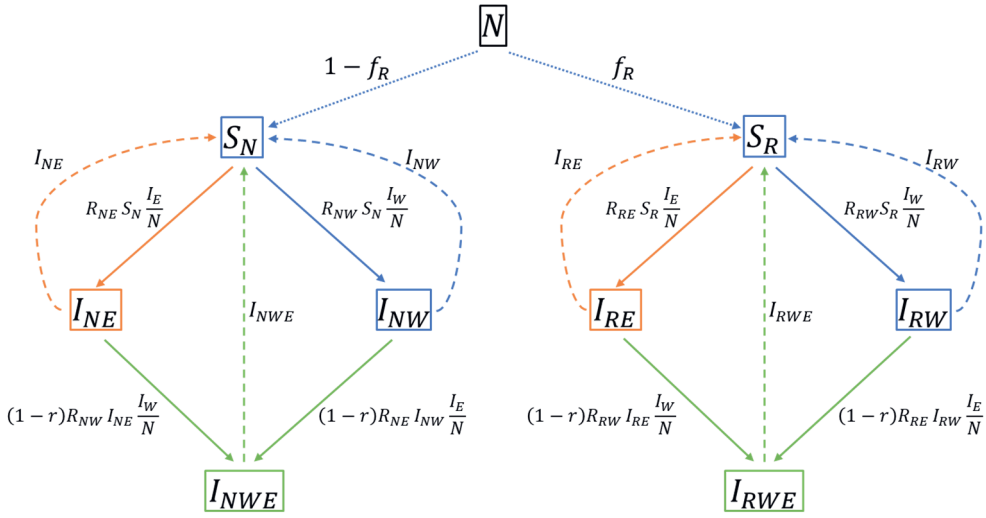


Figure 5.2 SIS-model with resistant and non-resistant hosts and wild-type and escape-type pathogens. SIS-model with resistant and non-resistant hosts (Blue subsystem), pathogens either of the wild type (subscript W) or that can escape host resistance (subscript E; Orange subsystem), and infection of hosts with both types of pathogen (Green subsystem). For explanation of symbols, see Table 5.1. I_W and I_E denote the total number of individuals infected with a wild-type and escape pathogen, respectively, and are the sum taken over both host types, where $I_W = I_{NW} + I_{RW}$ and $I_E = I_{NE} + I_{RE}$ respectively.

When properties of an individual or a pathogen affect the transmission of an infectious disease, they essentially alter the values of the underlying parameters, i.e., β and/or α, and thus R_0, since these parameters fully encompass the characteristics of the infection in the SIS-model (for examples, see also [26-28]). To include an effect of host resistance on transmission, we decrease R_0 by decreasing β (note that it does not matter for the results whether a decrease in R_0 originates from a reduction in β or an increase in α; see Discussion). Furthermore, we assume that, once infected, both non-resistant and resistant hosts are equally likely to infect susceptible individuals. This means that there is no difference in the "infectivity per unit of time" between resistant and non-resistant hosts.

Now that we have defined how host resistance acts on transmission of the infection, the next step is to define the genetic model for host resistance. Given that derivations for transmission models with many levels of host resistance

rapidly become very complex, we will use one of the simplest genetic models for diploid individuals. We assume that resistance is defined by a single locus, which is either fully dominant or fully recessive, such that the host population consists of two types: resistant (with frequency f_R) and non-resistant ($1 - f_R$). Thus, in this model, we have two types of hosts, resistant (R) and non-resistant (N), and one type of pathogen, the wild-type pathogen (W). Using this genetic model, we can set up a SIS-model with two types of hosts, resistant hosts (subscript R) and non-resistant hosts (subscript N; this is the blue subsystem in Figure 5.2 and the full model equations are in Appendix 5.6.1). The number of susceptible hosts of each type in the local population are denoted by S_R and S_N. Analogously, the number of hosts of each type that are infected with the wild-type pathogen (subscript W) are denoted by I_{RW} and I_{NW}.

Transmission involves a pair of individuals: an infected donor individual, and a susceptible recipient individual. Here, we assume that the pair-wise transmission rate parameter depends on the resistance genotype of the recipient but not on the genotype of the donor. This is equivalent to the absence of genetic variation in infectivity. Since we use $\alpha = 1$, the transmission rate parameters are directly equal to the basic reproduction ratios (Equation 5.3). In a contact between an infected and a susceptible (i.e., non-infected) host, these basic reproduction ratios (R_{NW} and R_{RW}) are defined for the recipient individual, because we model variation in resistance. The subscript (N or R) thus reflects the genotype of the susceptible host (W refers to the wild-type pathogen).

Just like the homogeneous SIS-model, the heterogeneous SIS-model also tends to an equilibrium. However, because of the difference in resistance between individuals, the endemic prevalence in this equilibrium is no longer equal to Equation 5.5 because non-resistant individuals are more likely to become infected than resistant individuals. Consequently, at equilibrium, the susceptible individuals are predominantly of the resistant type, which are less likely to become infected. As a result, the overall endemic prevalence at equilibrium is a bit lower than in the model without heterogeneity (DIEKMANN AND HEESTERBEEK 2000; SPRINGBETT *et al.* 2003) and is equal to:

$$P_W^* = \frac{I_{NW}^* + I_{RW}^*}{N} \tag{5.6}$$

and is reached when both host types have reached their equilibrium (BIJMA *et al.* 2022), i.e.:

$$R_{NW}S_N^* \frac{I_W^*}{N} = I_{NW}^* \text{ and } R_{RW}S_R^* \frac{I_W^*}{N} = I_{RW}^* \tag{5.7}$$

Solving Equations 5.6 and 5.7 using $S_N^* = (1 - f_R)N - I_{NW}^*$ and $S_R^* = f_R N - I_{RW}^*$, results in equilibrium solutions for the endemic prevalence in both host types and for the overall endemic prevalence in the population. The resulting equations are complex and are given in Appendix 5.6.2, together with detailed derivations (Equations 5.12 to 5.18). We will use figures to illustrate the results.

5.2.4 Results for the heterogeneous SIS-model with host resistance to wild-type infection

We consider the situation where the wild-type pathogen is endemic in a non-resistant host population, while the infection is absent in a population where all individuals have the resistant genotype due to herd immunity ($R_0 < 1$). This represents the most beneficial situation for genetic selection for host resistance, because it will lead to eradication of the infection (in this section, escape mutants are ignored). This situation corresponds to $R_{NW} > 1$, so that $R_0 > 1$ when $f_R = 0$, i.e. the infection is endemic in a non-resistant host population, and $R_{RW} < 1$, so that $R_0 < 1$ when $f_R = 1$, i.e. the infection is absent in a resistant host population. Note that $R_{RW} < 1$ results in the infection to be absent because of herd immunity, but this does not imply that resistant individuals cannot get infected at all (see also results below). We will also show a situation where $R_{RW} > 1$, to illustrate what will happen when the resistance is not sufficient to fully eradicate the infection.

In a population with both resistant and non-resistant individuals, the prevalence and whether the infection is present or not, depend on the frequency of resistant hosts (f_R). Figure 5.3 shows the endemic prevalence of the infection (P_W^*) in a population that consists of a mix of resistant and non-resistant hosts, as a function of f_R, and for three values of the basic reproduction ratio for resistant hosts (R_{RW}). The basic reproduction ratio for non-resistant hosts (R_{NW}) was set to 1.5. Figure 5.3 clearly shows that the overall prevalence decreases with increasing frequency of resistant hosts, at a rate that depends on the value of R_{RW}, i.e., prevalence

decreases faster when R_{RW} is lower. At a low frequency of resistant hosts, virtually all infected hosts are from the non-resistant type (P^*_{NW}; dashed line in Figure 5.3), which makes sense given the low frequency of resistant hosts and their lower susceptibility. Note that prevalence is expressed relative to the total population size and not to the number of individuals of a given type, *i.e.,* as I^*_{RW}/N rather than $I^*_{RW}/(f_R N)$. At a higher frequency of resistant hosts (e.g., $f_R = 0.25$), a larger fraction of the infections occurs in resistant hosts (P^*_{RW}; dashed-dotted line in Figure 5.3). This happens even when the reproduction ratio of resistant hosts is lower than 1, because the overall R_0 is higher than1, leading to maintenance of the infection in the population. For R_{RW} of 0.1 and 0.8, the overall prevalence decreases to 0 at a certain f_R. This is the frequency of resistant hosts above which the infection is expected to die out, because the greater resistance of the population reduces the overall reproduction ratio of the infection to less than 1 (herd immunity). For R_{RW} of 1.1, this point does not exist, because the basic reproduction ratio in a fully resistant population is still above the threshold of 1, implying that the infection will persist in the population.

For $R_{RW} < 1$, the f_R at which the infection dies out can be found by realising that the overall R_0 should be higher than 1 for the infection to persist in the population. For our model, the overall basic reproduction ratio is the average of the type-specific reproduction ratios, weighed by the frequencies of both types (DUSHOFF AND LEVIN 1995), i.e.:

$$R_0 = (1 - f_R)R_{NW} + f_R R_{RW} \tag{5.8}$$

Solving $R_0 = 1$ for f_R results in the upper limit of f_R, below which the infectious disease can persist in the population, i.e.:

$$f_{R_{max}} = \frac{R_{NW} - 1}{R_{NW} - R_{RW}} \tag{5.9}$$

Using the values for R_{NW} and R_{RW} from Figure 5.3, we find solutions for $f_{R_{max}}$ of $0.5/1.4 = 0.36$ for $R_{RW} = 0.1$ and $0.5/0.7 = 0.71$ for $R_{RW} = 0.8$. These results represent the minimum frequency of hosts with the resistant genotype in a population that is required to eradicate an infection, which is relevant for selection.

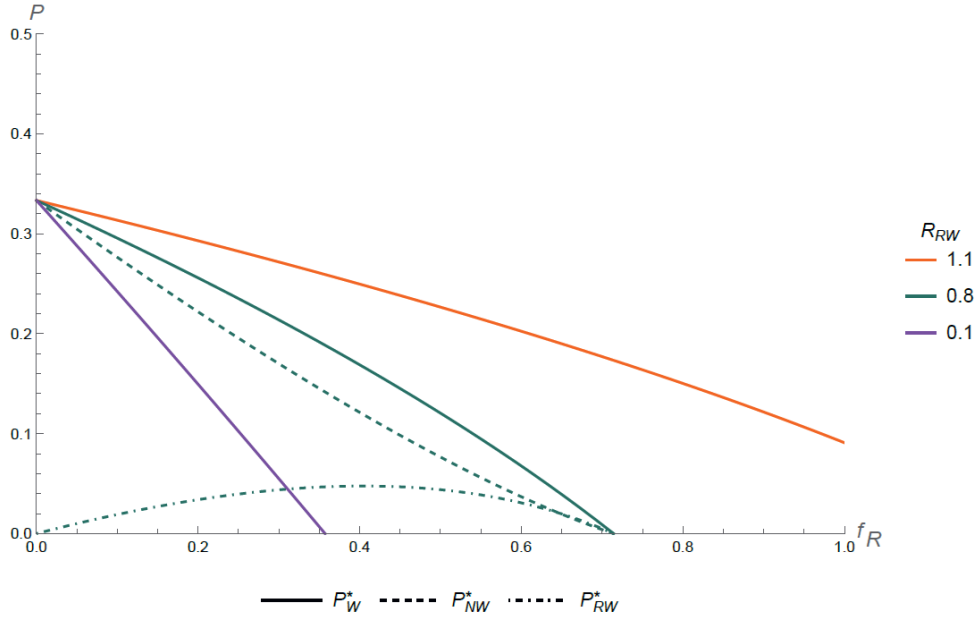


Figure 5.3 Prevalence of the infection with the wild-type pathogen. Equilibrium prevalence of the infection with the wild-type pathogen ($P_W^* = \frac{I_W^*}{N}$), as a function of the frequency of resistant hosts f_R for three different values of R_{RW}, $R_{NW} = 1.5$, such that the endemic prevalence of the wild type in a fully non-resistant population is 0.33 (Equation 5.5). For $R_{RW} = 0.8$, the fractions of infected resistant ($P_{RW}^* = \frac{I_{RW}^*}{N}$) and non-resistant ($P_{NW}^* = \frac{I_{NW}^*}{N}$) hosts are also shown.

In a closed population, $f_{R_{max}}$ also sets an upper bound to the invasion possibility of escape mutants, because the infection with the wild-type pathogen dies out when $f_R > f_{R_{max}}$. When the wild-type infection has died out, mutants cannot develop anymore in a closed local population, simply because there are no wild-type pathogens to mutate from. The next section provides further details on the important role of $f_{R_{max}}$ when the model is expanded by allowing for the development of pathogen mutants that can escape host resistance.

5.2.5 Heterogeneous SIS-model with introduction of an escape mutant

When we want to include pathogen mutants that can escape host resistance into our model, where resistance is not absolute, pathogen escape entails that the mutant pathogen is better able to infect resistant hosts than the wild-type

pathogen. We will incorporate one escape mutant into the model, such that there are two pathogen types, a wild-type pathogen (W) and an escape mutant (E), each of which can infect both host types, albeit the escape mutant will more easily infect a resistant host than the wild-type pathogen. This does not mean that we only consider one type of mutant, it merely means that we will look at the competition between a certain mutant and the wild-type pathogen one at a time. Furthermore, we do not consider the probability that an escape mutation occurs but focus on the case where escape mutants are present, because a mutation will occur sooner or later, as long as the wild-type pathogen is present in the population.

In this section, we assume that infection with one of the two pathogen types offers full cross-resistance against infection with the other type, such that hosts can be infected with only one pathogen type at a time (this assumption will be relaxed in the next section). It means that we consider two additional types of infected individuals, i.e., N and R host types that are infected with the escape (E) mutant, such that we have four types of infected individuals in total (the orange subsystem in Figure 5.2; note that we still have two types of susceptible individuals). We will model transmissibility of escape mutants by including two additional reproduction ratios in our model, such that transmission between a pair of individuals no longer depends only on the type of susceptible host (N or R), but also on the type of pathogen that is carried by the infected host (W or E). These two reproduction ratios are for the escape mutant infecting non-resistant or resistant susceptible hosts, denoted by R_{NE} and R_{RE}, respectively. Selection pressure on the pathogen population is then determined by the four reproduction ratios and the frequency of resistant hosts in the population.

Mutants typically arise in a host population that is endemically infected with the wild-type pathogen. Thus, to determine whether an escape mutant can spread in the host population, we need to derive an expression for the reproduction ratio of the escape mutant in a host population in which the wild-type pathogen is at endemic equilibrium. We will call this the invasion reproduction ratio of the escape mutant (R_{INV}). When $R_{INV} > 1$, the escape mutant can invade a host population that is endemically infected with the wild-type pathogen. When $R_{INV} < 1$, escape mutants might occur, but they cannot spread.

We can derive R_{INV} by applying the definition of R_0, but this time for a population where the wild-type pathogen is endemic, instead of for a fully susceptible population. To find R_{INV}, given full cross-resistance (i.e. hosts that are infected with one of the pathogen types cannot simultaneously get infected with the other type), we need to multiply the basic reproduction ratios of the escape mutant with the fraction of susceptible individuals (*i.e.,* all individuals of a given genotype that are not yet infected with the wild-type pathogen). These fractions are given by $(1 - f_R) - P^*_{NW}$ for non-resistant hosts, and by $f_R - P^*_{RW}$ for resistant hosts. Then, the reproduction ratio for invasion of the escape mutant into a host population that is endemically infected with the wild-type pathogen becomes:

$$R_{INV} = R_{NE}[(1 - f_R) - P^*_{NW}] + R_{RE}(f_R - P^*_{RW}) \tag{5.10}$$

R_{INV} thus depends on f_R and the four reproduction ratios, two of which are shown explicitly in Equation 5.10 (R_{NE}, R_{RE}), while the other two (R_{NW} and R_{RW}) are implicit in the expressions for the endemic equilibrium prevalences P^*_{NW} and P^*_{RW} (see Equations 5.16 and 5.17 in Appendix 5.6.2).

5.2.6 Results for the heterogeneous SIS-model with introduction of an escape mutant

Figure 5.4 shows R_{INV} (dashed line, right y-axis) as a function of f_R for reproduction ratios $R_{NW} = 1.5$, $R_{NE} = 1.1$, $R_{RE} = 1.5$, and $R_{RW} = 0.8$ (Figure 5.4a) or 0.1 (Figure 5.4b). The endemic prevalence of the wild-type pathogen (P^*_W; solid line; left y-axis) is shown in both panels as a function of f_R, similar to Figure 5.3. We chose to simulate $R_{RW} < 1$, such that the wild-type pathogen will die out in a fully resistant population (see previous section). This is visible in Figure 5.4 as the solid line that decreases to $P^*_W = 0$. By assumption, the escape mutation comes with a cost in fitness for the pathogen, such that the escape mutant will be outcompeted by the wild-type pathogen in a non-resistant population. However, due to its adaptation to the resistant host, it can persist in a fully resistant population, in contrast to the wild-type pathogen. This is visible in Figure 5.4 as the increasing dashed line.

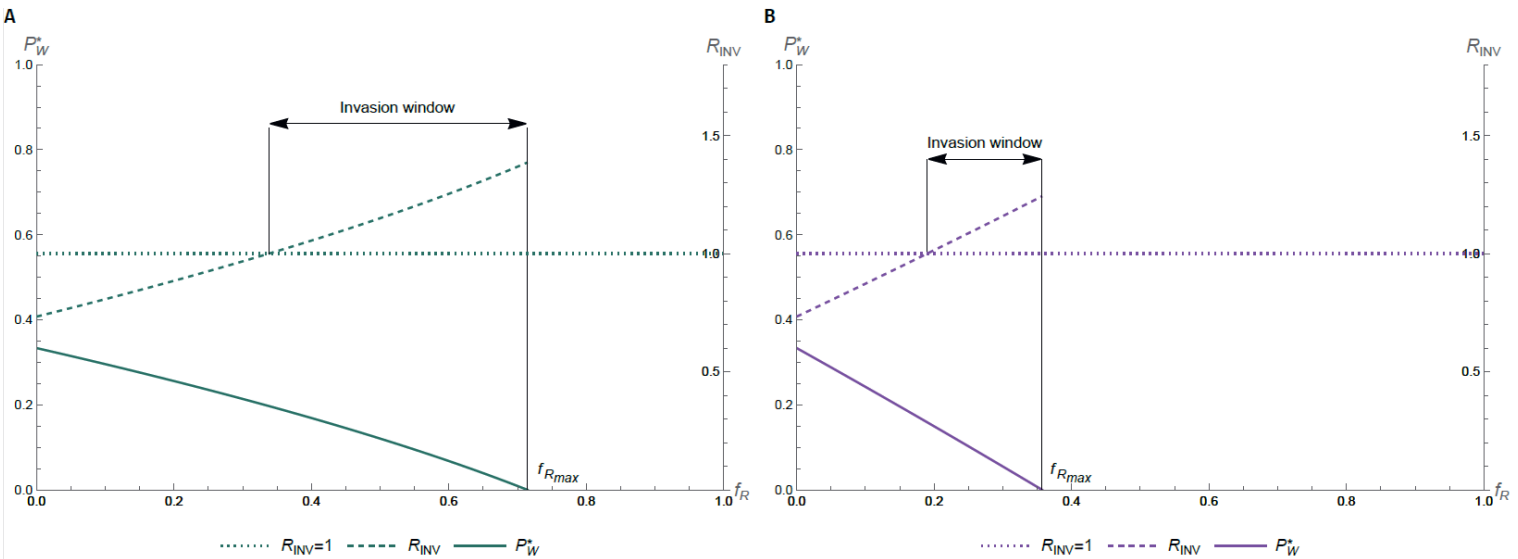


Figure 5.4 Invasion window with full cross resistance. Reproduction ratio for invasion of an escape mutant in a population endemically infected with the wild-type pathogen as a function of f_R (R_{INV}, dashed line, right y-axis), for $R_{RW} = 0.8$ (panel (**a**)) and $R_{RW} = 0.1$ (panel (**b**)). The dotted line shows the threshold value for R_{INV} of 1. Hosts can be infected with only one type of pathogen at a time (full cross resistance). Prevalence of the wild-type pathogen as a function of f_R is given by the solid line, indicated on the left y-axis, comparable to Figure 5.3. The window in which there is a risk of invasion of the escape mutant is indicated. $R_{NW} = 1.5, R_{NE} = 1.1, R_{RE} = 1.5$.

At a certain f_R, the reproduction ratio for invasion of the escape mutant (R_{INV}; dashed line in Figure 5.4) crosses the threshold $R_{INV} = 1$ (indicated by the dotted line in Figure 5.4), above which the escape mutant can invade the population. This is a critical point that represents the lower bound of the invasion window, there is a risk of escape mutants invading the (partly) resistant host population when the frequency of resistant hosts is greater than this lower bound. Together with the upper bound of the wild-type pathogen dying out, as defined in the previous paragraph, the point $R_{INV} = 1$ determines the critical range of the frequency of resistant hosts in which an escape mutant pathogen can both arise ($P_W^* > 0$) and invade ($R_{INV} > 1$). This range of f_R is indicated as the 'invasion window' in Figure 5.4.

Comparing panels (a) and (b) of Figure 5.4, we see that a stronger effect of the resistance gene of the host (lower R_{RW}) not only results in a lower f_R that is required to eradicate the wild-type pathogen ($0.36 < 0.71$), but also in a narrower invasion window, because of a lower upper bound. A stronger effect of the resistance gene (*i.e.*, lower R_{RW}) thus reduces the risk of invasion of escape mutants because it results in a smaller invasion window. A smaller invasion window can be passed more quickly by artificial selection of the host population.

In summary, in this section we have derived an expression for the reproduction ratio for the invasion of an escape mutant in hosts that are endemically infected with the wild-type pathogen, assuming that hosts can be infected with only one pathogen at a time (full cross-resistance). In the next section, we will relax this assumption and expand our model to allow for infection of hosts with both types of pathogen at the same time.

5.2.7 Heterogeneous SIS-model, allowing for double infections

Although the two pathogen types are closely related, infection with the one type may not provide full resistance to infection with the other type. Coexistence of different strains (incomplete cross-resistance) is common for many bacteria (e.g. STOESSER *et al.* 2015; DAVIDSON *et al.* 2016; NIGSCH *et al.* 2021). However, for viruses such as influenza A, infection with the wild-type pathogen may give substantial resistance to infection with the escape mutant and *vice versa* (EPSTEIN AND PRICE 2010). If cross-resistance is not complete, hosts that are already

infected with one pathogen type can get infected with another as well. Thus, the possibility of double infection leads to an additional transmission route, from singly-infected to doubly-infected hosts. If some cross-resistance is present, this leads to a lower rate for a second infection compared to the rate of first infection. We will model this effect by including a cross-resistance factor $(1-r)$ in the transmission rate from singly- to doubly-infected hosts (the green subsystem in Figure 5.2). Parameter r takes values between 0 and 1; 0 for no cross-resistance and 1 for full cross-resistance (no double infections occur). For example, non-resistant hosts that are infected with the wild-type pathogen get infected with the escape mutant at rate $(1-r)R_{NE}I_{NW}\frac{I_E}{N}$. Once infected, we assume that doubly-infected hosts are equally likely to infect a susceptible host with one of the two pathogen types as singly-infected hosts.

To determine whether an escape mutant can invade under these conditions, we need to adapt Equation 5.10 because susceptible individuals as well as individuals that are already infected with the wild-type pathogen can become infected by the escape mutant. The equation for the invasion reproduction ratio of an escape mutant in a population that is endemically infected with the wild-type pathogen then becomes:

$$R_{INV} = R_{NE}[(1-f_R) - P^*_{NW} + (1-r)P^*_{NW}] + R_{RE}[f_R - P^*_{RW} + (1-r)P^*_{RW}] \quad (5.11)$$

This expression allows us to investigate how invasion of an escape mutant is affected by incomplete cross-resistance.

5.2.8 Results for the heterogeneous SIS-model that allows for double infections

Figure 5.5 shows the effect of incomplete cross-resistance on the invasion window of the escape mutant. Figure 5.5 is similar to Figure 5.4 (which implicitly has $r = 1$), but now R_{INV} is shown for three levels of cross-resistance (0, 0.5 and 1). Reproduction ratios are the same as those used in Figure 5.4. Incomplete cross-resistance ($r < 1$) increases the width of the invasion window. When $r = 0$, the window covers the whole range in f_R where the wild-type pathogen is present. This happens because R_{NE} is greater than 1 and, when $r = 0$, all hosts infected with the wild-type pathogen can still get infected with the escape mutant as well,

because, in this case, there is no competition between the pathogen types. As without double-infected hosts, a lower R_{RW} (Figure 5.5b) decreases the width of the invasion window. In the next section, we will further investigate how the reproduction ratios affect the width of the invasion window at different levels of cross-resistance.

5.2.9 Factors affecting the width of the invasion window

As shown in the previous paragraphs, the invasion opportunity of escape mutants in a host population that is genetically selected for resistance is restricted by two bounds: (1) a lower bound that represents the degree of resistance of the population at which the invasion reproduction ratio (R_{INV}) becomes greater than 1, and (2) an upper bound that represents the point above which the wild-type pathogen dies out ($f_{R_{max}}$, Equation 5.9). Since we obtained an expression for $f_{R_{max}}$, we tried to solve $R_{INV} = 1$ for f_R algebraically, to determine the effect of the four reproduction ratios and the level of cross-resistance on the width of the invasion window. Unfortunately, this resulted in a closed form solution only when there is either no cross-resistance or full cross-resistance ($r = 0$ or $r = 1$). Thus, we will investigate the width of the invasion window numerically by taking one set of values for the four reproduction ratios as the default scenario ($R_{NW} = 1.5, R_{RW} = 0.8, R_{NE} = 1.1, R_{RE} = 1.5$), and then vary one of them at a time (except R_{NW}). For cross-resistance, r will take values of 0, 0.5, and 1.

Figure 5.6a shows the size of the invasion window as a function of the frequency of resistant hosts and the basic reproduction ratio of the wild-type pathogen in resistant hosts (R_{RW}). We can see that the window becomes smaller with decreasing R_{RW}, as was already visible in Figures 5.4 and 5.5. As stated before, this implies that a stronger effect of the resistance gene (*i.e.,* lower R_{RW}) decreases the risk of invasion of an escape mutant, because the wild-type pathogen goes extinct sooner. The decreasing width of the window with a lower R_{RW} is mainly caused by the effect on $f_{R_{max}}$ (solid dark green line in Figure 5.6a), which becomes closer to 1 with increasing R_{RW}. Hence, a change in R_{RW} mainly impacts the upper bound of the invasion window.

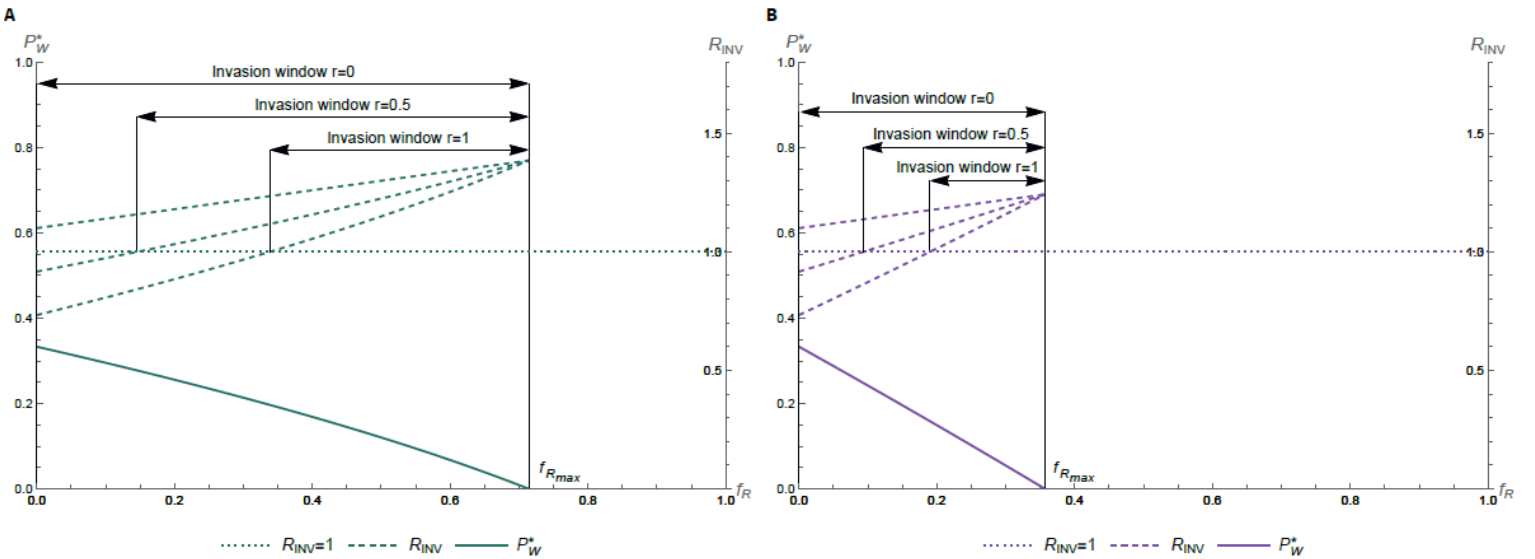


Figure 5.5 Invasion window for different levels of cross-resistance. Reproduction ratio for invasion of an escape mutant in a population endemically infected with the wild-type pathogen as a function of f_R (R_{INV}, dashed lines, right y-axis), for cross resistance (r) levels of 0, 0.5 and 1 and for $R_{RW} = 0.8$ (panel (a)) and $R_{RW} = 0.1$ (panel (b)). The dotted line shows the threshold value for R_{INV} of 1. Prevalence of the wild-type pathogen as a function of f_R is given by the solid line, indicated on the left y-axis. The invasion window for the different cross-resistance levels is indicated. $R_{NW} = 1.5$, $R_{NE} = 1.1$, $R_{RE} = 1.5$.

The lower bound is less affected by a change in R_{RW}; and especially for low r, the line $R_{INV} = 1$ is almost vertical in Figure 5.6a. This happens because at lower levels of cross-resistance, there is less competition between the two pathogen types, since (part of) the hosts that are infected with one pathogen type can also become infected with the other type. In the most extreme case, when there is no cross-resistance ($r = 0$), the lower bound is not affected by the frequency of resistant hosts (Figure 5.6a). This means that R_{RW}, and thus the endemic equilibrium prevalence of the wild-type pathogen, has no effect on the risk of invasion of an escape mutant when $r = 0$.

We restricted the y-axis of Figure 5.6a to a maximum of 1, because the pathogen can persist in resistant hosts when $R_{RW} > 1$, and it is thus impossible to eradicate the wild-type pathogen. Genetic selection for infectious disease resistance will in that case not be sustainably beneficial because escape mutants will eventually occur, even at a high frequency of resistant hosts. Thus, to prevent invasion of escape mutants, R_{RW} needs to be less than 1, either through the effect size of the resistance genes or by taking additional measures in combination with genetic selection to achieve R_{RW} less than 1.

Figure 5.6b shows the size of the invasion window as a function of the frequency of resistant hosts and the basic reproduction ratio of the escape mutant in non-resistant hosts (R_{NE}). The value of R_{NE} relative to R_{NW} reflects the fitness costs for the escape mutant in a non-resistant host population, relative to the wild-type pathogen. If R_{NE} is much lower than R_{NW}, the escape mutant spreads much less in non-resistant hosts than the wild-type pathogen, i.e., the escape mutation comes with high fitness costs (low *y*-axis value). If R_{NE} is close or equal to R_{NW}, differential fitness costs are low or absent and the spread of the escape mutant in non-resistant hosts is similar to the spread of the wild-type pathogen (high *y*-axis value). Consequently, the size of the invasion window increases as R_{NE} moves closer to R_{NW}. At a certain R_{NE}, the invasion window covers the whole range of f_R in which the wild-type pathogen can persist. When there is full cross-resistance (top dashed line, $r = 1$), this point occurs when R_{NE} is equal to R_{NW} (= 1.5), meaning that the escape mutant spreads equally well as the wild-type pathogen in non-resistant hosts. With no cross-resistance (bottom dashed line, $r = 0$), this point occurs when R_{NE} is equal to 1. If there is no cross-resistance, the presence

or absence of the wild-type pathogen has no effect on the spread of the escape mutant and, thus, a basic reproduction ratio greater than 1 is sufficient for the mutant to persist. We can also see in Figure 5.6b that the invasion window stays relatively small for lower values of R_{NE} and only covers a range in f_R of about 0.05 when R_{NE} is close to 0. This indicates that it is extremely difficult for escape mutants to invade when they experience high fitness costs in non-resistant hosts.

Figure 5.6c shows the size of the invasion window as a function of the frequency of resistant hosts and of the basic reproduction ratio of the escape mutant in resistant hosts (R_{RE}). The size of the invasion window increases with R_{RE}, but less so when cross-resistance is high. If there is a little cross-resistance, the mutant can already invade if R_{RE} is slightly less than 1 (top dashed lines for $r = 0.5$ and $r = 1$ in Figure 5.6c), because $R_{NE} > 1$ (1.1) and $R_{RW} < R_{RE}$ in that case, *i.e.,* host resistance is more effective against the wild-type pathogen than against the escape mutant. Thus, in a very small range of f_R, "escape" mutants might invade even though their reproduction ratio in resistant hosts is less than 1. The increasing slope of the line $R_{INV} = 1$ when there is no cross-resistance (bottom dashed line in Figure 5.6c, $r = 0$), mainly results from R_{NE} being greater than 1. Figure 5.6b shows that at an R_{NE} of 1.1 and no cross-resistance, the invasion window covers the full range in f_R from 0 to $f_{R_{max}}$. So, in a population that consists of mainly non-resistant hosts (left side of Figure 5.6), the escape mutant will be well able to invade, as explained in a previous paragraph. At a higher frequency of resistant hosts, as with the wild-type pathogen, the transmission of the escape mutant in resistant hosts becomes more and more determining for the invasion probability, indicated by R_{RE}. If R_{RE} is less than 1, host resistance is not only effective against the wild-type pathogen but also against the "escape" mutant, which is evident for the increase in the lower bound of the invasion window (the line $R_{INV} = 1$) with increasing f_R for $r = 0$ in Figure 5.6c. The combination of reproduction ratios of the escape mutant in non-resistant and resistant hosts, together with cross-resistance, thus determines its opportunity to invade.

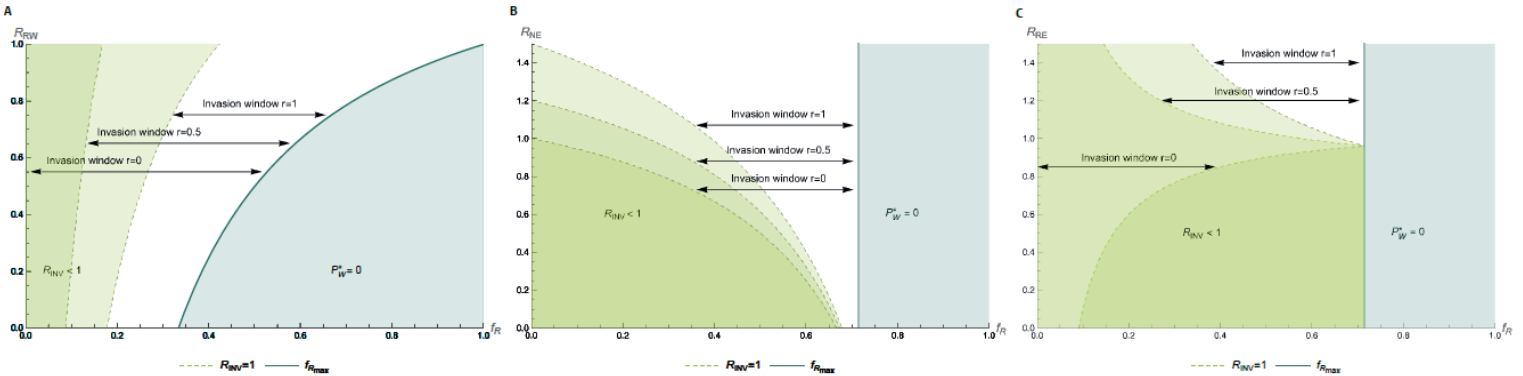


Figure 5.6 Size of the invasion window for different values of R_{RW}. Value of f_R for the lower bound of the invasion window, $R_{INV} = 1$, as a function of the basic reproduction ratio of the wild-type pathogen in resistant hosts (R_{RW}) (panel (**a**)), the basic reproduction ratio of the escape mutant in non-resistant hosts (R_{NE}) (panel (**b**)), and the basic reproduction ratio of the escape mutant in resistant hosts (R_{RE}). For levels of cross resistance of 0, 0.5 and 1 (dashed, light green lines). In panel (**a**), for cross resistance of 0, the line lies on the y-axis. The value of f_R from which the wild-type pathogen dies out (upper bound of the invasion window) is shown by the solid green line. The invasion window for the different cross-resistance levels is indicated by the arrows. For $R_{NW} = 1.5$, $R_{RW} = 0.8$ (panels (**b**) and (**c**)), $R_{NE} = 1.1$ (panels (**a**) and (**c**)), and $R_{RE} = 1.5$ (panels (**a**) and (**b**))

5.3 Discussion

Using a mathematical model of disease transmission, we showed that an infectious disease can be eradicated by increasing the frequency of genetically-resistant hosts to a sufficient level, if the reproduction ratio of the wild-type pathogen in resistant hosts is less than 1. Eradication was achieved by herd immunity in the local population, when R_0 falls below a value of 1, which requires neither that all individuals in the local population are resistant, nor that resistant individuals are fully resistant. The minimum frequency of resistant hosts needed to eradicate an infection ($f_{R_{max}}$, Equation 5.9) is a function of the reproduction ratios of the pathogen in resistant and non-resistant hosts.

As long as the wild-type pathogen is not eradicated (i.e., the frequency of resistant hosts is not high enough), escape mutants of the pathogen can arise and invade the population. To determine whether escape mutants can invade or not, we derived a reproduction ratio for escape mutants in a population that is endemically infected with the wild-type pathogen (R_{INV}, Equation 5.11). Invasion of escape mutants then depends on the reproduction ratios of the wild-type pathogen in resistant and non-resistant host types (through the endemic prevalence), the frequency of resistant hosts, and the basic reproduction ratios of the escape mutant in non-resistant and resistant hosts. We identified an 'invasion window', i.e., the range for the frequency of resistant hosts within which there is a risk of escape mutant invasion. The lower bound of this window is the frequency of resistant hosts above which the reproduction ratio of the invading escape mutant (R_{INV}) becomes greater than 1. The upper bound of the invasion window is the frequency of resistant hosts above which the wild-type pathogen dies out ($f_{R_{max}}$). The invasion window is wider, implying more opportunity for escape mutants to invade, when the resistance allele has a smaller effect and when hosts can get infected by both types of pathogen at the same time (i.e., cross-resistance is not complete).

5.3.1 Model Assumptions

In our model, we assumed that host resistance was determined by a single, fully dominant or fully recessive locus. This assumption resulted in a relatively simple model, with only two types of hosts, which helped in the interpretation of the results. Although the assumption of a single locus determining resistance might only be realistic for very few cases, our predictions will also hold if a quantitative trait locus (QTL) explains a considerable part of the genetic variation, preferably strong enough to bring R_{RW} below 1. Such a major QTL was for instance found for infectious pancreatic necrosis (IPN) resistance in Atlantic salmon (MOEN *et al.* 2009). In most other cases, infectious disease resistance is likely a polygenic trait that is affected by many QTL, each explaining only a small part of the genetic variation, as found for instance by genome-wide association studies (GWAS) for mastitis and digital dermatitis resistance in dairy cattle (TIEZZI *et al.* 2015; BIEMANS *et al.* 2019b). With polygenic resistance, eradication of the wild-type pathogen will also be needed to prevent invasion of escape mutants. Hence, our results are relevant for that case also. However, if polygenic resistance comes with multiple resistance mechanisms, pathogens will be less likely to evolve escape mutations to all mechanisms, and predictions become more complex. We will further elaborate on this in the section 'Implications for animal breeding' below.

Key to the general patterns that emerge from our models are the values for the reproduction ratios. We assumed $R_{NW} > 1$, such that the wild-type pathogen is endemic in non-resistant hosts, and $R_{RW} < 1,$ such that resistance is effective against the wild-type pathogen. Thus, before starting selection for increased host resistance, it is important to know the values of R_{NW} and R_{RW}. Values for R_{NW} for a specific infection are probably already available or can be estimated from field data, e.g., based on the endemic prevalence of the infection (Equation 5.5). Values for R_{RW} could be obtained from transmission experiments, such as those that are used to estimate the effectivity of vaccines (DE JONG AND KIMMAN 1994; VELTHUIS *et al.* 2003; VAN DER GOOT *et al.* 2005; DOESCHL-WILSON *et al.* 2018), or from knowledge of QTL-effects on resistance (susceptibility) or recovery, for example from GWAS.

Estimating the basic reproduction ratios for the escape mutant is more difficult, since these mutants have not yet emerged when selection starts. Nevertheless, in our opinion, some general predictions of the typical properties of escape mutations can be made. We argue that the wild-type pathogen has likely evolved over a long time to be adapted to the non-resistant host. This implies that a typical escape mutant will probably be less well adapted to non-resistant hosts, and thus will have a lower reproduction ratio than the wild-type pathogen in these hosts ($R_{NE} < R_{NW}$). In other words, the escape mutation typically comes with a cost in pathogen fitness in non-resistant hosts, which is expected and often observed for such mutations (KRAAIJEVELD AND GODFRAY 1997; KOSKELLA 2018). In resistant hosts, the escape mutant will, by definition, have a higher reproduction ratio than the wild-type pathogen ($R_{RE} > R_{RW}$). However, since the two pathogens are closely related and the wild-type pathogen is adapted to the non-resistant host, it is probably unlikely that R_{RE} will exceed the reproduction ratio of the wild-type pathogen in non-resistant hosts ($R_{RE} \leq R_{NW}$).

We modelled pathogen characteristics via the transmission rate parameter (β), assuming a recovery rate (α) of 1, which allowed us to present the results in terms of reproduction ratios. Since most livestock diseases of interest to animal breeding have a low mortality, we assumed that the pathogen does not kill the host. Thus, the infectious period does not end with death of the host, but with its recovery. A brief investigation showed that the reproduction ratios fully capture variation in both the transmission and recovery rate parameters in our SIS-model (results not shown). Thus, for the possibility of a mutant to invade, it does not matter whether a lower reproduction ratio in resistant hosts is caused by reduced transmissibility of the pathogen (lower β) compared to in non-resistant hosts, by a greater recovery rate of the host (higher α), or by a combination of both. This result implies that the assumption of $\alpha = 1$ does not restrict the generality of our results, but merely served to simplify the mathematics. Furthermore, our results are not restricted to situations in which a trade-off between recovery and transmissibility exists; given the reproduction ratios used in our model, results do not depend on the nature of such a trade-off, although evolution of pathogen virulence may depend on this trade-off (see below). With incomplete cross-resistance, the generality of our results with respect to variation in the transmission and recovery rate parameters holds if the recovery rate parameter of hosts infected with both

pathogen types is identical to the recovery rate parameter of hosts that are infected with the escape mutant only. Further research is needed to assess whether this condition is realistic.

Another important assumption underlying our results is that there is no continuous import of infectious material or infectious animals from outside the herd. If infectious material frequently enters from outside, minor outbreaks of the wild-type pathogen might still occur in a herd that is genetically selected for higher resistance. Here, a minor outbreak refers to the introduction of an infection in a population leading to only a few infected individuals, after which the infection dies out (DE JONG 1995). Escape mutants may then also emerge and invade from these minor outbreaks. The assumption of a closed or semi-closed local population holds or can hold for many livestock diseases, depending on farm management. Even if pathogens are transmitted through the environment, the population might still be closed, as long as the infectious material is excreted in the environment by animals from the same farm. This is for instance most likely the case for some important pathogens that cause mastitis and claw infection digital dermatitis (KLAAS AND ZADOKS 2018; ORSEL *et al.* 2018), but not for bovine tuberculosis, where badgers or other wildlife species are an important environmental infection source (BOEHM *et al.* 2009).

5.3.2 Pathogen evolution and eradication

We deliberately ignored the evolutionary dynamics in the pathogen population, which have been studied extensively (e.g. MAY AND ANDERSON 1990; FRANK 1991; DUSHOFF AND LEVIN 1995; VAN BAALEN AND SABELIS 1995; REGOES *et al.* 2000; GANUSOV *et al.* 2002; YATES *et al.* 2006; CHABAS *et al.* 2018). In these studies, interest is usually in host mortality, or virulence, of the pathogen at its evolutionary optimum, and results show that whether or not a mutant strain will emerge and invade strongly depends on the model assumptions. Our interest, however, was to identify the range of host population resistance within which an escape mutant can invade. This different interest mainly follows from the human control on genetic selection, animal movement, and isolation or removal of infectious individuals in livestock populations, as opposed to human or natural populations. Moreover, the more closed and localised structure of livestock populations makes eradication of an infection, especially locally, much more feasible in livestock than

in human populations, which is illustrated by the disease-free status of many countries and regions for animal infections such as the Pseudorabies virus (Aujeszky's disease) and bovine tuberculosis (EUROPEAN COMMISSION 2020). Our recent results show that such eradication by genetic selection is much more promising than commonly believed (HULST *et al.* 2021; BIJMA *et al.* 2022). Hence, similar to the use of vaccination, genetic selection in livestock can, at least in principle, be used to eradicate a pathogen from the (local) population. When the objective is eradication, interest is not in the evolution of pathogens to an optimum, but in avoiding the invasion of escape mutants before (local) eradication has been achieved.

Our model with two types of hosts is essentially comparable to models that are used to investigate the evolutionary consequences of vaccination on the pathogen. Such models have received much more attention in the literature than models of the evolutionary consequences of artificial genetic selection of the host population. GANDON AND DAY (2007) mention two main approaches that are taken in such evolutionary vaccination studies: (1) the modelling of the invasion of escape strains of the pathogen, similar to our approach, and (2) the modelling of evolutionary changes in the virulence of the pathogen. We will discuss some examples of both approaches below and point at important similarities and differences with our study.

Concerning the first approach, GANDON AND DAY (2007) show expressions for the critical vaccination coverage that is required to eradicate an infection, analogous to our $f_{R_{max}}$ (Equation 5.9), and for the reproduction ratio of invasion of an escape mutant, analogous to our R_{INV} (Equation 5.11). MAN *et al.* (2021) conclude that replacement of the wild type human papillomavirus by a mutant strain depends on the trade-off between the amount of cross-immunity (our parameter r) and cross-protection of the vaccine against the mutant strain (i.e. the ratio $\frac{R_{RE}}{R_{RW}}$ in our model), which closely corresponds to our findings.

As discussed above, an important difference between our study and many other studies on pathogen escape is that we focus on eradication of the infection. In the model of MCLEAN (1995), who investigated why escape mutants did not arise in several vaccination campaigns, the vaccine-favoured strain (i.e. the escape

mutant) takes over sooner or later, such that the infection is never eradicated. Similarly, MAGORI AND PARK (2014) investigated how different types of imperfect vaccines influence the rate of spread of mutant strains with a fixed vaccination coverage of 0.25. In their model, the infection was not eradicated, because the reproduction ratio of the wild-type pathogen was still above 1 at their assumed vaccination coverage. The main reason why these vaccination studies have not considered eradication is probably because they focus on human infections, where (local) eradication is much less feasible than in livestock (see above).

The second approach that has been used to model evolutionary changes in pathogen virulence, differs more from the approached used in our study. The main interest in that approach is in pathogens that may evolve to become more harmful after vaccination of the host population. A prominent example from livestock of such a change is the evolution of Marek's disease virus, a highly contagious virus of poultry. Vaccination of poultry flocks with a vaccine that did not stop transmission of the virus resulted in a new variant of the virus that was more harmful to both vaccinated and unvaccinated host (READ *et al.* 2015). An example from the human population is the evolution of the bacterium that causes whooping cough (pertussis) following vaccination with a vaccine that reduces disease symptoms but does not prevent transmission of the bacterium (MILLER AND METCALF 2019). In both these cases, the pathogen became more harmful after the vaccination campaign, which relates back to the virulence-transmission trade-off, which is extensively discussed in the evolutionary literature (see ALIZON *et al.* 2009 for a review). When vaccination reduces clinical symptoms but does not stop transmission, more transmissible strains of the pathogen can evolve, without paying the cost of killing the host rapidly. However, when the aim of the disease control strategy is to eradicate infection, as is the focus in our study, difficulties associated with increased virulence of the pathogen are prevented. Nevertheless, the above examples stress the importance of further research into the sustainability of selection for tolerance or resilience (DOESCHL-WILSON *et al.* 2012; BERGHOF *et al.* 2019b; KNAP AND DOESCHL-WILSON 2020; BAI AND PLASTOW 2022), where the goal is to reduce the impact of disease on production and not necessarily to stop transmission.

5.3.3 Implications for animal breeding

Our results show that a window of invasion for escape mutants can be identified for the frequency of resistant hosts in a local population. Typically, this invasion window does not cover the entire range of the frequency of resistant hosts. Thus, an implication of the results (e.g. Figure 5.4) is that, in a closed population, the frequency of animals with the resistance gene should be increased as rapidly as possible through the window, *i.e.,* from the lower bound where $R_{INV} > 1$ to the upper bound of eradicating the wild-type pathogen ($f_{R_{max}}$). This allows the infection to be eradicated while minimizing the risk of invasion of escape mutants that nullify the effects of selection. Note that an increase of resistant hosts can also lead to a considerable decrease in prevalence, even when the infection is not totally eradicated. Although this seems desirable at first sight, because the number of individuals affected by the infection is largely reduced, it introduces the risk of invasion of escape mutants and this risk remains in the long term.

Here, a rapid increase in the frequency of resistant hosts refers to a rate of increase that is faster than the rate at which escape mutants can develop in the pathogen population. Consequently, current rates of genetic improvement of host resistance may be fast enough when the target infection concerns a slowly evolving macroparasite. For example, selection for a single resistance gene of sheep against gastro-intestinal nematode infection, might not cause adaptation of the nematodes for up to 20 years (KEMPER *et al.* 2013). However, for rapidly evolving microparasites, the current speed of genetic improvement of host resistance will likely not be fast enough. The strategies that we discuss in the following section to bridge the invasion window may therefore be most relevant to microparasitic infections.

One can think of different strategies to bridge the invasion window. Local livestock populations that are relatively small and largely isolated from other such populations are most suitable for such strategies, because of the possibility to control animal movement and contacts between farms. A possibility to achieve a high frequency of resistant hosts in a local population could be to base herd composition on the resistance of the individual animals, such that herds are either resistant enough to eradicate the infection (*i.e.*, the frequency of the resistant hosts is greater than $f_{R_{max}}$) or that the frequency of resistant hosts stays below

the point above which the escape mutant can invade (i.e. R_{INV} stays less than1). This is probably most feasible in cases where the breeding population is separated from the production population, and where all animals in production herds are replaced at once, such as in poultry and sometimes in fattening pigs, as opposed to dairy cattle for example. Although the increase in resistance will necessarily be more gradual in the nucleus selection lines of the former species, the risk of escape mutants emerging seems limited since the selection lines are usually free from clinically important infectious diseases (specific pathogen-free).

When replacement occurs gradually, as in cattle, preventing escape mutants from invading will be much more challenging. If feasible, one could consider separating resistant from non-resistant animals, as in the previous strategy, such that, in fact, two herds are created; one in which the infection with the wild-type pathogen is still endemic but escape mutants cannot invade because the level of resistance is too low, and one in which the wild-type pathogen is eradicated because of the high level of resistance.

Another option would be to increase (local) population resistance rapidly by putting high selection pressure on the resistance trait or by sorting the genetic material, *e.g.,* by using only the most resistant sires on a certain farm. A complicating factor when breeding an endemically infected population for eradication of the pathogen, is that the infection can still be present for some time after the population reaches the point when eradication is expected, before it dies out (BISHOP AND MACKENZIE 2003). During that period, a risk of invasion of escape mutants continues because there is substantial selection pressure on the pathogen population to escape the high frequency of resistant hosts. Finding and removing or isolating infected individuals could then be a way to accelerate eradication. A similar approach is taken when eradicating infections by vaccination (e.g. HENDERSON 1976; HENDERSON 2011; MARINER *et al.* 2012; ROEDER *et al.* 2013). Note that the ongoing infection as discussed above is not visible in our deterministic model. Similarly, but with a more favourable outcome, the infection might, by chance, die out before the point where eradication is expected.

To reach eradication as soon as possible, it is key that R_{RW} is as low as possible (see Figure 5.3). Although it might be difficult to influence the size of the effect of resistance genes, it is important that the effect is at least large enough to decrease

R_{RW} below 1. Thus, when for example multiple QTL for host resistance are identified, selection should preferably increase their frequency simultaneously, instead of one by one. It might also help to include multiple traits in the breeding goal that are focused on reducing infection prevalence, i.e., infectivity and recovery, in addition to only resistance, as we modelled here. Moreover, other interventions against the infection can assist in decreasing R_{RW} rapidly to a value less than 1, such as removal of infected animals, vaccination, or general hygienic measures. Thus, a combination of genetic selection for infectious disease resistance and other interventions will be most effective, both in eradicating the wild-type pathogen and in preventing escape mutant invasion.

All strategies described above are also relevant to gene-edited resistance, including the use of other interventions to assist in decreasing the reproduction ratio to a value below 1 (PETERSEN *et al.* 2022). Resistance to the porcine reproductive and respiratory syndrome virus is a clear example where gene-editing might be an interesting strategy for infection control (PETERSEN *et al.* 2022), although it is questionable whether many more of such cases exist.

As discussed by BISHOP AND MACKENZIE (2003) , it is probably more difficult for pathogens to find escape strategies for polygenic disease resistance traits than for monogenic traits, especially when escape from polygenic resistance requires multiple mutations in the pathogen. This expectation is supported by the results of a selection experiment of KEMPER *et al.* (2009) in which nematodes showed no adaptation to long-term exposure to polygenic resistant sheep. This phenomenon is more widely observed and exploited in plant breeding, where sustainable strategies for disease resistance breeding consist of combining multiple resistance genes in a cultivar, preferably with different mechanisms of action (MUNDT 2018). Combining multiple genes with different mechanisms of resistance may again be most feasible for species with a separate breeding population. It requires that the breeding population is not exposed to the pathogen before all individuals have at least two genes with different mechanisms of resistance, otherwise pathogens would still be able to evolve resistance to each mechanism one-by-one. However, our quantitative predictions are not easily translated to host resistance based on multiple mechanisms, since this would require a considerable expansion of the

model to account for the different mechanisms of resistance and pathogen escape strategies.

In livestock genetic improvement, selection for an index consisting of estimated breeding values for many traits has become the standard (HAZEL *et al.* 1994). For ordinary quantitative genetic traits, such multi-trait index selection (HAZEL 1943) is superior to selection for several traits, one trait at a time, for example over generations (tandem selection,) or within generation using independent selection thresholds for each trait (independent culling levels) (YOUNG 1961). Multi-trait selection typically creates small steps of improvement for each trait. However, when the breeding goal includes resistance to an infectious disease, gradual improvement by multi-trait selection will cause the host population to remain in the invasion window for a long time, such that emerging escape mutants might have ample opportunity to invade. Thus, our results give at least some reason to think that multi-trait selection might be suboptimal when infectious disease resistance is part of the breeding goal. The best documented and undisputed case where long-term weak selection against pathogens has resulted in the evolution of escape mutants, is the long-term use of low doses of antibiotics. That practice has strongly supported the evolution of antibiotic-resistant bacteria, which poses a threat to the availability of effective antibiotics in the long term (KHACHATOURIANS 1998; OLOFSSON AND CARS 2007; SANDEGREN 2014). In a sustainable strategy, resistance is strong enough to bring the reproduction ratio of the wild-type pathogen below 1 ($R_{RW} < 1$). This resembles the use of high doses of antibiotics, which is known to be effective in preventing the evolution of antibiotic resistant strains (OLOFSSON AND CARS 2007). For animal breeding, this would correspond to applying the total genetic selection differential for infectious disease resistance within a single or a few generations, at least locally, instead of accumulating it gradually over many generations of multi-trait selection. Although the similarity to the use of low doses of antibiotics clearly questions the sustainability of multi-trait selection when resistance is part of the breeding goal, further (experimental) research is definitely needed here.

5.4 Conclusions

Here, we showed that genetic selection for infectious disease resistance determined by a single locus in a closed livestock population typically provides an opportunity for escape mutants of pathogens to invade the host population. Given the reproduction ratios of the infectious pathogens, we identified the range for the frequency of resistant hosts in the population within which there is this risk of invasion. As long as there is no import of infectious material from outside the closed population, this 'invasion window' extends from the frequency of resistant hosts at which the reproduction ratio of the escape mutant in a host population that is endemically infected with the wild-type pathogen becomes greater than 1, until the frequency of resistant hosts above which the wild-type pathogen is expected to die out. To prevent invasion of escape mutants in closed populations, the frequency of resistant hosts should thus be increased faster through this window than escape mutants can develop in the pathogen population. The current, very gradual, multi-trait selection approach in animal breeding might thereby pose a risk of invasion of escape mutants. A possible strategy to prevent the invasion of escape mutants is to place animals into herds based on their genetic resistance, such that herds stay out of the invasion window on either side. Combining such a sustainable selection strategy with other interventions that reduce transmission of the pathogen will help to prevent invasion of escape mutants and to eradicate an infection.

5.5 Acknowledgements

The authors thank Eline Vink for helping with the design of the multi-type transmission model and Andrea Doeschl-Wilson and Jack Dekkers for their comments that led to considerable improvement of the manuscript.

5.6 Appendix

5.6.1 Model equations

$$\frac{dS_N}{dt} = (I_{NW} + I_{NE} + I_{NWE}) - R_{NW}S_N\frac{(I_{NW} + I_{RW})}{N} - R_{NE}S_N\frac{(I_{NE} + I_{RE})}{N}$$

$$\frac{dS_R}{dt} = (I_{RW} + I_{RE} + I_{RWE}) - R_{RW}S_R\frac{(I_{NW} + I_{RW})}{N} - R_{RE}S_R\frac{(I_{NE} + I_{RE})}{N}$$

$$\frac{dI_{NE}}{dt} = R_{NE}S_N\frac{(I_{NE} + I_{RE})}{N} - (1-r)R_{NW}I_{NE}\frac{(I_{NW} + I_{RW})}{N} - I_{NE}$$

$$\frac{dI_{NW}}{dt} = R_{NW}S_N\frac{(I_{NW} + I_{RW})}{N} - (1-r)R_{NE}I_{NW}\frac{(I_{NE} + I_{RE})}{N} - I_{NW}$$

$$\frac{dI_{NWE}}{dt} = (1-r)R_{NW}I_{NE}\frac{(I_{NW} + I_{RW})}{N} + (1-r)R_{NE}I_{NW}\frac{(I_{NE} + I_{RE})}{N} - I_{NWE}$$

$$\frac{dI_{RE}}{dt} = R_{RE}S_R\frac{(I_{NE} + I_{RE})}{N} - (1-r)R_{RW}I_{RE}\frac{(I_{NW} + I_{RW})}{N} - I_{RE}$$

$$\frac{dI_{RW}}{dt} = R_{RW}S_R\frac{(I_{NW} + I_{RW})}{N} - (1-r)R_{RE}I_{RW}\frac{(I_{NE} + I_{RE})}{N} - I_{RW}$$

$$\frac{dI_{RWE}}{dt} = (1-r)R_{RW}I_{RE}\frac{(I_{NW} + I_{RW})}{N} + (1-r)R_{RE}I_{RW}\frac{(I_{NE} + I_{RE})}{N} - I_{RWE}$$

5.6.2 Prevalence of wild-type pathogen in resistant and non-resistant hosts

At equilibrium:

$$R_{NW}S_N^*\frac{I_W^*}{N} = I_{NW}^* \; and \; R_{RW}S_R^*\frac{I_W^*}{N} = I_{RW}^* \tag{5.12}$$

Given that:

$$S_N^* = (1 - f_R)N - I_{NW}^*,$$

$$S_R^* = f_R N - I_{RW}^*,$$

$$\text{and } \frac{I_W^*}{N} = P_W^*,$$

$$R_{NW}P_W^*\left((1 - f_R)N - I_{NW}^*\right) = I_{NW}^* \; and \; R_{RW}P_W^*(f_R N - I_{RW}^*) = I_{RW}^* \tag{5.13}$$

Solving for $\frac{I_{NW}^*}{N} = P_{NW}^*$ and $\frac{I_{RW}^*}{N} = P_{RW}^*$ yields:

$$P_{NW}^* = \frac{R_{NW}P_W^*\,(1 - f_R)}{1 + R_{NW}P_W^*} \tag{5.14}$$

$$P_{RW}^* = \frac{R_{RW}P_W^* f_R}{1 + R_{RW}P_W^*} \tag{5.15}$$

Where $P^* = P^*_{NW} + P^*_{RW}$. We can solve these two equations with two unknowns to get expressions for the prevalence as function of the reproduction ratios and f_R:

$$P^*_{NW} = \frac{(1-f_R)\left(-R_{RW}+R_{NW}\left(-1+R_{RW}+R_{RW}\sqrt{\frac{1}{R_{NW}^2}+\frac{-2+4f_R-\frac{2}{R_{RW}}}{R_{NW}}+\frac{1-2R_{RW}-4f_RR_{RW}+R_{RW}^2}{R_{RW}^2}}\right)\right)}{R_{RW}+R_{NW}\left(-1+R_{RW}+R_{RW}\sqrt{\frac{1}{R_{NW}^2}+\frac{-2+4f_R-\frac{2}{R_{RW}}}{R_{NW}}+\frac{1-2R_{RW}-4f_RR_{RW}+R_{RW}^2}{R_{RW}^2}}\right)} \tag{5.16}$$

$$P^*_{RW} = \frac{f_R\left(-R_{RW}+R_{NW}\left(-1+R_{RW}+R_{RW}\sqrt{\frac{1}{R_{NW}^2}+\frac{-2+4f_R-\frac{2}{R_{RW}}}{R_{NW}}+\frac{1-2R_{RW}-4f_RR_{RW}+R_{RW}^2}{R_{RW}^2}}\right)\right)}{-R_{RW}+R_{NW}\left(1+R_{RW}+R_{RW}\sqrt{\frac{1}{R_{NW}^2}+\frac{-2+4f_R-\frac{2}{R_{RW}}}{R_{NW}}+\frac{1-2R_{RW}-4f_RR_{RW}+R_{RW}^2}{R_{RW}^2}}\right)} \tag{5.17}$$

$$P^*_W = \frac{1}{2}\left(-\frac{1}{R_{NW}}+\frac{-1+R_{RW}+R_{RW}\sqrt{\frac{1}{R_{NW}^2}+\frac{-2+4f_R-\frac{2}{R_{RW}}}{R_{NW}}+\frac{1-2R_{RW}-4f_RR_{RW}+R_{RW}^2}{R_{RW}^2}}}{R_{RW}}\right) \tag{5.18}$$

Chapter 6

Estimation of genetic variance and breeding values for infectious disease susceptibility from longitudinal data using generalized linear mixed models

Andries D. Hulst[1,2], R. Pong-Wong[3], A. Doeschl-Wilson[3], Piter Bijma[2], Mart C.M. De Jong[1]

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

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

[3]The Roslin Institute, University of Edinburgh, Midlothian EH25 9RG, UK.

Abstract

Recent theoretical work shows that the potential of genetic selection to reduce the prevalence of infectious diseases is much larger than expected, due to indirect genetic effects that arise in the transmission process. However, to fully benefit from these indirect effects, we need to estimate genetic parameters and breeding values, which requires statistical methods tailored to the transmission process. Here, we evaluate Generalised-Linear-Mixed Models (GLMMs) to estimate genetic parameters and breeding values for host susceptibility from simulated epidemics. Applying GLMMs to longitudinal records of individuals' infection status, the impact of sampling interval, population structure, infection characteristics and model formulation was assessed. The results show that a GLMM fitted to longitudinal records on individual binary infection status can produce accurate and unbiased estimates of genetic variance for susceptibility to an infectious disease. Of the data requirements, the time interval between consecutive observations on individual infection status was the main factor affecting estimation, while group size had limited effect. The required observation interval depends on the infection and recovery rates of individuals. The GLMM thus seems an accurate and relatively easy implementable model to estimate genetic parameters for susceptibility, hence it might considerably improve the potential of genetic selection to reduce the impact of infectious diseases in livestock.

6.1 Introduction

Genetic selection of animals against infectious diseases, as an addition or alternative to other control measures, has attracted interest for several decades (e.g. BISHOP *et al.* 2010). The common approach in animal breeding is to measure infectious disease resistance as the binary infection status of the animal (healthy/infected = 0/1), and analyse these data either with a regular linear model or a so-called threshold model (e.g. PÉREZ-CABAL *et al.* 2009; WELDERUFAEL *et al.* 2017). While linear models treat the binary observation as a continuous Normally-distributed variable, threshold models use a GLMM-like approach to link the binary observation to an underlying normally distributed trait (DEMPSTER AND LERNER 1950; GIANOLA 1982; DE VILLEMEREUIL *et al.* 2016). Importantly, both approaches do not incorporate the transmission dynamics of infectious diseases.

Recent theoretical work, however, shows that these transmission dynamics cause indirect genetic effects, resulting in a much larger potential of selection for lower infectious disease prevalence than expected from classical quantitative genetic models (HULST *et al.* 2021; BIJMA *et al.* 2022). These models are expressed in terms of susceptibility, rather than disease resistance. However, susceptibility is merely the opposite of disease resistance; individuals with a high susceptibility have a low resistance, and vice versa. The expectation from classical quantitative genetic models is that genetic reduction of infection prevalence becomes impossible at lower prevalence, because additive genetic variance on the observed binary scale diminishes when the prevalence of the binary trait goes to zero (ROBERTSON 1950). In full contrast, the indirect genetic effects on infection prevalence that arise because of transmission make it in principle possible to eradicate an infection using genetic selection (HULST *et al.* 2021; BIJMA *et al.* 2022).

Since one of the host traits affecting the transmission process is susceptibility to infection (ANCHE *et al.* 2014; ANACLETO *et al.* 2015), accurate estimates of genetic variation and breeding values for susceptibility are needed for correct prediction of response to selection in infection prevalence. A statistical model tailored to the transmission process is needed, because the disease status of an individual animal also depends on its exposure to infectious material, e.g. from infected herd mates (HULST *et al.* 2021). An animal with any particular susceptibility will be, on average, more likely to become infected itself when it is in a herd with highly

susceptible animals, and less likely when it is in a herd with lowly susceptible animals, which would result in different estimates for the breeding value of the individual. The discrepancy is caused by a difference in exposure between the two herds; the prevalence of the infection, and thus exposure to infectious material, will be higher in a herd with more susceptible animals than in a herd with less susceptible animals, just because the highly susceptible individuals are more likely to be infected and thereby also more likely to infect others. Thus, for accurate estimation of genetic variance and breeding values for susceptibility, we need a method that accounts for the exposure of each individual to infectious herd mates.

Generalised-Linear-Mixed-Models (GLMMs) that are founded in epidemiological theory and were developed to estimate transmission parameters, have been shown to be able to estimate genetic parameters and individual genetic effects for susceptibility (VELTHUIS *et al.* 2003; ANCHE *et al.* 2015; BIEMANS *et al.* 2019a). Results of BIEMANS *et al.* (2019a) suggest that the GLMM can relatively accurately estimate breeding values for susceptibility, even with a limited amount of data. A benefit of the GLMM is that it is available in standard animal-breeding software (i.e. ASREML (GILMOUR *et al.* 2015)) and requires relatively limited computation time, which is relevant for practical implementation. However, the performance of the GLMM for estimation of genetic parameters for susceptibility has never been properly validated.

In this study, we therefore evaluate how precision of variance estimates and prediction accuracies for susceptibility, estimated with a GLMM, depend on the genetic variance in susceptibility, population structure, available data, and model formulation. For the evaluation, we use longitudinal data from simulated epidemics and the GLMM implemented into ASREML software. Longitudinal data is required because a key feature of the GLMM is that it accounts for exposure of susceptible individuals to infectious material and the correctness of this estimate of exposure depends on the frequency of observing it. The effect of size of the genetic variance in susceptibility, and transmissibility of the infection (R_0) on the precision of genetic parameters and breeding value accuracy for susceptibility will be investigated. With respect to data requirements, focus lies on observation interval and contact-group size. The results provide relevant insights in how different types of disease data and trial designs impact on the precision of genetic

variance estimates and accuracy of estimated breeding values for host susceptibility.

6.2 Methods

To be able to evaluate the genetic parameter estimates for host susceptibility obtained from GLMMs, we simulated epidemics of an infectious disease using a Susceptible-Infectious-Recovered (SIR) epidemiological model. In an SIR-model, animals can be in one of three states; Susceptible (i.e., they can become infected), Infected, or Recovered. Infected animals are assumed to be immediately infectious, and recovered animals represent animals that are immune and can thus not be infected again (Figure 6.1). We simulated a population consisting of contact groups within which the infection can be transmitted. In each contact group, four random individuals were infected to start the epidemic. Infection and recovery times of all individuals were recorded. From these known infection and recovery times, data on individual disease status (susceptible, infected or recovered) was sampled using a certain sampling interval. This mimics a real-life setup, where exact infection and recovery times are usually unknown, but individual infection status can be sampled at given intervals. The resulting longitudinal data on individuals' infection status was then used as input for estimation of genetic parameters and breeding values. To assess the performance of the GLMM and data requirements, we simulated different scenarios with varying genetic variance in susceptibility, different average transmissibility of the infection, group size, and sampling interval. Furthermore, we investigated the effect of including different model terms (e.g., a fixed group effect to account for differences in exposure between groups) in the GLMM.

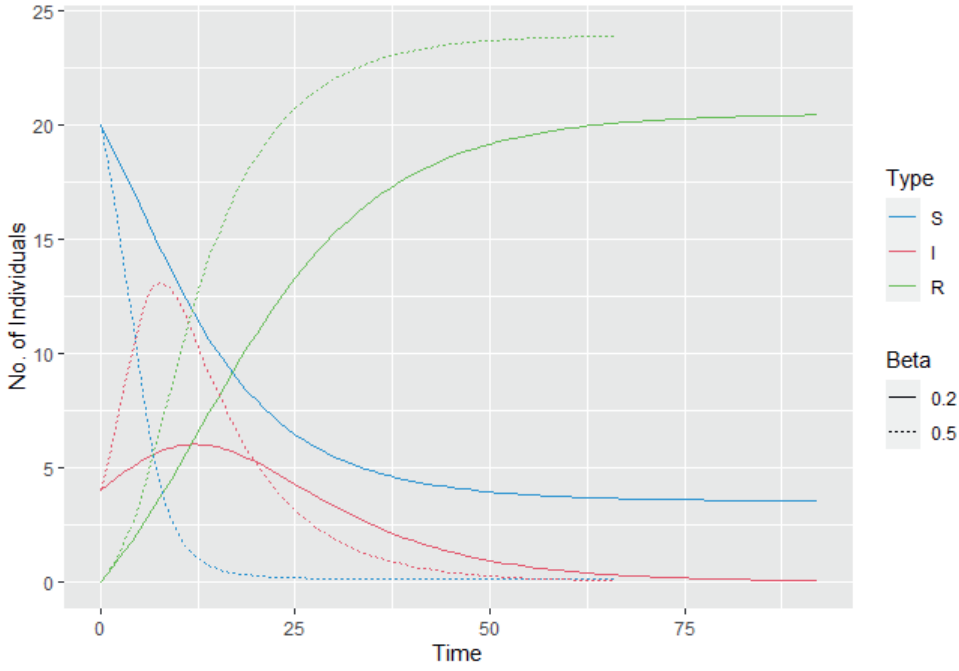


Figure 6.1 Typical epidemic curves for a contact group of 24 individuals with transmission rate parameter (β) of 0.2 and 0.5, and recovery rate parameter (α) of 0.1 (such that R_0 is 2 and 5). Dying out of the epidemic is represented by stopping of the line.

6.2.2 Data simulation using an epidemiological model

Epidemics were simulated using a stochastic SIR model in R (R CORE TEAM 2020), assuming that individuals only differed in their susceptibility to infection. A Doob-Gillespie algorithm was used to simulate the epidemics as a series of stochastic events that follow a Poisson process, as described e.g. in (POOLEY *et al.* 2020). Briefly, the Doob-Gillespie algorithm represents stochasticity in Markovian compartmental (e.g., SIR) models (i.e., models for which the transition rates depend solely on the current state of the system).

Two events can happen in the SIR-model: infection and recovery. For an individual i, infection, i.e., the transition from the susceptible (S) to the infectious (I) state happens at a rate

$$S_i \rightarrow I: \beta \gamma_i \frac{I}{N} \tag{6.1}$$

where β denotes the average transmission rate parameter, γ_i the susceptibility of individual *i*, and $\frac{I}{N}$ the proportion of infectious individuals in the contact group.

The second event, recovery, the transition of an individual from infectious (I) to recovered (R), happens at an average rate

$$I \rightarrow R: \alpha \tag{6.2}$$

where α denotes the recovery rate parameter, which was gamma-distributed with shape-parameter 3 and rate 0.3, such that the average recovery rate was 0.1, resulting in an average duration of the infectious period of $\frac{1}{\alpha}$ = 10 days. A gamma-distributed recovery rate was preferred over the more conventional exponential distribution, because it results in a more biologically realistic distribution of the infectious period (POOLEY *et al.* 2020).

Individuals' susceptibility (γ) was simulated in the same way as in HULST *et al.* (2021) , i.e., assuming and infinitesimal genetic model, and following a lognormal distribution,

$$\gamma_i = e^{A_{\gamma_i} + E_{\gamma_i}} \tag{6.3}$$

where A_{γ_i} denotes the additive genetic part of individual susceptibility, and E_{γ_i} the permanent environmental part of individual susceptibility. Both A_γ and E_γ were normally distributed, with mean zero and variance $h^2_{\log(\gamma)}\sigma^2_{P_{\log(\gamma)}}$ for the genetic part, and $(1 - h^2_{\log(\gamma)})\sigma^2_{P_{\log(\gamma)}}$ for the permanent environmental part. Here $\sigma^2_{P_{\log(\gamma)}}$ denotes the total "phenotypic" variance in the logarithm of susceptibility, and $h^2_{\log(\gamma)}$ the heritability of the logarithm of susceptibility (i.e., on the underlying normal scale). We refer to $\sigma^2_{P_{\log(\gamma)}}$ as a phenotypic variance, because it is the sum of a genetic and environmental part, even though phenotypic values for susceptibility are not directly observable. Thus, A_{γ_i} is the breeding value, as common in quantitative

genetics, for the logarithm of susceptibility, and γ is lognormally distributed. One of the advantages of the lognormal distribution of γ is that the transmission rate will never become negative (which is not possible).

Furthermore, for convenience, we want the mean of γ to be 1, such that our parameter β in Equation 6.1 is the actual mean transmission rate parameter. However, the mean of the lognormal distribution is $e^{\mu+\frac{1}{2}\sigma^2}$, and is thus greater than 1 for any variance larger than 0. For small variance the deviation of 1 is small, but for large variance the deviation becomes quite problematic for the comparison of scenarios (e.g., for $\sigma^2_{P_\gamma} = 1$ the mean of the lognormal is 1.65). Thus, for each simulation we divided β by the mean of the lognormal distribution, given the simulated phenotypic variance in the logarithm of γ, such that the mean transmission rate (β) was corrected for the effect of simulated variance on the mean level of transmission.

6.2.3 Population structure

The population was simulated with a full-sib half-sib structure, where 96 unrelated sires were mated to 5 unrelated dams each, producing 10 offspring per mating, resulting in a total offspring population of 4800 animals. Offspring breeding values for susceptibility (A_{γ_i}) were simulated according to the infinitesimal model (FISHER 1918) (Equation 6.3). These offspring were randomly allocated to groups of size *N*. Epidemics were simulated within groups, starting with infection of four randomly selected seeder individuals, and no transmission between groups. At fixed timepoints with interval Δt, the infection status of each individual was recorded. Since the epidemics in the contact groups had a limited time span because at some point there are no infected individuals left (Figure 6.1), the number of time points recorded was not the same in all scenarios but depended on the duration of the epidemic (*T*) and the sampling interval (Δt). Thus, the total number of observed timepoints per scenario was $\frac{T}{\Delta t}$. Hence, scenarios with smaller Δt had more observations than scenarios with larger Δt.

6.2.4 Statistical analysis with GLMM

We used a Generalised-Linear-Mixed-Model tailored to infectious disease data, following BIEMANS *et al.* (2019a), VELTHUIS *et al.* (2003), LIPSCHUTZ-POWELL *et al.* (2014), and ANCHE *et al.* (2015). The GLMM models the probability of a susceptible individual, *i*, to become infected in a certain time interval,

$$P_i(t) = 1 - e^{-\beta\gamma_i\frac{I}{N}\Delta t} \tag{6.4}$$

where $P_i(t)$ is the probability that susceptible individual *i* becomes infected in interval Δt, and $e^{-\beta\gamma_i\frac{I}{N}\Delta t}$ is the zero term of the Poisson distribution, representing the probability for *i* to escape infection. The term in the exponent is the negative of the mean number of infections of *i* during Δt, which is the product of the average transmission rate parameter β, the susceptibility γ_i of *i*, the number of infected herd mates at the start of the interval $\frac{I}{N}$, and the length of the interval (Δt). Whether or not an individual becomes infected in an interval, *y* = 0/1, follows a binomial distribution with binomial total 1 and probability $P_i(t)$; $y_i \sim Bin(1, P_i(t))$. For binary infection data resulting from a Poisson process, as simulated above, the appropriate link function to connect the expected value of *y* to the explanatory variables is the complementary log-log (cloglog) (VELTHUIS *et al.* 2003). A GLMM tailored to longitudinal records of individual binary infection status, therefore, follows from applying the cloglog transformation to Equation 6.3, giving:

$$cloglog\big(P_i(t)\big) = \log(\beta) + \log(\gamma_i) + \log\left(\frac{I}{N}\Delta t\right) \tag{6.5}$$

where $P_i(t)$ is the expected probability of infection of susceptible individual i, $\log(\beta)$ is the intercept, $\log(\gamma_i)$ the logarithm of the susceptibility of individual *i*, and $\log\left(\frac{I}{N}\Delta t\right)$ an offset, i.e., an explanatory variable with a known value. The offset accounts for the known exposure of individual *i* to infected herd mates, which depends on the fraction of herd mates that are infected, I/N, and the duration of the sampling interval (Δt). Thus, the input data for the GLMM consists of, for each individual that is susceptible at the beginning of an interval, whether that susceptible individual becomes infected in that interval (whether it is a case), and the fraction of infected herd mates ($\frac{I}{N}$) it was exposed to in the previous interval.

To determine the effect of different model terms on the estimation of genetic variance and breeding values for susceptibility, we fitted four different GLMMs to the simulated data of individual binary infection status recorded at time intervals Δt, in ASREML (GILMOUR *et al.* 2015). Table 6.1 gives an overview of the models. Model 1 can be considered as the 'full model':

$$cloglog\left(P_{ikt}(t)\right) = \mathrm{c}_0 + \mathrm{Group}_\mathrm{k} + \mathrm{Time}_\mathrm{T} + Group_k.Time_T + A_{\ln_\gamma,i} + E_{P\ln_\gamma,i} + \log\left(\frac{\mathrm{I}}{\mathrm{N}}\Delta\mathrm{t}\right) \qquad (model\ 1)$$

Where c_0 is the intercept, which will give the estimate for $\log(\beta)$, $\mathrm{Group}_\mathrm{k}$ and Time_T are fixed effects for group and observation period, $Group_k.Time_T$ is a random interaction term between group and observation period, $A_{\ln_\gamma,i}$ is the random genetic effect for log-susceptibility with underlying pedigree relationship matrix, $E_{P\ln_\gamma,i}$ the random permanent effect for susceptibility, and $\log\left(\frac{\mathrm{I}}{\mathrm{N}}\Delta\mathrm{t}\right)$ an offset, as in Equation 6.4. Here, the logarithm of susceptibility $\log(\gamma)$ (Equation 6.4) is separated into a random genetic and a random permanent environmental effect, with $\log(\gamma) \sim \mathrm{N}(0, \mathbf{A}\sigma^2_{\mathrm{A}_{\log(\gamma)}} + \sigma^2_{\mathrm{E}_{\log(\gamma)}})$. Where **A** is the pedigree relationship matrix among animals.

Note that although we did not simulate systematic group and time effects, we fit a fixed group effect in model 1 and 2 and a random interaction effect between group and time in models 1 to 3. The motivation for this is mainly found in BIEMANS *et al.* (2019a), where a significant group.time variance was estimated, even with fitted fixed group and time effects (similar to our model 1). Furthermore, in empirical data, it is beforehand not known whether differences between group and time intervals account for a substantial part of the observed variation, thus it seems reasonable to evaluate the effect of fitting a term for period and group in the model. Next to the genetic term for susceptibility, all models contained a permanent individual random effect, because we repeatedly observed the infection status of each individual, such that each interval yielded an observation for each susceptible individual with corresponding offset.

Table 6.1. Overview of fitted models

Model	Fixed	Random		
1	Group, Time	IDgenetic	IDpermanent	Group*Time
2	Group	IDgenetic	IDpermanent	Group*Time
3		IDgenetic	IDpermanent	Group*Time
4		IDgenetic	IDpermanent	

6.2.5 Investigated scenarios

We simulated different scenarios to investigate the performance of the GLMM in estimation of genetic parameters and breeding values for susceptibility. The varied factors were related to data design, i.e., sampling interval and group size, and biological inputs, i.e., the transmission rate parameter (β), and phenotypic variance and heritability of susceptibility. For each of the variables we defined a default value, and next varied one variable at a time, while the others maintained their default value. Table 6.2 gives an overview of the scenarios.

For the sampling interval the approach was slightly different. There, we tested multiple values in each scenario, to see if the other variables had an effect on the optimum sampling interval. Faster epidemics, for example, represented by a higher value of β, may need a shorter sampling interval to get accurate estimates for susceptibility.

Group size was not expected to have a large effect on accuracy of estimation for susceptibility (ANACLETO *et al.* 2015). However, by simulating different group sizes, we want to provide insights into estimation of susceptibility both from experimental (i.e., small groups) and field data (large groups). An important limitation of small groups might be the risk of stochastic fade-out of the epidemic. This is less likely to occur if the epidemics is kick-started by a large number of seeder individuals. However, for these seeders no information on susceptibility can be recorded, which might negatively affect accuracy of estimation (e.g., with group size 12, one-third of the population is a seeder in our simulations).

Table 6.2. Overview of scenarios

Variable	Default value(s)	Other simulated values
Sampling interval (Δt; days)	1	3, 5, 10, 15, 20, 30
Group size (*N*)	24 (*200 groups*)	12 (*400*), 48 (*100*), 96 (*50*), 240 (*20*), 480 (*10*)
Beta ($\beta;\ day^{-1}$)	0.2	0.12, 0.15, 0.4, 0.5, 1
Duration of infectious period (days)	10	-
Phenotypic variation in susceptibility ($\sigma^2_{P_\gamma}$)	0.25	0.1, 0.5, 1
Heritability of log(γ) $h^2_{\log(\gamma)}$[1]	0.5	0, 1

[1]The simulated values of phenotypic variation and heritability correspond to maximum values of observed-scale heritability (BIJMA *et al.* 2022) between 0.0125 and 0.125, the default scenario to 0.03.

Empirical estimates for heritability and phenotypic variance in susceptibility, as defined above, are sparse. However, in HULST *et al.* (2021) we showed that observed-scale heritability estimates of 0.02 and 0.05 for binary infection status from a linear model, which are common values, correspond to a genetic variation in susceptibility between 0.10 and 0.25. Thus, we simulated a genetic variance of 0.125 in our default scenario, and simulate a range of values around that. Together with the different simulated heritabilities these will give an idea of the effect of genetic and non-genetic variation on the precision of estimates.

Given that we simulated a fixed mean duration of the infectious period of 10 days, variation in infection transmission was simulated through beta only. Our default value of $\beta = 0.2/day$ corresponds to a basic reproduction ratio (R_0) of 2 (reflecting infections where on average ~50% of individuals will ultimately become infected). From BIJMA *et al.* (2022) we know that genetic selection has largest potential for infections with a lower R_0, because indirect effects (dramatically) increase with decreasing R_0. Note that a β below 0.1/day would result in an R_0 below 1, and thus in no outbreak of infection. All our simulated values are therefore above 0.1/day.

6.3 Results

We start this section by comparing the precision and bias of parameter estimates for the different models under the default scenario. Based on this comparison, we choose the best performing model to investigate the effect of the different variables on parameter estimation and accuracy of breeding values. Note that all estimates are the mean of 100 replicates, with corresponding standard errors.

6.3.1 Models

There was a considerable difference in accuracy of estimation of β between the four statistical models (Table 6.3). Model 1 considerably underestimated β, whereas models 3 and 4 yielded unbiased and accurate estimates. This indicates that models 3 and 4 are well able to capture the mean level of transmission, while this is less the case for models 1 and 2. Although the estimate from model 2 was not biased, standard errors on β for models 1 and 2, were relatively large compared to those for model 3 and 4.

For genetic variance of susceptibility, the results were similar (Table 6.3). Models 1 and 2 had slightly larger standard errors than 3 and 4. Furthermore, model 1 considerably overestimated the genetic variance. For permanent environmental variance the pattern was similar: (severe) overestimation in model 1, less so in model 2, while model 3 and 4 resulted in best and most accurate estimates. Including fixed effects for group and particularly for time apparently leads to

biased and inaccurate estimates of variances and β, even though the fixed effects were not statistically significant in both models.

Even though it was not simulated, all models that included a random group.time interaction term gave significant estimates for this variance. This suggests that such variance, which arises in the stochastic transmission processes, accounts for a significant part of the variation in individuals' infection status. With longer sampling intervals the estimated group.time variance increased considerably (Figure 6.2A). A comparison of model 3 and model 4 (Figure 6.1B) shows that, if a group.time interaction is not included in the model (model 4), part of its variance ends up in the estimate for genetic variance in susceptibility, leading to overestimation of this variance with larger sampling intervals. Addition of a group.time random term to the GLMM is thus especially important when the sampling interval is relatively large. However, since the other estimates are not significantly affected by the group.time variance with lower intervals (see the full overlap between model 3 and 4 for interval 1 in Figure 6.1B), adding it to the model seems advisable with any sampling interval.

Based on the above results, we decided to fit a model that includes a random group.time effect, but without fixed group and time effects, which corresponded to model 3. In the remainder of the results we will therefore focus on the estimates from this model.

6.3.2 Sampling interval

Figure 6.3A shows that the genetic variance in susceptibility was slightly overestimated if there was no genetic variance simulated ($h^2_{\log(\gamma)} = 0$). However, this was expected since variance estimates were forced to be non-negative in ASREML. Estimates did not seem affected by sampling interval in the absence of simulated genetic variation. When genetic variance was simulated, only a sampling interval of 1 day resulted in an unbiased estimate. For sampling intervals 3, 5, and 10 the variances were slightly overestimated, while for sampling interval above 15, they were underestimated. The largest overestimation occurred at interval 5, and the largest underestimation at interval 30. Especially for a simulated genetic variance of 0.25 ($h^2 = 1$), the underestimation with sampling intervals of 20 and 30 was considerable. The precision of estimates was quite

similar across sampling intervals, with confidence intervals with a range of 0.08 for $h^2 = 1$, 0.07 for $h^2 = 0.5$, and 0.05 for $h^2 = 0$ (Figure 6.3B).

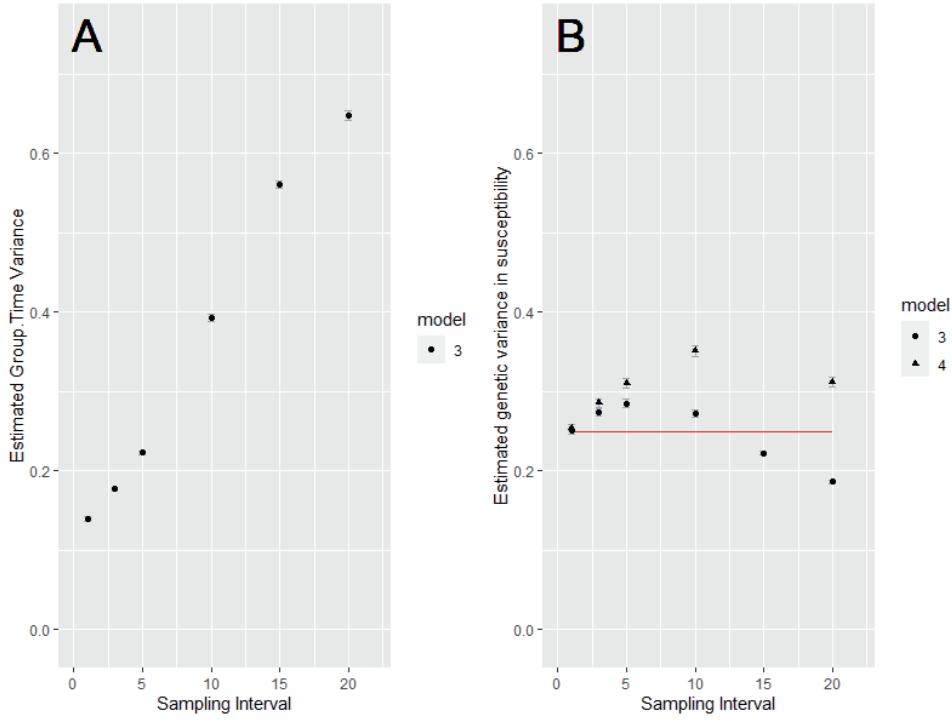


Figure 6.2. A. Group.Time variance for different sampling intervals. **B.** Genetic variance in susceptibility for different sampling intervals, estimates from model 3 and 4. The red line indicates the simulated genetic variance. All other input values as in the default scenario.

6.3.4 Simulated phenotypic variance

Figure 6.4 shows that the GLMM is able to pick up the trend in simulated phenotypic variance and heritability in susceptibility, especially with a sampling interval of 1. With a sampling interval of 10, simulated genetic variances of 0.5 and 1 were significantly underestimated, irrespective of simulated heritability. With a sampling interval of 1 this underestimation also occurred, but to a lesser extent. Again, absence of genetic variation was accurately identified (i.e., confidence interval overlaps zero only when the true variance is zero), a sampling

interval of 1 resulted in unbiased estimates for genetic variance of 0.25 and lower. Absolute precision decreased with increased simulated variance (Figure 6.4B). However, relative to the size of the estimate, precision increased.

6.3.5 Transmission rate parameter (β)

Figure 6.5 shows the effect of β on genetic variance estimates. Higher values of β (and thus higher R_0), result in epidemics where transmission goes faster, so that more individuals get infected per unit of time, and a greater proportion of the individuals becomes ultimately infected. Higher value of β generally led to increasing overestimation of the genetic variance in susceptibility. However, the pattern was inconsistent in the absence of simulated genetic variance (simulated heritability of 0), where overestimation decreased with increasing β. With simulated heritability of 1 (genetic variance 0.25), a low value of β of 0.12 and 0.15 resulted in a slightly underestimated genetic variance. Nevertheless, deviations from the simulated value were relatively small, except for $\beta = 1$ ($R_0 = 10$, reflecting an infection affecting on average ~90% of the individuals). Precision of estimates increased with simulated β, and was clearly lower for β of 0.12 and 0.15 (Figure 6.5B).

Table 6.3. Mean estimates of epidemiological parameters and variance components obtained by the four models for the default scenario. Estimates followed by their standard errors. True values are indicated in the header between brackets.

Model	β (0.20)	$\sigma^2_{A_\gamma}$ (0.125)	$\sigma^2_{E_\gamma}$ (0.125)	σ^2_{GT} (0)
1	0.14 ± .008	0.201 ± .0060	1.117 ± .0110	0.10 ± .004
2	0.19 ± .012	0.143 ± .0043	0.359 ± .0074	0.14 ± .004
3	0.20 ± .001	0.124 ± .0034	0.150 ± .0052	0.14 ± .004
4	0.20 ± .001	0.125 ± .0034	0.177 ± .0053	NA

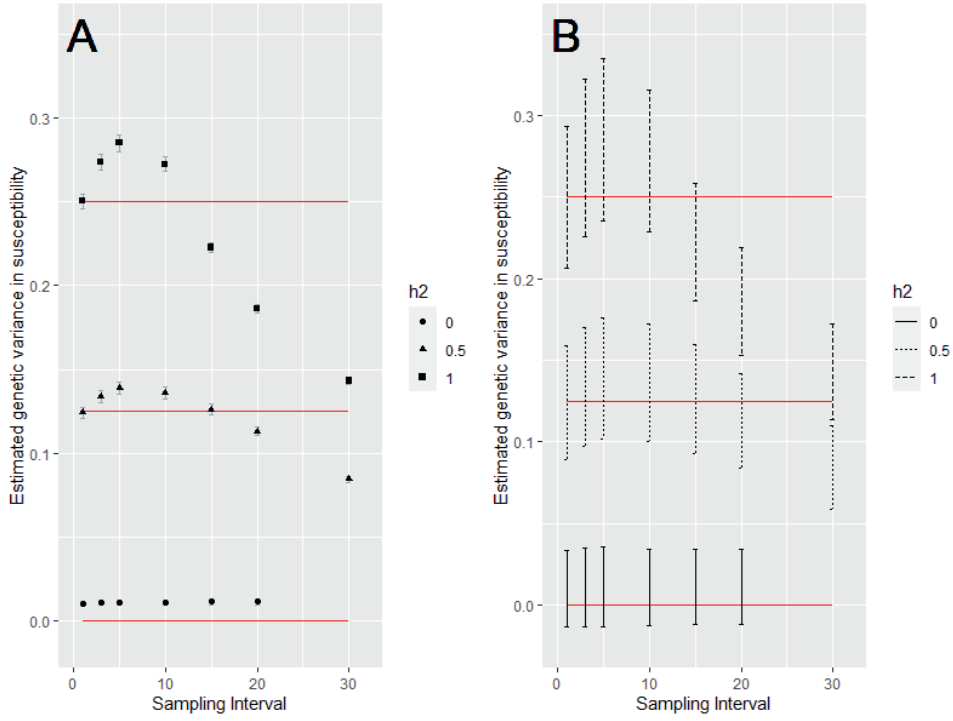


Figure 6.3. Bias in mean estimates of genetic variance in susceptibility (panel A) and mean confidence intervals of these estimates (panel B) for different sampling intervals and three values of heritability. Other variables as in the default scenario. Red lines indicate the simulated genetic variance, grey error bars in panel A indicate standard errors of the mean of 100 replicates. Note the missing estimate for interval 30, $h^2 = 0$, due to convergence problems.

6.3.6 Group size

Figure 6.6 shows that group size had a minor effect on the bias and precision of estimates of genetic variance in susceptibility. There seemed to be a little overestimation for some group sizes, but there was no clear pattern. The precision of the estimates was slightly lower for group size 12 (Figure 6.6B), which is likely caused by the relatively large proportion of seeder individuals (without observation) in that scenario.

6.3.7 Breeding value accuracy and bias

Figure 6.7A shows that the accuracy of estimated breeding values (EBVs), being the correlation between true and estimated breeding values, was little affected by the sampling interval. Higher heritability led, as can be expected, to higher accuracy, as did higher simulated genetic variance (Figure 6.7B). Despite the low observed-scale heritability of binary disease status, which was either 0.02 or 0.05 (see Table 6.2), the accuracies of EBVs for susceptibility suggest that meaningful response to selection can be obtained. Accuracy increased with the transmission rate parameter but reached a maximum at a certain level (Figure 6.7C). An intuitive explanation is that outbreak size is smaller with lower β (Figure 6.1), leaving a larger fraction of the population uninfected. There is no difference in susceptibility estimate between these uninfected individuals, thus breeding value estimation will be less accurate if a larger fraction of the population stays uninfected. The effect of group size on accuracy is similar as that of β (Figure 6.7D). In our design, each contact group had four seeder individuals that started the infection, irrespective of group size. Thus, the number of individuals for which meaningful records for susceptibility could be obtained, was larger in the larger contact groups, just because the total number of seeder individuals was lower in that case.

Next to the accuracy of sire EBVs, we also calculated accuracy for dam and offspring EBVs. The patterns for dam and offspring EBV accuracy were very similar to those in Figure 6.7 and align fully with our expectations based on population structure. Our population was simulated according to a full-sib half-sib design, with 50 offspring per sire, 10 per dam, and records only on the offspring. Therefore, the accuracy of offspring EBVs was higher than of dam EBVs, because each offspring had, next to own records, information on 4 full-sibs, and 45 half-sibs, whereas dams only had information from 10 offspring.

Figure 6.8 shows the bias in estimated sire breeding values for the logarithm of susceptibility for different sampling intervals. The pattern of bias in prediction accuracy is very similar to that of the genetic variance estimates (Figure 6.3), with no bias at interval 1, overestimation at intervals 3 to 10, and underestimation from interval 15. This holds for all scenarios investigated above.

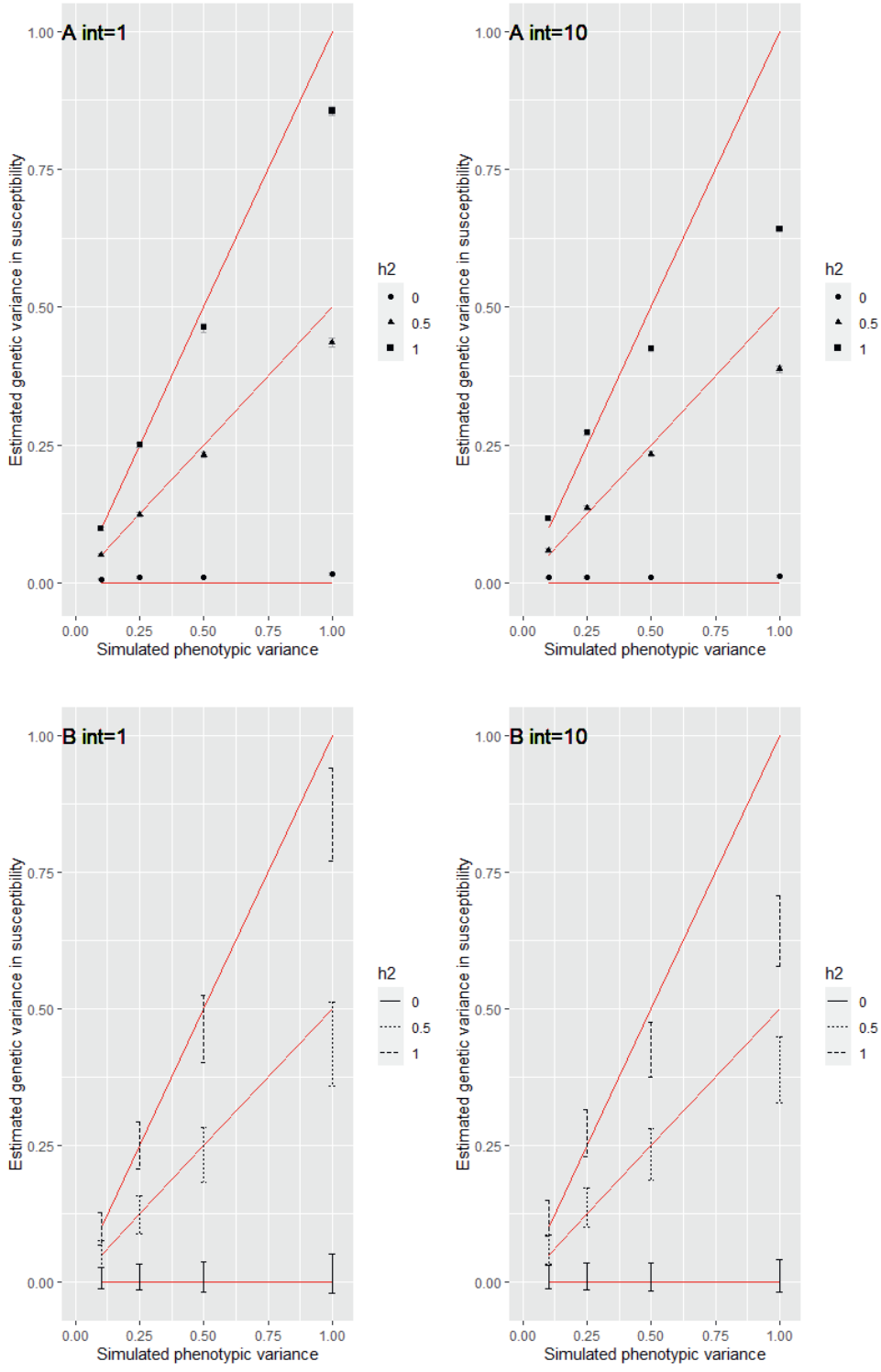

A int=1
A int=10
B int=1
B int=10
Estimated genetic variance in susceptibility
Simulated phenotypic variance
h2
0
0.5
1
0.00
0.25
0.50
0.75
1.00

Figure 6.4 (previous page) Bias in mean estimates of genetic variance in susceptibility (panel A) and mean confidence intervals of these estimates (panel B) for different simulated values of phenotypic variance in susceptibility, three values of heritability, and sampling interval of 1 (left panel) and 10 (right panel). Other variables as in the default scenario. Red lines indicate the simulated genetic variance, grey error bars in panel A indicate standard errors of the mean of 100 replicates.

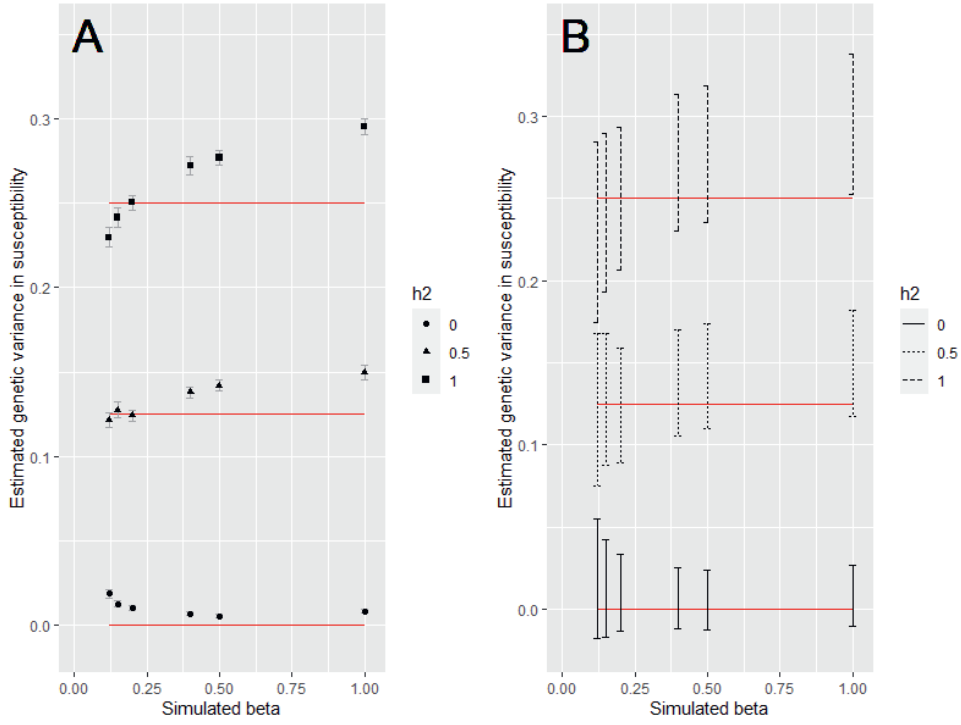


Figure 6.5. Bias in mean estimates of genetic variance in susceptibility (panel A) and mean confidence intervals of these estimates (panel B) for different simulated values of the average transmission rate parameter β and three values of heritability. Other variables as in the default scenario. Red lines indicate the simulated genetic variance, grey error bars in panel A indicate standard errors of the mean of 100 replicates.

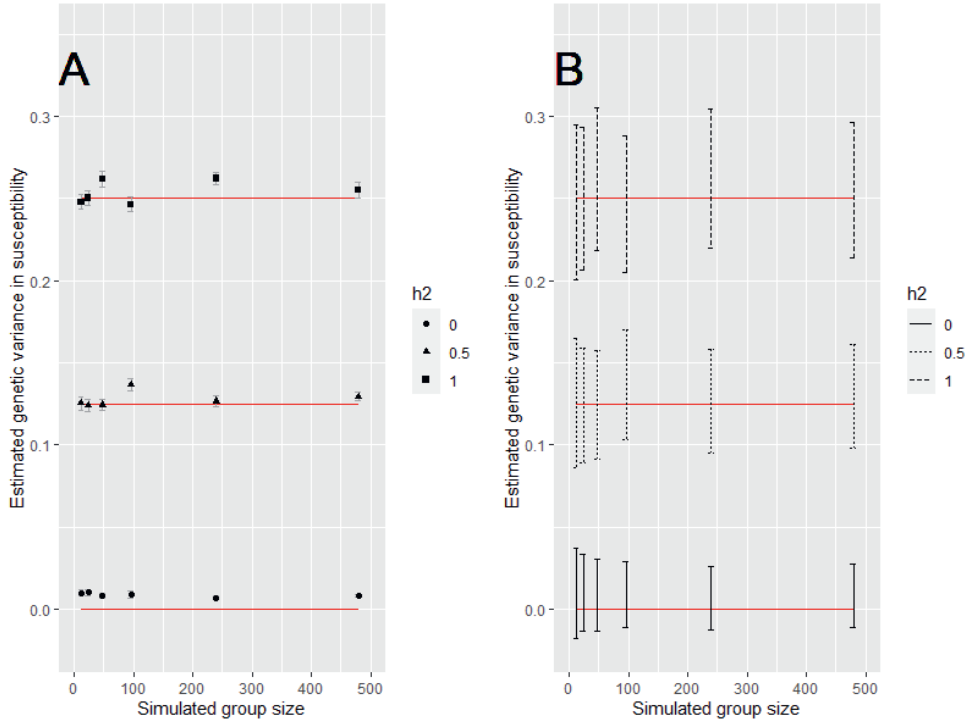


Figure 6.6 Bias in mean estimates of genetic variance in susceptibility (panel A) and mean confidence intervals of these estimates (panel B) for different simulated group size and three values of heritability. Other variables as in the default scenario. Red lines indicate the simulated genetic variance, grey error bars in panel A indicate standard errors of the mean of 100 replicates.

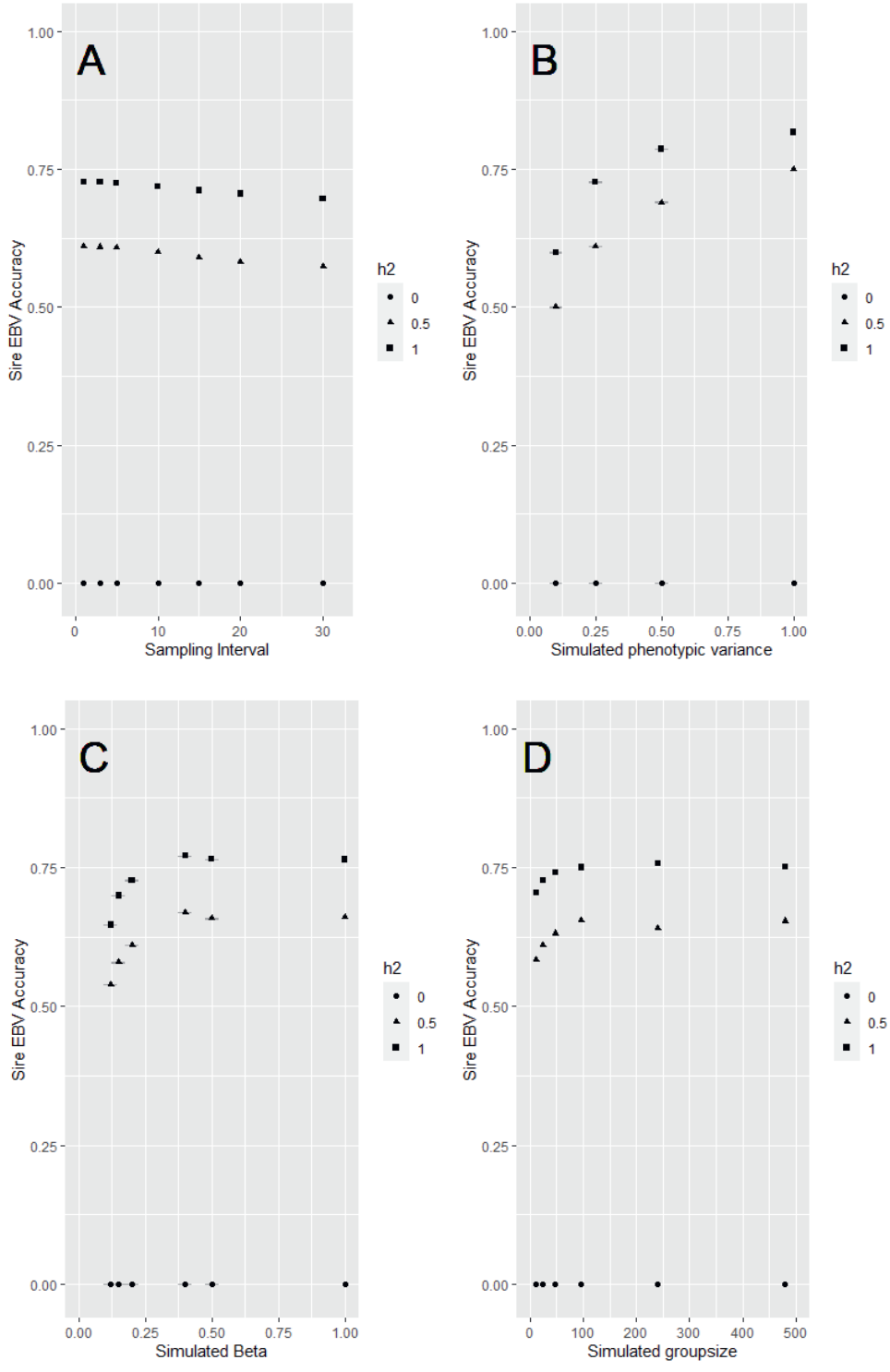

A
B
C
D
Sire EBV Accuracy
Sampling Interval
Simulated phenotypic variance
Simulated Beta
Simulated groupsize
h2
0
0.5
1
1.00
0.75
0.50
0.25
0.00
0
10
20
30
0.00
0.25
0.50
0.75
1.00
100
200
300
400
500

Figure 6.7 (previous page) Accuracy of sire estimated breeding values (EBV), for heritability of 0, 0.5, and 1. A. Different sampling intervals B. Different phenotypic variance C. Different simulated beta D. Different contact group size. Other values as in the default scenario (Table 6.2).

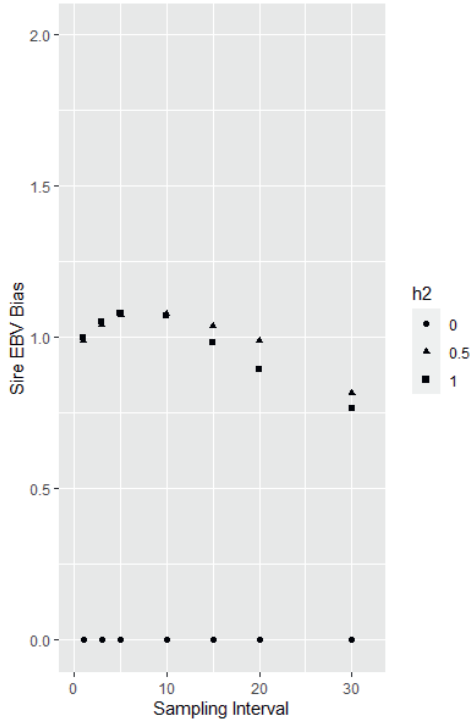


Figure 6.8 Regression coefficient of regression of estimated sire breeding value for the logarithm susceptibility on true sire breeding value for the logarithm of susceptibility, for different sampling intervals. Other values as in the default scenario. A regression coefficient of 1 means no bias in EBV, a regression coefficient >1 means that breeding values are overestimated, a regression coefficient of <1, means underestimation of breeding values.

6.4 Discussion

Here we showed that a GLMM is able to produce accurate and almost unbiased estimates of genetic variance for susceptibility to an infectious disease if longitudinal records of individuals' infection status are available. The sampling interval was the main factor affecting estimation, especially with a larger simulated genetic variance. In general, the shorter the sampling interval, the better the estimates.

Sampling interval is important because it determines the quality of the information. The input for the GLMM is binary data on susceptible individuals. In each interval, the observation per susceptible individual can either be a zero (not infected in the interval), or a one (infected in the interval). To accurately determine the susceptibility of an individual, each observation is corrected for the exposure to infectious individuals by including the offset in the model. However, if the sampling interval is too large, the information in the offset gets outdated and might therefore be incorrect. Either because additional individuals became infected after the start of the interval, resulting in an offset that underestimates exposure, or because individuals recovered after the start of the interval, giving an offset that overestimates exposure. Furthermore, a larger sampling interval will result in less accurate observation of the (differences in) infection times of individuals, which likely leads to a lower estimated genetic variation. For example, in a very long interval, nearly all individuals may become infected, making it impossible to discriminate between them. The group.time random effect captures (part) of the extra variation that arises due to the incorrect offset (see Figure 6.2), because that is estimated based on what is actually happening (i.e. the number of new infections) and not, like the offset, based on the fraction of infected individuals in the previous interval.

To get more grip on the sampling interval that is needed for accurate estimation of genetic parameters, we could look at the minimum expected interval between events (Figure 6.9), events being infection of susceptible individuals and recovery of infected individuals. The minimum expected interval depends on the transmission rate of the epidemic (β), the average recovery rate (α), the expected maximum number of infected individuals (I_{max}), derived deterministically, and the number of susceptible individuals in the contact group (S):

$$Min.expected\ interval = \frac{1}{\frac{\beta SI_{max}}{N} + \alpha I_{max}} \quad (6.6)$$

The effect of increase in β on the minimum expected inter-event interval seems to be visible as bias in the parameter estimates (Figure 6.5), where from $\beta = 0.4$ the genetic variance was overestimated. Figure 6.9 shows that the minimum expected interval decreased from ~15 days for $\beta = 0.2$ to ~3 days for $\beta = 0.4$, but it only decreases below 1 for a β of 1. Thus, the minimum expected inter-event interval as calculated here cannot fully explain this overestimation for $\beta = 0.4$. However, it is important to note that the actual minimum interval may be much smaller than the expected value, due to stochastic events, and thus that the chosen sampling interval is preferably shorter than the minimum expected interval.

The model comparison (Table 6.3) showed that the inclusion of fixed effects for group and time in the model led to biased and inaccurate estimates for β and the genetic variance in susceptibility. The overestimation of genetic variance is in contrast to our experience that the addition of a fixed effect in the model could only lead to a reduction in observed variance. The underestimation and large uncertainty in the estimate of β for these models, i.e. much larger standard errors than for the models without fixed effects, show that the models with fixed effects poorly picked up the mean level of transmission, which likely causes errors in estimation of the individual susceptibility effect and thereby probably has led to the overestimated susceptibility variance. This observation is important, since inclusion of fixed herd, year, and season effects in models of genetic analysis of livestock is common. Our results indicate that care should be taken with keeping such effects in a GLMM when they are not significant, since they might affect the variance estimates.

The usual approach for genetic analysis of infectious disease data is to analyse binary disease/infection status or an indicator trait (e.g. somatic cell count) with linear or GLMM-like threshold models (e.g. PÉREZ-CABAL *et al.* 2009; WELDERUFAEL *et al.* 2017; MARTIN *et al.* 2018). Importantly, these threshold models are not based on epidemiological principles. As far as we know, the only study so far that estimated genetic parameters for infectious disease susceptibility from real data

using exposure in a GLMM tailored to the transmission process is by BIEMANS *et al.* (2019a). Corresponding to our results of the group.time effect, they found a significant farm*period variance. In contrast to our results, however, dropping this effect from the model led to a (non-significant) decrease in estimated genetic susceptibility variance, instead of an increase. Their estimated susceptibility variance of 0.55 at a sampling interval of half the infectious period (which corresponds to 5 in our case), might be underestimated if we look at our results (see Figure 6.4). However, as indicated before, theoretical results suggest that a genetic variance in susceptibility of 0.25 is already at the higher end (HULST *et al.* 2021; BIJMA *et al.* 2022). Definitely, more empirical estimates of genetic variance in susceptibility using exposure as an offset in a multiplicative model are needed to get a better idea of realistic values.

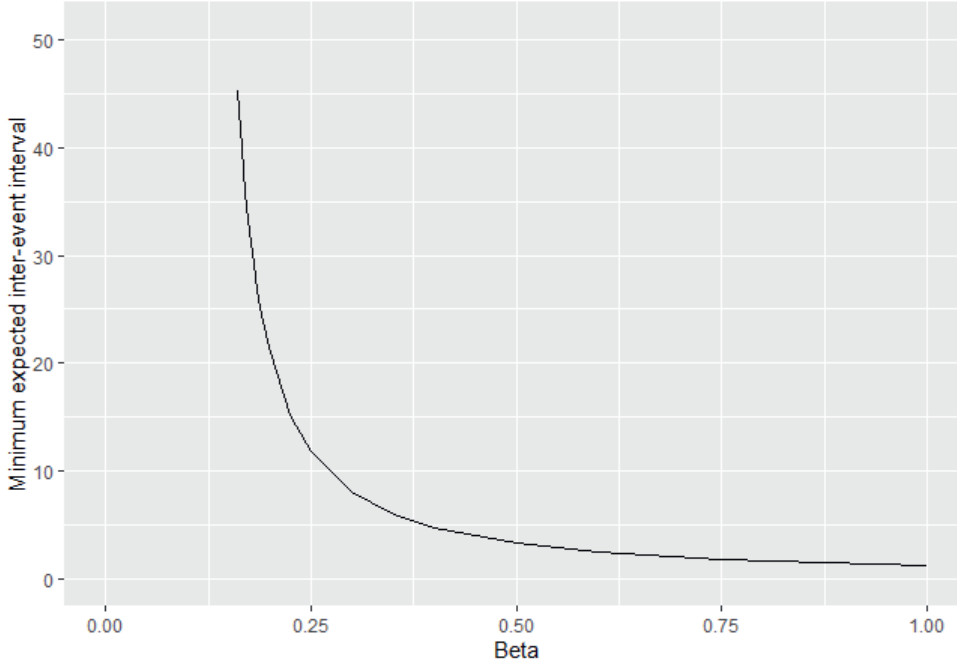


Figure 6.9 Minimum expected inter-event interval per individual in days as a function of beta, for an average duration of the infectious period of 10, as used in this paper. An event being infection of a susceptible individual or the recovery of an infected individual.

ANACLETO *et al.* (2015) did a study similar to ours, with simulated infection data, but a different, Bayesian method to estimate genetic parameters. They showed that group size has very little effect on prediction accuracy for susceptibility, which corresponds to our results in Figure 6.6. Furthermore, our results align with their indication that prediction accuracy starts to decline with increasing sampling interval, in particular when the interval becomes larger than the average duration of the infectious period.

Next to susceptibility, other host genetic traits affect the basic reproduction ratio of an infectious disease and are therefore relevant to selection. For example, infectivity, which is the propensity of an animal to infect herd mates once it has become infected, and recoverability, which is the ability of an animal to recover from infection. Given that there is evidence for genetic variation in both of these traits (ANACLETO *et al.* 2019; BIEMANS *et al.* 2019a; PRENTICE *et al.* 2022), two questions seem relevant for further investigation: (1) How does genetic variation in infectivity and recoverability, or other host genetic traits affecting disease transmission, if not accounted for in the analysis, affect estimates for susceptibility, and (2) how to obtain accurate estimates for infectivity and the other traits.

GLMMs can be used to estimate genetic effects for infectivity (ANCHE *et al.* 2015; BIEMANS *et al.* 2019a). However, previous studies show that estimates for infectivity effects are much more sensitive to experimental design than the estimates for susceptibility (e.g. ANCHE *et al.* 2015; POOLEY *et al.* 2022a). Nevertheless, the potential of selection on infectivity to reduce the prevalence of infectious diseases in livestock might be even larger than the potential of susceptibility, because genetic variation infectivity has not yet been harvested by breeders (TSAIRIDOU *et al.* 2018). Thus, there seems opportunity for further research into the suitability of GLMMs to estimate genetic parameters for infectivity and the corresponding data requirements.

Our results indicate that genetic parameters for susceptibility can accurately be estimated using GLMMs, at least when there is only variation present in susceptibility. Accuracy of estimation is not much affected by infection characteristics, but it requires frequent sampling of the infection status of individuals. The sampling interval should at least not exceed the average duration

of the infectious period, but is preferably much shorter, to prevent bias in the estimates of genetic variance. In practice, achieving such small sampling intervals is labour-intensive and might be unfeasible, especially in large populations. However, current developments in automated phenotyping might lead to change on this point. Investments in such systems seem justifiable, given the financial and societal impact of infectious diseases in livestock. In conclusion, the GLMM seems a robust, accurate and relatively easy implementable model to estimate genetic parameters for susceptibility. Thereby, it might give an important contribution to better quantification of the potential of genetic selection to reduce the impact of livestock infectious diseases.

Chapter 7

General Discussion

7.1 Introduction

This thesis reveals a critical difference in aim and approach between animal breeding and epidemiology in controlling infectious diseases. In animal breeding programs, the rate of genetic change per trait is typically small, due to a large number of traits in the breeding goal and antagonistic correlations between those traits. For infectious disease traits, this corresponds to small reductions in endemic prevalence. Such small reductions are in full contrast to the goals in veterinary epidemiology. There, the aim of an infection control program is typically to locally eradicate the infectious disease, and thus to achieve large reductions in prevalence. Here, 'local' can be farm, regional, or national level and eradication is typically achieved by applying a combination of measures, such as vaccination, trade restrictions, or test and removal (e.g. EUROPEAN COMMISSION 2020).

Chapter 5 shows that the current approach in animal breeding typically creates an 'invasion window' for mutants of the pathogen that can escape the genetic resistance of hosts. In this window, these mutants might spread through the population and thereby diminish the gradual reductions in prevalence that were achieved through genetic selection. Therefore, I argue that the goal for selection against infectious diseases in animal breeding should be changed from gradual reductions in prevalence to eradication. In that way, the approach in animal breeding corresponds to the approach in epidemiology.

From an animal breeding perspective, it might seem impossible to resemble the epidemiological approach. First, because accepted quantitative genetic theory predicts that breeding a binary trait to a prevalence of 0, i.e. eradication, is impossible (ROBERTSON 1950). Second, because if achieving large reductions in prevalence in a few generations of selection is possible at all, this might be undesirable in a breeding program with many traits due to unfavourable correlated responses in other traits and high inbreeding. However, the results in this thesis indicate otherwise. In Chapters 2 and 3 I showed that eradication is, in principle, possible using genetic selection, because of indirect effects that occur in the transmission of infectious diseases. These indirect effects are very similar to the mechanism underlying herd immunity following vaccination. The next section of this general discussion further elaborates on the indirect genetic effects in infection transmission. As shown in Chapter 4, the indirect effect can be increased

by combining genetic selection with other infectious disease interventions, such that less genetic change in an infectious disease trait is required to achieve large reductions in infection prevalence.

However, changing the breeding goal for infectious disease traits from 'gradual reduction in the prevalence of the infection' to 'eradication of the infection', requires specific changes to a breeding program (e.g., traits used, phenotypes collected) and to the way genetic material is distributed into (commercial) livestock populations. In Chapter 6, I developed important insights into how breeding values and genetic parameters for infection susceptibility can be estimated accurately from longitudinal infection data. There, I showed that a relatively short observation interval and a statistical model that accounts for exposure are needed for unbiased and accurate estimation, while group size and the transmission characteristics of the disease had little effect. In this general discussion, I further explore what such a restructured breeding program could look like, which traits could be selected for, what kind of phenotypes are needed, how these phenotypes can be collected, and how herds should be composed both in the breeding and production population. Finally, I will discuss the implications of the findings in this thesis for future research and more broadly for agriculture and society.

7.2 Indirect genetic effects in infectious disease transmission

The indirect genetic effect caused by infectious disease transmission is a recurring theme in this thesis. Indirect effects are well known in infectious disease epidemiology from amplifying the effect of infectious disease interventions and causing herd immunity (Chapter 4; (FINE 1993)). An indirect effect occurs for interventions that decrease the chance of an individual to become infected or its propensity to infect others. It arises because of a reduced infection pressure in the population, which decreases the probability of all individuals to become infected. This reduced infection probability of all individuals feedbacks again on the infection pressure in the population.

Indirect genetic effects are effects of the genotype of an individual on the trait value of other individuals (GRIFFING 1967; MOORE *et al.* 1997; MUIR 2005; BIJMA

2014). An example of an indirect genetic effect can be found in selection for increased individual survival of laying hens, which might lead to increased population mortality (CRAIG AND MUIR 1996; MUIR 2005). Here, the favourable genotype for individual survival has an unfavourable indirect genetic effect on survival of other individuals in the group. The indirect genetic effect of susceptibility, discussed in this thesis, is thus an effect of the genetic susceptibility of an individual on the probability to become infected of its herd mates. In contrast to laying hen survival, both the direct and indirect genetic effect are in the same direction, i.e., selection for decreased individual susceptibility will lead to an overall reduction of infection probability in the population.

In Chapter 2, the indirect effect of susceptibility causes the large response to selection in infection prevalence and eventually leads to eradication of the infection following simple sire selection. In Chapter 3, the size of this indirect effect is quantified relative to the ordinary breeding values of infectious disease status (the direct effect). The total breeding value for prevalence, which includes the indirect effect, is a factor of the inverse of the endemic prevalence larger than the breeding value for disease status. Thus, for endemic prevalence of one in four (0.25), the total effect of selection for reduced susceptibility on infection prevalence is four times greater than predicted by the breeding value for disease status. In chapter 4, the result of chapter 3 is placed in a general epidemiological context, and I show that the factor inverse of the endemic prevalence gives a good approximation of the overall effect of any infectious disease intervention, irrespective of intervention coverage and effect. Chapter 5 deals with coevolution of the pathogen and determination of the invasion window. However, also in this chapter the indirect effect of susceptibility comes into play when the upper limit of the invasion window is defined: the point of eradication. In chapter 6, the indirect effects of susceptibility are the main motivation for the use of a GLMM that corrects for exposure (i.e., the infection pressure in the herd) for estimation of genetic parameters and breeding values for susceptibility.

The notion that genetic selection on susceptibility leads to indirect genetic effects is in itself not new. Back in 1997, BISHOP AND STEAR (1997) identified a 1.7 times larger response to selection for resistance of sheep to gastro-intestinal parasites than predicted by quantitative genetic models. As they wrote, this was due to the

epidemiological dynamics of the parasite infection. Here, resistance is basically an inverse definition of the same trait as susceptibility, with low resistance being high susceptibility and high resistance being low susceptibility. Further demonstration of the indirect effect of susceptibility was given by NIEUWHOF *et al.* (2009), who stated that selection for resistance, including the epidemiological dynamics, has the potential to double the expected selection response in prevalence of foot rot in sheep. In a theoretical paper on defining the heritable variation in R_0, ANCHE *et al.* (2014) showed that selection of related individuals leads to higher response in R_0, because the indirect effect of susceptibility is (partly) included in selection in that way.

Nevertheless, most emphasis when it comes down to improving genetic selection against infectious diseases has gone to selection on infectivity, the propensity of an infected individual to infect others (LIPSCHUTZ-POWELL *et al.* 2012; ANCHE *et al.* 2014; ANACLETO *et al.* 2015; BIEMANS *et al.* 2017; TSAIRIDOU *et al.* 2018; ANACLETO *et al.* 2019; TSAIRIDOU *et al.* 2019). Unlike susceptibility, which has both a direct and indirect effect on the infection status of an individual animal, the effect of infectivity on infection status is fully indirect. This implies that selection of an individual for reduced infectivity will not directly affect its own probability of getting infected, only the probability of herd mates. Likewise, breeding values for infectivity are not easy to estimate with linear models of infection status, but require specific designs with respect to group size and relatedness within and between groups (BIJMA 2010; ANCHE *et al.* 2015; POOLEY *et al.* 2022a).

An interesting motivation for the interest in infectivity is found in evolutionary theory by LIPSCHUTZ-POWELL *et al.* (2012). The evolutionary argument is that natural selection has exhausted genetic variation in susceptibility, because susceptibility is a component of an individual's own fitness. Since infectivity is not a component of the own fitness of an individual, but only of its groupmates, genetic variation in infectivity is less affected by natural selection. This leads to the expectation that, if accurate estimates of breeding values can be obtained, selection for infectivity yields larger response in infection prevalence than selection for susceptibility. A related argument is that susceptibility has already been selected for in breeding programs that select for lower infectious disease status (LIPSCHUTZ-POWELL *et al.* 2012). However, results of such selection are

typically estimated to be limited, because of the low heritability of disease status. This low expectation also raises doubts about whether selection on susceptibility will be effective and might have limited the selection effort put into the trait. The results in Chapter 2 of this thesis indicate otherwise. The typical low heritability estimates for infectious disease status actually correspond to a considerable genetic variation in susceptibility, and indirect genetic effects lead to considerably increased response to selection in prevalence.

Given the considerable indirect effect of susceptibility, especially at lower endemic prevalence, and given that the indirect effect of susceptibility is fully correlated to the direct effect, questions arise whether the current selection practice already benefits from the additional indirect response that is expected when animals are effectively selected for lower susceptibility. Signs of such additional response are scarce in literature (see e.g., the examples in Chapter 2), but further investigation is required for a detailed answer. Several hypotheses for this lack of examples have been proposed in this thesis. Pathogen adaptation might play a role, given that genetic progress is typically slow in a multi-trait selection program, with unfavourable correlations between resistance and production traits (e.g. RUPP AND FOUCRAS 2010). Another problem is that farm or herd effects, which are usually fitted in genetic models, can mask the susceptibility of individual animals. When a herd consists of predominantly lowly susceptible animals, the indirect genetic effects lead to a (much) lower than average prevalence of the infection in that herd. However, the herd effect attributes this indirect genetic effect to non-genetic herd factors, such that the breeding value for susceptibility of the animals in this herd is overestimated, and the genetic potential underestimated. When a herd consists of predominantly highly susceptible individuals, the herd effect has the opposite effect, leading to overestimation of the genetic potential.

7.3 Genetic variation in disease traits

A critical condition for the large potential of genetic selection for lower infectious disease prevalence is the availability of genetic variation in susceptibility (and potentially in other traits). Since susceptibility is an unobservable trait, a method tailored at identifying variation in susceptibility from observed infection data, e.g., using a GLMM with an offset that corrects for exposure as in chapter 6, is needed to get estimates. Since these methods are relatively new and not widely implemented, empirical estimates for genetic variation in susceptibility are still scarce. The only example so far seems to be from data on digital dermatitis in dairy cattle, by BIEMANS *et al.* (2019a), which estimates a genetic variance of 0.55 for susceptibility on the log-scale. This variance of 0.55, with a phenotypic mean of 1, corresponds to a genetic coefficient of variation of 74%. On the exponential scale this variance of 0.55 roughly corresponds to a 20-fold difference in susceptibility between the 5% least and 5% most susceptible animals, which is considerable and close to the upper bound of genetic variance in susceptibility as identified in Chapter 2.

Nevertheless, numerous studies use classical quantitative genetic methods on infection status data to estimate heritability and genetic variation in infectious disease status (e.g. PÉREZ-CABAL *et al.* 2009; WELDERUFAEL *et al.* 2017; MARTIN *et al.* 2018). Typically, such studies find a heritability between 0.02 and 0.10. In Chapter 2, I showed that these typical heritability values correspond to a substantial genetic variation in susceptibility. Next to susceptibility, the heritable variation in infection status might also arise from genetic variation in recoverability, which determines the duration of the infectious period. The effect of both traits on R_0, and thus on endemic prevalence, is similar. Consequently, the in Chapter 2 identified genetic variation in susceptibility should actually be interpreted as genetic variation in susceptibility, recoverability, or a combination of the two.

Genetic variation might also be present in other epidemiological traits that affect infection transmission but are not directly captured by a linear model of infection status, such as infectivity. For infectivity, BIEMANS *et al.* (2019a) estimated a genetic variance of a magnitude 50 times the susceptibility variance, which seems immense, but at least provides a strong indication that genetic variation for

infectivity of digital dermatitis is present in dairy cows. Furthermore, experimental studies of fish showed evidence for the presence of genetic differences in infectivity (Doeschl-Wilson *et al.* 2018; Anacleto *et al.* 2019). In contrast to infectivity and susceptibility, the presence of genetic variation in recoverability has less clearly been demonstrated with empirical data. Nevertheless, one of the studies in fish showed genetic variation in infection 'endurance', which also entailed the duration of the infectious period (Anacleto *et al.* 2019). Definitely, more empirical estimates of genetic variation in susceptibility and the other traits are needed to unwind the true potential of genetic selection for decreased infection prevalence. The results in chapter 6 indicate that for susceptibility estimates can be obtained from both experimental and field data, but that repeated observations with a short observation interval are needed to correctly link infection to exposure and thus for accurate and unbiased estimation.

7.4 Resistance vs. Resilience vs. Tolerance

Three traits are typically discussed when the aim is to reduce the impact of infectious diseases in livestock by using genetic selection: resistance, tolerance, and resilience (Doeschl-Wilson and Kyriazakis 2012; Bishop and Woolliams 2014; Knap and Doeschl-Wilson 2020). Resistance is defined as the ability of an animal to prevent itself from becoming infected or to clear itself from infection, which has an impact on transmission of the pathogen and on the clinical outcome in the animal (Råberg *et al.* 2007; Bishop and Woolliams 2014). Susceptibility, which is used throughout this thesis, is essentially the reciprocal of resistance (highly resistant individuals have a low susceptibility and vice versa), with the same effects on transmission and clinical outcome. Tolerance is different from resistance, because the focus of selection for tolerance is reducing pathogenic outcome of infection (i.e. clinical signs), with no emphasis on the consequences for transmission (Råberg *et al.* 2007; Bishop and Woolliams 2014). A common definition of resilience is the ability to maintain performance (production) under challenges from the environment (Albers *et al.* 1987; Berghof *et al.* 2019b). With this definition, resilience is more or less a 'black box' trait, where the underlying mechanisms and the effects on pathogen transmission are typically unknown. Furthermore, the challenges from the environment that determine resilience often

not only include pathogens, but also other environmental factors such as heat stress and poor air quality, and management or handling stressors.

Question here is which of the three traits is best to use in a breeding program that has the goal to reduce the impact of infectious disease in the population. The advantage of resistance over tolerance is that it not only reduces clinical signs in the resistant individuals, but also has an effect on transmission of the pathogen. This reduced transmission results in an indirect effect through lower prevalence of the infection in the population, such that also fewer resistant individuals have a lower probability of becoming infected. Chapters 2, 3 and 4 show that this indirect effect can be considerable and can even be used to locally eradicate the infection. An argument used against selection for resistance is that resistant animals exert selection pressure on the pathogen by reducing its transmission (e.g. BISHOP AND STEAR 2003; GUY *et al.* 2012). This selection pressure will sooner or later lead to the evolution of adaptations of the pathogen that enable it to escape the genetic resistance of animals and thereby to diminish the effects of genetic selection. In chapter 5, I showed that pathogen escape can be prevented by increasing the level of resistance of the exposed host population fast or at once to a reproduction ratio below 1, such that the pathogen dies out locally before adapting to the selected host population. However, such a strategy is totally different from the current approach in animal breeding and would require considerable changes to the breeding infrastructure, which will be discussed later.

Tolerant animals, on the other hand, are often assumed to exert no selection pressure on the pathogen population, because they still fully contribute to transmission. A common expectation is thus that selection for tolerance will not be counteracted by adaptations of the pathogen (RÅBERG *et al.* 2009; DOESCHL-WILSON AND KYRIAZAKIS 2012). Although this reasoning might seem logical, it overlooks a very important aspect of the coevolution between host and pathogen. The classical way of thinking about such coevolution is that pathogens eventually evolve to be non-harmful to their hosts, because they depend on the host for reproduction and it is thus most beneficial for pathogen fitness to keep the host alive as long as possible (e.g. MAY *et al.* 1983). In other words, pathogens that kill their host also spoil their own chance to reproduce. Nevertheless, there are currently still numerous pathogenic parasites, some of which have been

pathogenic for thousands of years (MAY *et al.* 1983; MAY AND ANDERSON 1990). An explanation for this existence of pathogenicity is that some degree of harm to the host is often beneficial for transmission of the pathogen, the so-called virulence-transmission trade-off (MAY *et al.* 1983). The classical example of this trade-off is Myxoma virus in European rabbits, which is transmitted by flies and mosquitos. The virus causes skin lesions and is highly lethal (FENNER AND WOODROOFE 1953). When introduced into a naïve rabbit population, mortality due to Myxoma virus is more than 99% (FENNER AND MARSHALL 1957). This mortality rapidly reduces to more intermediate levels, but not further to zero (FENNER AND WOODROOFE 1965). The pathogen is picked up by the vector species from skin lesions, which are thus essential for the pathogen to be transmitted. The trade-off is that if skin lesions become too severe, the host (the rabbit) rapidly dies, and the virus can no longer transmit. Hence, long term evolution of pathogens has resulted in some optimum level of virulence, corresponding to a maximum effective reproduction ratio (MAY *et al.* 1983). In addition, competition between pathogen strains within the individual (killer-milker dilemma) might lead to more virulent pathogens (VAN BAALEN AND SABELIS 1995). A slowly transmitting pathogen that does little or no harm to its host might be outcompeted by a pathogen that induces a stronger immune response or kills the host but transmits much faster than the harmless pathogen. Although the fast-transmitting pathogen might achieve less transmission in total, it spoils the chances of transmitting for the harmless pathogen in the same host by killing the host.

Since selection for tolerant hosts essentially aims to make pathogens harmless to the host, it might be counteracted by the two evolutionary mechanisms described above. Thus, apart from reducing disease in the host population, increased toleration of the pathogen by the host might open up opportunities for more harmful mutants of the pathogen to invade the host population. Such harmful strains can outcompete the tolerated strain by achieving higher transmission through the initiation of clinical signs or by killing the host and thereby spoiling the host environment for the tolerated strain. In any case, the outcome will be that selection for tolerance is counteracted by adaptations in the pathogen that lead to the occurrence of clinical signs in tolerant animals again. Furthermore, the disease might be (much) more severe in the less tolerant animals due to the new, more harmful, strains of the pathogen that spread through the host population.

This phenomenon has been observed in practice in vaccination campaigns against pertussis (whooping cough) in humans, and Marek's disease in chickens (READ *et al.* 2015; MILLER AND METCALF 2019). In the latter case, mortality of unvaccinated birds increased from about 80% to nearly 100%, due to virulence-increasing adaptations of the virus. Thus, the general assumption that selection for tolerance will not be counteracted by evolution of the pathogen largely neglects the complex coevolution between pathogen and host, and thereby leads to an overestimated potential of selection for tolerance.

Next to tolerance and resistance, resilience can be considered a third possible trait to select for when the goal is to reduce the impact of infectious diseases in livestock populations. Heritable variation in resilience has been identified in various species and based on different production traits (ELGERSMA *et al.* 2018; BERGHOF *et al.* 2019a; PUTZ *et al.* 2019; POPPE *et al.* 2021). Typically, resilience is measured as the consistency of production over time, in a challenging environment. Animals that are classified resilient show less fluctuations in production and recover faster from production drops (ALBERS *et al.* 1987; BERGHOF *et al.* 2019b). If infections are part of the environmental challenge, such resilient animals can either be tolerant or resistant to infection, or a combination of the two (KNAP AND DOESCHL-WILSON 2020). In contrast to the typical definition of resistance and tolerance, however, resilience is not an infection-specific trait. This makes it an interesting trait to select for from a practical point of view, because livestock species are usually challenged by multiple pathogens in their commercial environment. However, because of the 'black box' character of the trait, it is often unclear whether the genetic variation in resilience actually corresponds to resistance or tolerance for multiple pathogens.

Interesting research to get more insight into the multiple pathogen background of resilience is that of a Natural Disease Challenge Model (NDCM) of pigs in Canada (PUTZ *et al.* 2019; BAI *et al.* 2020). Here, pigs reared in nucleus environments with high health status (pathogen free), were over a period of several years introduced into a commercial, dirty, environment, and there exposed to herd mates infected with a variety of pathogens that typically occur in such environments. One of the outcomes of this experiment was the identification of a general resilience trait with moderate heritability. Resilience being the variation in individual feed intake over

time, with the motivation that pigs eat less when they are ill (PUTZ *et al.* 2019). In other words, genetics affects the feed intake of individual pigs on a farm where they are confronted with a large variety of pathogens. Also in this experiment, however, the exact mechanisms of resilience are often unknown, might very well differ between individuals, and likely also contain response to other environmental factors than pathogens. Thus, when selection is for a resilience trait, it will be difficult to manage the risk of pathogen adaptation. The strategies I proposed in chapter 5 will, for example, be more difficult to implement if part of the animals continues to shed pathogens, without showing any clinical signs of infection. Such asymptomatic shedders make it harder to eradicate an infection from a population, leaving opportunity for the emergence and invasion of escape mutants of the pathogen.

A potential solution to the above mentioned problems with resilience would be to shift from individual resilience to a concept of herd resilience (DOESCHL-WILSON *et al.* 2021). The principle of herd resilience seems especially relevant and beneficial for infectious diseases. Essentially, this approach aims to reduce the basic reproduction ratio (R_0) of the infection in the local population, preferably below 1. Thus, the results of this thesis might contribute to increased herd resilience. Evidently, selection for reduced susceptibility with effort to make use of the occurring indirect genetic effects, the common thread of this thesis, also results in a reduction of R_0, and thus in an increased herd resilience. Selection for reduced infectivity might act as a powerful addition to this. Another possible trait that fits in the herd resilience concept is to select for improved response to vaccination. Results of challenge tests provide a first indication that genetic variation in vaccine response exists (DUNKELBERGER *et al.* 2022). Also here, this is particularly useful if a vaccine reduces the transmissibility of the infection and might be risky if the vaccine only reduces the clinical outcome of infection. As noted by DOESCHL-WILSON *et al.* (2021), an increased effect of vaccination enhances (local) eradication programmes, i.e. to come to resilient herds. Furthermore, a combination of interventions will be much more powerful than (the sum of) separately applied interventions, in decreasing prevalence as well as in preventing pathogen adaptation, as discussed in Chapters 4 and 5.

7.5 Collection of phenotypes

One of the challenges in selection for reduced susceptibility (and potentially infectivity) is the collection of phenotypic data of sufficient quality, e.g., with a short observation interval (Chapter 6), and on a sufficient number of animals. In commercial animal breeding programmes, the selection candidates represent high economic value and are therefore kept in environments with high biosecurity, i.e., specific-pathogen free in pigs and poultry. Consequently, it is not possible to collect infectious disease phenotypes on the selection candidates themselves. Thus, the phenotypic information of selection candidates needs to come from relatives, which could be either close relatives excluded from selection (e.g., siblings) that are exposed to a pathogen in a transmission experiment, or the descendants in the commercial environment, likely exposed to all relevant pathogens. Here, genomic selection is of large added value, because it decreases the need for continuous phenotypic information.

Note that there is an important difference between challenge test, which are already conducted in animal breeding, especially in aquaculture, and transmission experiments that I suggest here. In a typical challenge test, every individual animal gets artificially infected with a certain pathogen, after which its performance under this pathogen challenge is assessed. In aquaculture, for example, the trait of interest in such experiments is usually mortality (ELASWAD AND DUNHAM 2018; ROBINSON *et al.* 2023). In transmission experiments, on the other hand, uninfected individuals are exposed to (artificially) infected group mates. In epidemiology, such experiments are generally used to estimate transmission parameters and the effect of vaccines on those (e.g. VAN DER GOOT *et al.* 2005). Which of the two experiments to prefer is highly dependent on the trait of interest. If interest is purely in performance of the animal while it is infected, a challenge test might be desired, because of the certainty that every individual gets infected. Still, care should be taken in interpretation of the results because artificial infection, e.g., through inoculation, might yield a different clinical outcome than natural infection. If interest is, however, in a trait that measures the (genetic) contribution of an individual to transmission, such as susceptibility as used in this thesis, a transmission experiment is to be preferred. Challenge tests will be uninformative for a transmission trait, because of the unobservable

differences in transmission (i.e., everyone gets infected at the same time). In Chapter 6, data from simulated transmission experiments was used to estimate genetic parameters and breeding values for susceptibility, which showed low bias and high accuracy. Importantly, individuals that do not get infected during the experiment also provide information on their susceptibility (i.e., low susceptibility). Furthermore, results of transmission experiments can be used to estimate genetic effects for infectivity (POOLEY *et al.* 2020; POOLEY *et al.* 2022b), which seems impossible from challenge tests.

Relatives of the selection candidates that are exposed in a transmission experiment with a certain pathogen, might act as a source of phenotypic information, but could also be used to construct a genomic reference population. This approach has several advantages over using commercial information. Firstly, there is a controlled environment with full control on group composition, i.e., relatedness within and between groups, and observation interval, and in which it is exactly known to which pathogen the animals are exposed. Then, the experimental design can potentially be optimised to obtain estimates for infectivity as well (BIJMA 2010; POOLEY *et al.* 2022a). Secondly, in poultry and pigs the commercial animals are all crossbreds, and estimating breeding values of pure-line animals based on cross-bred information might be challenging, dependent on the genetic correlation between purebred and crossbred performance (BIJMA AND VAN ARENDONK 1998). Lastly, it might be desired to select for increased resistance to infections which are normally not present in the commercial environment, but form a severe problem if they are introduced, such as highly pathogenic avian influenza. Controlled transmission experiments in biosecure facilities are virtually the only possibility to estimate genetic parameters and breeding values for such infectious diseases, although the required number of animals for estimation of genetic parameters will make it challenging to find a suitable facility. An advantage of the use of phenotypes of animals in the commercial environment is that these are the environments in which the genetic improvement is relevant and measurable. Often, the animals are exposed to different, potentially interacting, pathogens, which might not be easy and very costly to mimic in an experimental setting.

The optimal solution will be largely species dependent. In dairy cattle, data will likely come exclusively from the commercial environment, but, in the absence of crossbreeding in most dairy cattle populations, without the problems of cross-bred data mentioned before. In aquaculture, on the other hand, large challenge tests, typically with family groups, are already conducted, also for infectious diseases. Here, size of the tested groups is typically large (>1000 animals). As shown by the results of chapter 6, however, it is very well possible to obtain accurate and unbiased genetic parameter estimates and breeding values for susceptibility from large groups (either commercial or in a testing environment), as long as the infection status of individual animals is frequently observed. For a correct estimation of infection transmission traits (such as susceptibility), however, the current challenge test, where every individual gets infected, should be changed to a transmission experiment (see above).

Frequent observation of infection status will be labour intensive, and routine collection of commercial data might therefore not be feasible at the moment. However, advancements in automated phenotyping using sensor data or video images might make frequent collection more feasible in the very near future. These methods will likely identify an infected animal based on the clinical manifestation of the infection (e.g. FERNANDES *et al.* 2020). For digital dermatitis, for example, image analysis might be used to identify claw lesions. Also, mobility data could act as an indicator trait of certain infections. A challenge with such approaches might be to reach a high enough sensitivity (i.e. truly infected animals are classified as infected) and specificity (i.e. truly uninfected animals are classified as uninfected), since both low sensitivity and specificity negatively affect the results of genetic selection (e.g. BISHOP AND WOOLLIAMS 2010; RAPHAKA *et al.* 2018). A further challenge if identification of infection is based on clinical signs is to make sure that selection really targets the probability of infection (i.e., resistance) and not only the manifestation of clinical signs (i.e., tolerance). Potential undesired effects of selection for tolerance were discussed in previous paragraph.

7.6 Implementation in practice

Two main findings of this thesis highlight the importance of herd composition for achieving optimal benefit of genetic improvement in infectious disease susceptibility. Here, with 'herd' I mean a group of animals that acts as a relatively isolated population in transmission of an infection. Thus, a herd can be a single farm, but also a single barn or division on a farm (e.g., when biosecurity measures are in place between barns/divisions), or a group of farms (e.g., when there is frequent exchange of animals between these farms). First, a herd should be composed in such a way that it optimally benefits from the indirect genetic effects in susceptibility and that local eradication is most feasible. The results in chapter 2 show that an animal with a low breeding value for susceptibility has smallest probability to become infected when all other animals in the herd have a low susceptibility as well. However, when part of the herd mates has a high susceptibility, this indirect effect diminishes, and the infection probability of the lowly susceptible animal considerably increases. Chapter 3 and 4 show that the relative importance of this indirect effect is equal to the inverse of the endemic prevalence. Thus, a herd composition that exploits indirect genetic effects becomes more and more beneficial if an infection has a lower prevalence.

Making optimal use of the indirect effect becomes even more important with respect to prevention of pathogen adaptation, as was shown in Chapter 5. To prevent pathogen adaptation, it is important to increase the resistance level of the population as fast as possible to achieve eradication of the infection. Due to the risk of pathogen adaptation, it can even be a waste of effort to start genetic selection for reduced susceptibility without implementing a thorough strategy to come to eradication of the pathogen from farms. This strategy should not only concern the way genetic material is brought into the commercial environment, but also other interventions that can be combined with breeding to come to faster eradication. As proposed in chapter 5, gradual improvement in susceptibility might not be sufficient if it is brought directly into commercial populations. An option could be to gradually improve susceptibility in the breeding population, down to a level for which immediate eradication can be expected ($R_0 < 1$), before the animals selected for reduced susceptibility are used to breed commercial stock. However, such a strategy is obviously only possible for species with a separate, isolated,

breeding population in which the target pathogen is absent, such as in poultry and pigs. For dairy cattle, more complicated approaches are needed, requiring, for example, sorting of the animals based on their susceptibility in different barns or farms, with biosecurity measures between the separate groups. Whether such an approach is realistic and desirable depends on farm size, but also on (economic) importance of the disease, and the (im)possibility to control it using other interventions.

Once local eradication is achieved, continuous new introduction of the pathogen on farm might impose a risk. Although such introductions will only lead to minor outbreaks, affecting a small number of animals, mutants of the pathogen that are able to escape host resistance might emerge during such minor outbreaks. Once emerged, they might spread through the population diminish the reduction in prevalence or eradication that was achieved through genetic selection. Whether such a new strain of the pathogen will eventually maintain itself probably depends on the primary source of infection. If this primary source is the livestock population, e.g. as seems the case for PRRSV in pigs (GALVIS *et al.* 2022), it is very likely that adapted strains will take over from the original (wild type) pathogen, especially when genetic resistance is spread throughout the livestock population. If the primary reservoir of infection is outside the livestock population, for example in wild birds as for avian influenza, replacement of the wild type with the escape strain of the pathogen might be less likely, because the escape strain lacks a selective advantage in the wildlife population. Of course, also here transmission of the escape strain between livestock herds needs to be prevented.

The success of several eradication programs for livestock infectious diseases in the European Union indicates that local eradication of pathogens is achievable in practice (EUROPEAN COMMISSION 2020), and that farmers and other parties in the livestock sector are willing and capable of adhering to such eradication programs. For infections that have not yet been successfully controlled, genetic selection could be a valuable addition to eradication strategies, when the approaches discussed in previous paragraphs are taken into account. To prevent the above-described problems with pathogen adaptation, it is important that there is control on the (genetic) composition of commercial herds and on (indirect) contacts of such herds with herds outside the eradication program, such that pathogens are

eradicated relatively rapidly on farm and do not re-emerge. If, for example, a pig farmer takes part in a chain-controlled eradication program for PRRS, but decides to buy genetic material from another breeding company that is not sufficiently PRRS resistant for eradication, this might not only lead to persistence of PRRS on the own farm, but might also open the invasion window for escape mutants of PRRSV, and thereby form a threat to the other farms in the program. The feasibility of such strong control on the genetic composition of production herds is highly species dependent. It will be high in aquaculture and poultry, where complete flocks typically consist of animals with the same genetic background, bought from (chain partners) of a single breeding company. To some extent this also holds for pigs, but definitely not for cattle, where farmers make more independent choices on which genetic material is brought into their herd.

7.7 Societal impact

The theory presented in this thesis shows that "correct" implementation of genetic selection for increased resistance to infectious diseases has large potential to reduce the impact of infectious diseases on the health and welfare of livestock. Furthermore, it will increase the economic efficiency of livestock farming, by reducing treatment costs and production losses. Importantly, the costs of infectious disease in livestock outweigh the costs of genetic improvement in production and reproduction traits multiple times (KNAP AND DOESCHL-WILSON 2020). An indirect benefit of improved genetic selection against infectious disease in livestock might be a decreased transmission of certain zoonotic infections to the human population.

Next to improved results of genetic selection in livestock populations, this thesis aims to contribute to better understanding of the indirect effects caused by infectious disease interventions, in particular through chapter 4. Such an improved understanding of indirect effects is of great relevance, also outside the livestock sector. In plants, genetic selection for resistance to infectious diseases is more widely applied than in animals, but hardly combined with knowledge from infectious disease epidemiology. Although the spread of infectious diseases in

plants might be more than in animals dependent on environmental factors (e.g. wind), indirect effects are relevant as well (e.g. SILVA *et al.* 2013). For example, a field of crops resistant to a certain infection might show very different incidence of that infection if neighbouring fields are planted with that same resistant variety, compared to when they are planted with a non-resistant variety of the same crop. Also in human medicine, indirect effects of infectious disease interventions are often not well understood or not optimally exploited. A clear example is the yearly vaccination campaign against influenza in many countries. In the Netherlands, for example, only elderly and other risk groups are eligible for vaccination, mainly to prevent severe clinical symptoms in those groups. However, multiple studies have shown that school-going children are the main source of transmission of influenza, and that applying the same number of vaccines to this group results in a larger, indirect, reduction of mortality and morbidity due to influenza in high-risk groups than the current strategy of vaccinating high-risk groups (LONGINI AND HALLORAN 2005; WORBY *et al.* 2015; EICHNER *et al.* 2017; TSANG *et al.* 2019).

7.8 Outlook

From the research presented in this thesis I conclude that genetic selection for decreased susceptibility to infectious disease has greater potential than always thought due to indirect effects that occur in transmission. These indirect effects are at least a factor of the inverse of the prevalence larger than the direct effect, showing that the additional reduction due to indirect effects increases rapidly with prevalence decreasing. Accurate and unbiased genetic parameters and breeding values for susceptibility can be obtained from both field and experimental data, as long as the observation interval is short enough. Pathogens might adapt to animals that are selected for increased resistance, but this can be prevented if the genetic change in animal resistance in the local population is fast and large enough, such that the pathogen has been eradicated before escape mutants of the pathogen emerged and invaded the animal population.

One of the remaining questions in relation to selection against infectious disease traits is the role of genetic variation in infectivity. As discussed before, evolutionary theory suggests that there might be (much) more genetic variation

in infectivity present compared to susceptibility, because the effect of infectivity on infection status of individuals is fully indirect. Furthermore, using this variation in genetic selection programs might considerably increase response to selection in infection prevalence (LIPSCHUTZ-POWELL *et al.* 2012; ANCHE *et al.* 2014; DOESCHL-WILSON *et al.* 2018; TSAIRIDOU *et al.* 2019; BIJMA *et al.* 2022). Research into estimation methods for genetic parameters and breeding values for infectivity, e.g. using GLMMs (ANCHE *et al.* 2015; BIEMANS *et al.* 2019a) or Bayesian inference (POOLEY *et al.* 2020; POOLEY *et al.* 2022b), and data requirements to get accurate and unbiased estimates out of such methods is therefore of great value. Results by POOLEY *et al.* (2022a), however, indicate that accurate estimation of infectivity, unlike susceptibility, requires specific designs with respect to group size and genetic composition, and seems therefore only feasible in experimental settings. However, genetic differences in infectivity might also affect accuracy and bias of estimates for susceptibility, in particular when these traits are correlated. Thus, two additional research questions in relation to infectivity might concern the correlation with susceptibility, and the effect on susceptibility estimation. Such insights are particularly important when collected data does not allow accurate estimation of infectivity, e.g., due to large group size, but can be used for estimation of susceptibility.

The results in this thesis are based on theoretical derivations, with limited empirical support. There are only very few empirical examples that compare expected to achieved selection response in infection prevalence (e.g. HERINGSTAD *et al.* 2007), despite that health traits have been part of breeding goals for several decades. Given that selection in a commercial setting typically results in a gradual response per trait, it might very well be that pathogens, by developing escape adaptations, have counteracted the selection effort. Nevertheless, this is still a hypothesis. Next to experimental work, research that compares genetic and phenotypic response in prevalence of certain infections in a variety of species might be of great value, both to confirm the large indirect effects found in this thesis and to get more insight into pathogen adaptation. For an empirical experiment to show the large indirect effects in selection for decreased susceptibility a single or a few generations of selection with a large selection differential in susceptibility (e.g., single trait selection) might be sufficient, as demonstrated in Chapter 2. An empirical experiment to show pathogen adaption,

however, requires many more generations of selection, with gradual improvement in susceptibility, and will therefore be more costly and less (or not) feasible. A potential alternative is to gradually decrease herd-level susceptibility by gradually adding lowly susceptible animals into a population with average susceptibility. Then, in the final stage of this gradual population, infection prevalence can be compared to a population that reached the same overall reduction in susceptibility in one step. In any case, a sufficient population size, either in a number of contact groups or as a 'commercial herd', is needed to make sure that the results are not affected by random dying out of the infection. Consequently, such experiments will be costly, both economic and in number of experimental animals, and will therefore not be easy to conduct. The results in this thesis however provide strong motivation for doing such experiments, showing that genetic selection of livestock for reduced prevalence of infectious disease has large potential and might make a considerable contribution to reducing the impact of infectious diseases in the livestock industry and beyond.

References

Albers, G. A. A., G. D. Gray, L. R. Piper, J. S. F. Barker, L. F. L. Jambre *et al.*, 1987 The genetics of resistance and resilience to Haemonchus contortus infection in young merino sheep. International Journal for Parasitology 17**:** 1355-1363.

Alizon, S., A. Hurford, N. Mideo and M. Van Baalen, 2009 Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. Journal of evolutionary biology 22**:** 245-259.

Anacleto, O., S. Cabaleiro, B. Villanueva, M. Saura, R. D. Houston *et al.*, 2019 Genetic differences in host infectivity affect disease spread and survival in epidemics. Scientific reports 9**:** 4924.

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

Anche, M. T., P. Bijma and M. C. M. De Jong, 2015 Genetic analysis of infectious diseases: estimating gene effects for susceptibility and infectivity. Genetics Selection Evolution 47**:** 85.

Anche, M. T., M. C. M. De Jong and P. Bijma, 2014 On the definition and utilization of heritable variation among hosts in reproduction ratio R 0 for infectious diseases. Heredity 113**:** 364.

AquaGen, 2012 Field results of QTL-innOva IPN – from egg to harvest, pp. in *AquaGen information letter*, Trondheim, Norway.

Bai, X., and G. S. Plastow, 2022 Breeding for disease resilience: opportunities to manage polymicrobial challenge and improve commercial performance in the pig industry. CABI Agriculture and Bioscience 3**:** 1-17.

Bai, X., A. M. Putz, Z. Wang, F. Fortin, J. C. Harding *et al.*, 2020 Exploring phenotypes for disease resilience in pigs using complete blood count data from a natural disease challenge model. Frontiers in Genetics 11**:** 216.

Bennett, R., 2003 The 'direct costs' of livestock disease: the development of a system of models for the analysis of 30 endemic livestock diseases in Great Britain. Journal of Agricultural Economics 54**:** 55-71.

Berghof, T. V., H. Bovenhuis and H. A. Mulder, 2019a Body weight deviations as indicator for resilience in layer chickens. Frontiers in genetics 10**:** 1216.

Berghof, T. V. L., M. Poppe and H. A. Mulder, 2019b Opportunities to Improve Resilience in Animal Breeding Programs. Frontiers in Genetics 9.

Biemans, F., P. Bijma, N. M. Boots and M. C. M. de Jong, 2018 Digital Dermatitis in dairy cattle: The contribution of different disease classes to transmission. Epidemics 23**:** 76-84.

Biemans, F., M. C. M. de Jong and P. Bijma, 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., M. C. M. de Jong and P. Bijma, 2019a Genetic parameters and genomic breeding values for digital dermatitis in Holstein Friesian dairy cattle: host susceptibility, infectivity and the basic reproduction ratio. Genetics Selection Evolution 51**:** 1-13.

Biemans, F., M. C. M. de Jong and P. Bijma, 2019b A genome-wide association study for susceptibility and infectivity of Holstein Friesian dairy cattle to digital dermatitis. Journal of Dairy Science 102**:** 6248-6262.

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

Bijma, P., 2011 A general definition of the heritable variation that determines the potential of a population to respond to selection. Genetics 189**:** 1347-1359.

Bijma, P., 2014 The quantitative genetics of indirect genetic effects: a selective review of modelling issues. Heredity 112**:** 61-69.

Bijma, P., 2020 The Price equation as a bridge between animal breeding and evolutionary biology. Philosophical Transactions of the Royal Society B 375**:** 20190360.

Bijma, P., A. D. Hulst and M. C. M. de Jong, 2022 The quantitative genetics of the prevalence of infectious diseases: hidden genetic variation due to indirect genetic effects dominates heritable variation and response to selection. Genetics 220.

Bijma, P., and J. Van Arendonk, 1998 Maximizing genetic gain for the sire line of a crossbreeding scheme utilizing both purebred and crossbred information. Animal Science 66**:** 529-542.

Bishop, S., and M. Stear, 1997 Modelling responses to selection for resistance to gastro-intestinal parasites in sheep. Animal Science 64**:** 469-478.

Bishop, S., and M. Stear, 1999 Genetic and epidemiological relationships between productivity and disease resistance: gastro-intestinal parasite infection in growing lambs. Animal Science 69**:** 515-524.

Bishop, S., and M. Stear, 2003 Modeling of host genetics and resistance to infectious diseases: understanding and controlling nematode infections. Veterinary parasitology 115**:** 147-166.

Bishop, S. C., R. F. Axford, F. W. Nicholas and J. B. Owen, 2010 *Breeding for disease resistance in farm animals*. CABI.

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

Bishop, S. C., and J. A. Woolliams, 2010 On the genetic interpretation of disease data. PloS one 5**:** e8940.

Bishop, S. C., and J. A. Woolliams, 2014 Genomics and disease resistance studies in livestock. Livestock science 166**:** 190-198.

Boehm, M., M. R. Hutchings and P. C. White, 2009 Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. PLoS One 4.

Chabas, H., S. Lion, A. Nicot, S. Meaden, S. van Houte *et al.*, 2018 Evolutionary emergence of infectious diseases in heterogeneous host populations. PLoS biology 16**:** e2006738.

Chae, C., 2021 Commercial PRRS modified-live virus vaccines. Vaccines 9**:** 185.

Clemens, J., R. Brenner, M. Rao, N. Tafari and C. Lowe, 1996 Evaluating new vaccines for developing countries: efficacy or effectiveness? Jama 275**:** 390-397.

Craig, J., and W. Muir, 1996 Group selection for adaptation to multiple-hen cages: beak-related mortality, feathering, and body weight responses. Poultry Science 75**:** 294-302.

Darwin, C., 1864 On the origin of the species by means of natural selection, or, The preservation of favoured races in the struggle for life, pp. D. Appleton and Company, New York.

Davidson, F. W., C. Ahlstrom, J. De Buck, H. G. Whitney and K. Tahlan, 2016 Examination of Mycobacterium avium subspecies paratuberculosis mixed genotype infections in dairy animals using a whole genome sequencing approach. PeerJ 4**:** e2793.

Davies, G., S. Genini, S. Bishop and E. Giuffra, 2009 An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. Animal 3**:** 415-436.

Davies, J., and D. Davies, 2010 Origins and evolution of antibiotic resistance. Microbiology and Molecular Biology Reviews 74**:** 417-433.

Davies, S. F. M., L. P. Joyner and S. B. Kendall, 1963 Coccidiosis. Coccidiosis.

De Jong, M. C. M., 1995 Mathematical modelling in veterinary epidemiology: why model building is important. Preventive veterinary medicine 25**:** 183-193.

De Jong, M. C. M., and T. G. Kimman, 1994 Experimental quantification of vaccine-induced reduction in virus transmission. Vaccine 12**:** 761-766.

De Villemereuil, P., H. Schielzeth, S. Nakagawa and M. Morrissey, 2016 General methods for evolutionary quantitative genetic inference from generalized mixed models. Genetics 204**:** 1281-1294.

Dempster, E. R., and I. M. Lerner, 1950 Heritability of threshold characters. Genetics 35**:** 212.

Diekmann, O., H. Heesterbeek and T. Britton, 2012 *Mathematical tools for understanding infectious disease dynamics*. Princeton University Press, Princeton, N.J etc.

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

Diekmann, O., J. A. P. Heesterbeek and J. A. Metz, 1990 On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of mathematical biology 28**:** 365-382.

Doeschl-Wilson, A., O. Anacleto, H. Nielsen, T. Karlsson-Drangsholt, M. Lillehammer *et al.*, 2018 New opportunities for genetic disease control: beyond disease resistance, pp. in *Proceedings of the World Congress on Genetics Applied to Livestock Production (Auckland)*.

Doeschl-Wilson, A., P. W. Knap, T. Opriessnig and S. J. More, 2021 Review: Livestock disease resilience: from individual to herd level. Animal 15**:** 100286.

Doeschl-Wilson, A., D. Lipschutz-Powell, O. Anacleto, G. Lough and A. Lengeling, 2014 New methods for capturing unidentified genetic variation underlying infectious disease in livestock populations, pp. in *Proceedings of 10th World Congress on Genetics Applied to Livestock Production, Vancouver, BC, Canada*.

Doeschl-Wilson, A. B., R. Davidson, J. Conington, T. Roughsedge, M. R. Hutchings *et al.*, 2011 Implications of host genetic variation on the risk and prevalence of infectious diseases transmitted through the environment. Genetics 188**:** 683-693.

Doeschl-Wilson, A. B., and I. Kyriazakis, 2012 Should we aim for genetic improvement in host resistance or tolerance to infectious pathogens? Frontiers in genetics 3**:** 272.

Doeschl-Wilson, A. B., B. Villanueva and I. Kyriazakis, 2012 The first step toward genetic selection for host tolerance to infectious pathogens: obtaining the tolerance phenotype through group estimates. Frontiers in genetics 3**:** 265.

Drolet, M., É. Bénard, M.-C. Boily, H. Ali, L. Baandrup *et al.*, 2015 Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. The Lancet Infectious Diseases 15**:** 565-580.

Dunkelberger, J., P. Mathur, E. Little, D. Hanson, S. Dee *et al.*, 2022 Evidence of naturally occurring genetic variation in host response to challenge with a highly pathogenic PRRSV strain, pp. 700-703 in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges*. Wageningen Academic Publishers.

Dushoff, J., and S. Levin, 1995 The effects of population heterogeneity on disease invasion. Mathematical Biosciences 128**:** 25-40.

Ebert, D., and P. D. Fields, 2020 Host–parasite co-evolution and its genomic signature. Nature Reviews Genetics 21**:** 754-768.

Eichner, M., M. Schwehm, L. Eichner and L. Gerlier, 2017 Direct and indirect effects of influenza vaccination. BMC infectious diseases 17**:** 1-8.

Elaswad, A., and R. Dunham, 2018 Disease reduction in aquaculture with genetic and genomic technology: current and future approaches. Reviews in Aquaculture 10**:** 876-898.

Elbers, A., J. Braamskamp, L. Dekkers, R. Voets, T. Duinhof *et al.*, 2000 Aujeszky's disease virus eradication campaign successfully heading for last stage in The Netherlands. Veterinary Quarterly 22**:** 103-107.

Elgersma, G. G., G. de Jong, R. van der Linde and H. A. Mulder, 2018 Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. Journal of Dairy Science 101**:** 1240-1250.

Epstein, S. L., and G. E. Price, 2010 Cross-protective immunity to influenza A viruses. Expert Review of Vaccines 9**:** 1325-1341.

España, G., Z. M. Cucunubá, H. Diaz, S. Cavany, N. Castañeda *et al.*, 2022 Designing school reopening in the COVID-19 pre-vaccination period in Bogotá, Colombia: A modeling study. PLOS Global Public Health 2**:** e0000467.

European Commission, 2020 Bovine and swine diseases situation 2019, pp. 5,12. European Commission.

Falconer, D. S., and T. F. C. MacKay, 1996 *Introduction to Quantitative Genetics*. Pearson Education Limited, Harlow, England.

Fatoba, A. J., and M. A. Adeleke, 2018 Diagnosis and control of chicken coccidiosis: a recent update. Journal of Parasitic Diseases 42**:** 483-493.

Fenner, F., and I. Marshall, 1957 A comparison of the virulence for European rabbits (Oryctolagus cuniculus) of strains of myxoma virus recovered in the field in Australia, Europe and America. Epidemiology & Infection 55**:** 149-191.

Fenner, F., and G. M. Woodroofe, 1953 The pathogenesis of infectious myxomatosis; the mechanism of infection and the immunological response in the European rabbit (Oryctolagus cuniculus). Br J Exp Pathol 34**:** 400-411.

Fenner, F., and G. M. Woodroofe, 1965 Changes in the virulence and antigenic structure of strains of myxoma virus recovered from Australian wild rabbits between 1950 and 1964. Australian Journal of Experimental Biology and Medical Science 43**:** 359-370.

Fernandes, A. F. A., J. R. R. Dórea and G. J. d. M. Rosa, 2020 Image analysis and computer vision applications in animal sciences: an overview. Frontiers in Veterinary Science 7**:** 551269.

Fine, P. E., 1993 Herd immunity: history, theory, practice. Epidemiologic reviews 15**:** 265-302.

Fisher, R. A., 1918 The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh 53**:** 399-433.

Frank, S. A., 1991 Ecological and genetic models of host-pathogen coevolution. Heredity 67**:** 73-83.

Galvis, J. A., C. A. Corzo, J. M. Prada and G. Machado, 2022 Modelling the transmission and vaccination strategy for porcine reproductive and respiratory syndrome virus. Transboundary and Emerging Diseases 69**:** 485-500.

Gandon, S., and T. Day, 2007 The evolutionary epidemiology of vaccination. Journal of the Royal Society Interface 4**:** 803-817.

Ganusov, V. V., C. T. Bergstrom and R. Antia, 2002 Within-host population dynamics and the evolution of microparasites in a heterogeneous host population. Evolution 56**:** 213-223.

Gardner, A., 2020 Price's equation made clear. Philosophical Transactions of the Royal Society B 375**:** 20190361.

Gianola, D., 1982 Theory and Analysis of Threshold Characters. Journal of Animal Science 54**:** 1079-1096.

Gillespie, D. T., 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of computational physics 22**:** 403-434.

Gilmour, A., B. Gogel, B. Cullis, S. Welham and R. Thompson, 2015 ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd.

Greenwood, J., and G. Yule, 1915 The Statistics of Anti-Typhoid and Anti-Cholera Inoculations and the Interpretation of such Statistics in General. Proceedings of the Royal Society of Medicine 8.

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.

Guy, S. Z. Y., P. C. Thomson and S. Hermesch, 2012 Selection of pigs for improved coping with health and environmental challenges: breeding for resistance or tolerance? Frontiers in genetics 3**:** 281.

Halloran, M. E., M. Haber, I. M. Longini, Jr and C. J. Struchiner, 1991 Direct and Indirect Effects in Vaccine Efficacy and Effectiveness. American Journal of Epidemiology 133**:** 323-331.

Halloran, M. E., M. Haber and I. M. Longini, Jr., 1992 Interpretation and Estimation of Vaccine Efficacy under Heterogeneity. American Journal of Epidemiology 136**:** 328-343.

Halloran, M. E., I. M. Longini Jr and C. J. Struchiner, 1999 Design and interpretation of vaccine field studies. Epidemiologic reviews 21**:** 73-88.

Havenstein, G., P. Ferket and M. Qureshi, 2003 Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. Poultry science 82**:** 1500-1508.

Hazel, L. N., 1943 The genetic basis for constructing selection indexes. Genetics 28**:** 476-490.

Hazel, L. N., G. E. Dickerson and A. E. Freeman, 1994 The selection index—then, now, and for the future. Journal of dairy science 77**:** 3236-3251.

Henderson, D. A., 1976 The eradication of smallpox. Scientific American 235**:** 25-33.

Henderson, D. A., 2011 The eradication of smallpox–an overview of the past, present, and future. Vaccine 29**:** D7-D9.

Henryon, M., P. M. Heegaard, J. Nielsen, P. Berg and H. R. Juul-Madsen, 2006 Immunological traits have the potential to improve selection of pigs for resistance to clinical and subclinical disease. Animal Science 82**:** 597-606.

Heringstad, B., G. Klemetsdal and T. Steine, 2007 Selection responses for disease resistance in two selection experiments with Norwegian red cows. Journal of dairy science 90**:** 2419-2426.

Hethcote, H. W., 1989 Three basic epidemiological models, pp. 119-144 in *Applied mathematical ecology*. Springer.

Hillestad, B., S. Johannessen, G. O. Melingen and H. K. Moghadam, 2021 Identification of a New Infectious Pancreatic Necrosis Virus (IPNV) Variant in Atlantic Salmon (Salmo salar L.) that can Cause High Mortality Even in Genetically Resistant Fish. Front Genet 12.

Hoes, J., P. J. Woestenberg, J. A. Bogaards, A. J. King, H. E. de Melker *et al.*, 2021 Population Impact of Girls-Only Human Papillomavirus 16/18 Vaccination in The Netherlands: Cross-Protective and Second-Order Herd Effects. Clinical Infectious Diseases 72**:** e103-e111.

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

Houle, D., 1992 Comparing evolvability and variability of quantitative traits. Genetics 130**:** 195-204.

Hulst, A. D., M. C. M. de Jong and P. Bijma, 2021 Why genetic selection to reduce the prevalence of infectious diseases is way more promising than currently believed. Genetics 217.

Jenner, E., 1800 *An inquiry into the causes and effects of the variolae vaccinae: a disease discovered in some of the western counties of England, particularly Gloucestershire, and known by the name of the cow pox*. printed, for the author, by Sampson Low...; and sold by Law... and Murray ….

Karlsson, E. K., D. P. Kwiatkowski and P. C. Sabeti, 2014 Natural selection and infectious disease in human populations. Nature Reviews Genetics 15**:** 379-393.

Keeling, M. J., and J. V. Ross, 2008 On methods for studying stochastic disease dynamics. Journal of the Royal Society Interface 5**:** 171-181.

Kemper, K., R. Elwin, S. Bishop, M. Goddard and R. Woolaston, 2009 Haemonchus contortus and Trichostrongylus colubriformis did not adapt to long-term exposure to sheep that were genetically resistant or susceptible to nematode infections. International journal for parasitology 39**:** 607-614.

Kemper, K. E., M. E. Goddard and S. C. Bishop, 2013 Adaptation of gastrointestinal nematode parasites to host genotype: single locus simulation models. Genetics Selection Evolution 45**:** 14.

Kennedy, D. A., and A. F. Read, 2017 Why does drug resistance readily evolve but vaccine resistance does not? Proceedings of the Royal Society B: Biological Sciences 284**:** 20162562.

Khachatourians, G. G., 1998 Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. Cmaj 159**:** 1129-1136.

Killeen, G. F., T. A. Smith, H. M. Ferguson, H. Mshinda, S. Abdulla *et al.*, 2007 Preventing Childhood Malaria in Africa by Protecting Adults from Mosquitoes with Insecticide-Treated Nets. PLOS Medicine 4**:** e229.

Klaas, I. C., and R. N. Zadoks, 2018 An update on environmental mastitis: Challenging perceptions. Transboundary and emerging diseases 65**:** 166-185.

Knap, P. W., and S. C. Bishop, 2000 Relationships between genetic change and infectious disease in domestic livestock. BSAP Occasional Publication 27**:** 65-80.

Knap, P. W., and A. Doeschl-Wilson, 2020 Why breed disease-resilient livestock, and how? Genetics Selection Evolution 52**:** 60.

Koskella, B., 2018 Resistance gained, resistance lost: An explanation for host-parasite coexistence. PLoS biology 16**:** e3000013.

Kraaijeveld, A. R., and H. C. J. Godfray, 1997 Trade-off between parasitoid resistance and larval competitive ability in Drosophila melanogaster. Nature 389**:** 278-280.

Lipschutz-Powell, D., J. A. Woolliams, P. Bijma and A. B. Doeschl-Wilson, 2012 Indirect genetic effects and the spread of infectious disease: are we capturing the full heritable variation underlying disease prevalence? PloS one 7**:** e39551.

Lipschutz-Powell, D., J. A. Woolliams and A. B. Doeschl-Wilson, 2014 A unifying theory for genetic epidemiological analysis of binary disease data. Genetics Selection Evolution 46**:** 15.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp and W. M. Getz, 2005 Superspreading and the effect of individual variation on disease emergence. Nature 438**:** 355.

Longini, I. M., Jr., and M. E. Halloran, 2005 Strategy for distribution of influenza vaccine to high-risk groups and children. Am J Epidemiol 161**:** 303-306.

Ma, J., and D. J. Earn, 2006 Generality of the final size formula for an epidemic of a newly invading infectious disease. Bulletin of mathematical biology 68**:** 679-702.

MacKenzie, K., and S. Bishop, 2001 Utilizing stochastic genetic epidemiological models to quantify the impact of selection for resistance to infectious diseases in domestic livestock. Journal of animal science 79**:** 2057-2065.

Magori, K., and A. W. Park, 2014 The evolutionary consequences of alternative types of imperfect vaccines. Journal of mathematical biology 68**:** 969-987.

Man, I., S. Vänskä, M. Lehtinen and J. A. Bogaards, 2021 Human papillomavirus genotype replacement: still too early to tell? The Journal of infectious diseases 224**:** 481-491.

Mariner, J. C., J. A. House, C. A. Mebus, A. E. Sollod, D. Chibeu *et al.*, 2012 Rinderpest eradication: appropriate technology and social innovations. Science 337**:** 1309-1312.

Martin, P., H. Barkema, L. Brito, S. Narayana and F. Miglior, 2018 Symposium review: Novel strategies to genetically improve mastitis resistance in dairy cattle. Journal of dairy science 101**:** 2724-2736.

May, R. M., and R. M. Anderson, 1990 Parasite—host coevolution. Parasitology 100**:** S89-S101.

May, R. M., R. M. Anderson, W. F. Bodmer and J. F. C. Kingman, 1983 Epidemiology and genetics in the coevolution of parasites and hosts. Proceedings of the Royal Society of London. Series B. Biological Sciences 219**:** 281-313.

McCullagh, P., 2018 *Generalized linear models*. Routledge.

McDonald, B. A., and C. Linde, 2002 The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica 124**:** 163-180.

McLean, A. R., 1995 Vaccination, evolution and changes in the efficacy of vaccines: a theoretical framework. Proceedings of the Royal Society of London. Series B: Biological Sciences 261**:** 389-393.

Mendel, G., 1866 *Versuche über Pflanzen-Hybriden*. Im Verlage des Vereines, Brünn.

Miller, I. F., and C. J. Metcalf, 2019 Vaccine-driven virulence evolution: consequences of unbalanced reductions in mortality and transmission and implications for pertussis vaccines. J R Soc Interface 16**:** 20190642.

Moen, T., M. Baranski, A. K. Sonesson and S. Kjøglum, 2009 Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. BMC Genomics 10**:** 368.

Moore, A. J., E. D. Brodie III and J. B. Wolf, 1997 Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution 51**:** 1352-1362.

More, S. J., B. Radunz and R. Glanville, 2015 Lessons learned during the successful eradication of bovine tuberculosis from Australia. Veterinary Record 177**:** 224-232.

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

Mundt, C. C., 2018 Pyramiding for resistance durability: Theory and practice. Phytopathology® 108**:** 792-802.

Nakagawa, S., P. C. Johnson and H. Schielzeth, 2017 The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. Journal of the Royal Society Interface 14**:** 20170213.

Neeteson-van Nieuwenhoven, A.-M., P. Knap and S. Avendaño, 2013 The role of sustainable commercial pig and poultry breeding for food security. Animal Frontiers 3**:** 52-57.

Neumann, E. J., J. B. Kliebenstein, C. D. Johnson, J. W. Mabry, E. J. Bush *et al.*, 2005 Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. Journal of the American Veterinary Medical Association 227**:** 385-392.

Niederwerder, M. C., C. J. Jaing, J. B. Thissen, A. G. Cino-Ozuna, K. S. McLoughlin *et al.*, 2016 Microbiome associations in pigs with the best and worst clinical outcomes following co-infection with porcine reproductive and respiratory syndrome virus (PRRSV) and porcine circovirus type 2 (PCV2). Veterinary Microbiology 188**:** 1-11.

Nieuwhof, G. J., J. Conington and S. C. Bishop, 2009 A genetic epidemiological model to describe resistance to an endemic bacterial disease in livestock: application to footrot in sheep. Genetics Selection Evolution 41**:** 19.

Nigsch, A., S. Robbe-Austerman, T. P. Stuber, P. D. Pavinski Bitar, Y. T. Gröhn *et al.*, 2021 Who infects whom?—Reconstructing infection chains of Mycobacterium avium ssp. paratuberculosis in an endemically infected dairy herd by use of genomic data. PLOS ONE 16**:** e0246983.

Olofsson, S. K., and O. Cars, 2007 Optimizing drug exposure to minimize selection of antibiotic resistance. Clinical Infectious Diseases 45**:** S129-S136.

Orenstein, W. A., R. H. Bernier, T. J. Dondero, A. R. Hinman, J. S. Marks *et al.*, 1985 Field evaluation of vaccine efficacy. Bulletin of the World Health Organization 63**:** 1055.

Orenstein, W. A., R. H. Bernier and A. R. Hinman, 1988 Assessing vaccine efficacy in the field: further observations. Epidemiologic reviews 10**:** 212-241.

Orsel, K., P. Plummer, J. Shearer, J. De Buck, S. D. Carter *et al.*, 2018 Missing pieces of the puzzle to effectively control digital dermatitis. Transboundary and emerging diseases 65**:** 186-198.

Pérez-Cabal, M. A., G. de los Campos, A. I. Vazquez, D. Gianola, G. J. M. Rosa *et al.*, 2009 Genetic evaluation of susceptibility to clinical mastitis in Spanish Holstein cows. Journal of Dairy Science 92**:** 3472-3480.

Petersen, G. E. L., J. B. Buntjer, F. S. Hely, T. J. Byrne and A. Doeschl-Wilson, 2022 Modeling suggests gene editing combined with vaccination could eliminate a persistent disease in livestock. Proceedings of the National Academy of Sciences 119**:** e2107224119.

Pooley, C., G. Marion, S. Bishop and A. Doeschl-Wilson, 2022a Optimal experimental designs for estimating genetic and non-genetic effects underlying infectious disease transmission. Genetics Selection Evolution 54**:** 59.

Pooley, C., G. Marion and A. Doeschl-Wilson, 2022b SIRE 2.0: a novel software tool to estimate host genetic variation in infectious disease transmission, pp. 1475-1478 in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges*. Wageningen Academic Publishers.

Pooley, C. M., G. Marion, S. C. Bishop, R. I. Bailey and A. B. Doeschl-Wilson, 2020 Estimating individuals' genetic and non-genetic effects underlying infectious disease transmission from temporal epidemic data. PLOS Computational Biology 16**:** e1008447.

Poppe, M., G. Bonekamp, M. L. van Pelt and H. A. Mulder, 2021 Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages. Journal of Dairy Science 104**:** 1967-1981.

Porta, M. S., S. Greenland, M. Hernán, I. d. S. Silva, J. M. Last *et al.*, 2014 A dictionary of epidemiology, pp. Oxford University Press, Oxford.

Prentice, J., C. Pooley, S. Tsairidou, R. P. Wong, O. Anacleto *et al.*, 2022 Transmission experiment in turbot shows high genetic variation in host infectivity affecting disease spread and survival, pp. 728-731 in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges*. Wageningen Academic Publishers.

Price, G. R., 1970 Selection and covariance. Nature 227**:** 520-521.

Price, P. W., 1980 *Evolutionary biology of parasites*. Princeton University Press.

Putz, A. M., J. C. Harding, M. K. Dyck, F. Fortin, G. S. Plastow *et al.*, 2019 Novel resilience phenotypes using feed intake data from a natural disease challenge model in wean-to-finish pigs. Frontiers in genetics 9**:** 660.

Quammen, D., 1996 *The song of the dodo : island biogeography in an age of extinctions*. Hutchinson, London.

R Core Team, 2020 R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Råberg, L., A. L. Graham and A. F. Read, 2009 Decomposing health: tolerance and resistance to parasites in animals. Philosophical Transactions of the Royal Society B: Biological Sciences 364**:** 37-49.

Raberg, L., D. Sim and A. F. Read, 2007 Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. Science 318**:** 812-814.

Raphaka, K., E. Sánchez-Molano, S. Tsairidou, O. Anacleto, E. J. Glass *et al.*, 2018 Impact of genetic selection for increased cattle resistance to bovine tuberculosis on disease transmission dynamics. Frontiers in veterinary science 5**:** 237.

Read, A. F., S. J. Baigent, C. Powers, L. B. Kgosana, L. Blackwell *et al.*, 2015 Imperfect vaccination can enhance the transmission of highly virulent pathogens. PLoS biology 13**:** e1002198.

Regoes, R. R., M. A. Nowak and S. Bonhoeffer, 2000 Evolution of virulence in a heterogeneous host population. Evolution 54**:** 64-71.

REX Consortium, 2013 Heterogeneity of selection and the evolution of resistance. Trends in Ecology & Evolution 28**:** 110-118.

Robertson, A., 1950 Proof that the additive heritability on the P scale is given by the expression Z2hx2/pq., pp. 212 in *Genetics*.

Robinson, N. A., D. Robledo, L. Sveen, R. R. Daniels, A. Krasnov *et al.*, 2023 Applying genetic technologies to combat infectious diseases in aquaculture. Reviews in Aquaculture 15**:** 491-535.

Roeder, P., J. Mariner and R. Kock, 2013 Rinderpest: the veterinary perspective on eradication. Philosophical Transactions of the Royal Society B: Biological Sciences 368**:** 20120139.

Rupp, R., and G. Foucras, 2010 Genetics of mastitis in dairy ruminants. Breeding for disease resistance in farm animals 183**:** 212.

Rushton, J., 2009 *The economics of animal health and production*. Cabi.

Salo, J., M. Hägg, M. Kortelainen, T. Leino, T. Saxell *et al.*, 2022 The indirect effect of mRNA-based COVID-19 vaccination on healthcare workers' unvaccinated household members. Nature Communications 13**:** 1162.

Sandegren, L., 2014 Selection of antibiotic resistance at very low antibiotic concentrations. Upsala journal of medical sciences 119**:** 103-107.

Silva, J. C. e., B. M. Potts, P. Bijma, R. J. Kerr and D. J. Pilbeam, 2013 Genetic control of interactions among individuals: contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. New Phytologist 197**:** 631-641.

Speksnijder, D., D. Mevius, C. Bruschke and J. Wagenaar, 2015 Reduction of veterinary antimicrobial use in the Netherlands. The Dutch success model. Zoonoses and public health 62**:** 79-87.

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

Stegeman, A., M. C. M. De Jong, A. Van Nes and A. Bouma, 1997 Dynamics of pseudorabies virus infections in vaccinated pig populations: a review. Veterinary quarterly 19**:** 117-122.

Stoesser, N., A. E. Sheppard, C. E. Moore, T. Golubchik, C. M. Parry *et al.*, 2015 Extensive within-host diversity in fecally carried extended-spectrum-beta-lactamase-producing Escherichia coli isolates: Implications for transmission analyses. Journal of Clinical Microbiology 53**:** 2122-2131.

Thompson-Crispi, K., A. Sewalem, F. Miglior and B. Mallard, 2012 Genetic parameters of adaptive immune response traits in Canadian Holsteins. Journal of dairy science 95**:** 401-409.

Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay and C. Maltecca, 2015 A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PloS one 10**:** e0114919-e0114919.

Tsairidou, S., A. Allen, G. Banos, M. Coffey, O. Anacleto *et al.*, 2018 Can we breed cattle for lower bovine TB infectivity? Frontiers in veterinary science 5**:** 310.

Tsairidou, S., O. Anacleto, J. Woolliams and A. Doeschl-Wilson, 2019 Enhancing genetic disease control by selecting for lower host infectivity and susceptibility. Heredity 122**:** 742.

Tsang, T. K., V. J. Fang, D. K. M. Ip, R. A. P. M. Perera, H. C. So *et al.*, 2019 Indirect protection from vaccinating children against influenza in households. Nature Communications 10**:** 106.

van Baalen, M., and M. W. Sabelis, 1995 The dynamics of multiple infection and the evolution of virulence. The American Naturalist 146**:** 881-910.

Van der Goot, J. A., G. D. Koch, M. C. M. De Jong and M. Van Boven, 2005 Quantification of the effect of vaccination on transmission of avian influenza (H7N7) in chickens. Proceedings of the National Academy of Sciences 102**:** 18141-18146.

Vanderplank, J. E., 1984 *Disease resistance in plants*. Elsevier.

Velthuis, A. G. J., M. C. M. De Jong, E. M. Kamp, N. Stockhofe and J. H. M. Verheijden, 2003 Design and analysis of an Actinobacillus pleuropneumoniae transmission experiment. Preventive Veterinary Medicine 60**:** 53-68.

Vleeshouwers, V. G., S. Raffaele, J. H. Vossen, N. Champouret, R. Oliva *et al.*, 2011 Understanding and exploiting late blight resistance in the age of effectors. Annual review of phytopathology 49**:** 507-531.

Vu, H. L. X., A. K. Pattnaik and F. A. Osorio, 2017 Strategies to broaden the cross-protective efficacy of vaccines against porcine reproductive and respiratory syndrome virus. Veterinary Microbiology 206**:** 29-34.

Wallace, A. R., 1869 *The Malay Archipelago : the land of the orang-utan and the bird of paradise : a narrative of travel, with studies of man and nature*. Macmillan, London.

Walsh, C., 2003 Where will new antibiotics come from? Nature Reviews Microbiology 1**:** 65-70.

Weinberg, G. A., and P. G. Szilagyi, 2010 Vaccine epidemiology: efficacy, effectiveness, and the translational research roadmap. The Journal of infectious diseases 201**:** 1607-1610.

Weiss, G. H., and M. Dishon, 1971 On the asymptotic behavior of the stochastic and deterministic models of an epidemic. Mathematical Biosciences 11**:** 261-265.

Welderufael, B. G., L. L. G. Janss, D. J. de Koning, L. P. Sørensen, P. Løvendahl *et al.*, 2017 Bivariate threshold models for genetic evaluation of susceptibility to and ability to recover from mastitis in Danish Holstein cows. Journal of Dairy Science 100**:** 4706-4720.

White, L. J., T. J. G. M. Lam, Y. H. Schukken, L. E. Green, G. F. Medley *et al.*, 2006 The transmission and control of mastitis in dairy cows: A theoretical approach. Preventive Veterinary Medicine 74**:** 67-83.

White, L. J., Y. H. Schukken, T. J. G. M. Lam, G. F. Medley and M. J. Chappell, 2001 A multispecies model for the transmission and control of mastitis in dairy cows. Epidemiology and Infection 127**:** 567-576.

Worby, C. J., S. S. Chaves, J. Wallinga, M. Lipsitch, L. Finelli *et al.*, 2015 On the relative role of different age groups in influenza epidemics. Epidemics 13**:** 10-16.

World Health Organization, 2023 Vaccines and immunization

Yates, A., R. Antia and R. R. Regoes, 2006 How do pathogen evolution and host heterogeneity interact in disease emergence? Proceedings of the Royal Society B: Biological Sciences 273**:** 3075-3083.

Young, S., 1961 A further examination of the relative efficiency of three methods of selection for genetic gains under less-restricted conditions. Genetics Research 2**:** 106-121.

Summary

Infectious diseases strongly affect the health and welfare of farmed animals and have large economic impact in the livestock sector. A variety of interventions is applied to combat infectious diseases in livestock, one of which is genetic selection. However, the potential of genetic selection to decrease the prevalence of an infectious disease is typically seen as limited, because of the low heritability of infectious disease traits. However, the quantitative genetic models used to estimate this heritability often ignore the transmission dynamics of infectious diseases. Previous work has shown that incorporating these epidemiological dynamics into prediction models considerably changes our ideas about the prospects of genetic selection against infectious diseases.

Another general problem with infectious disease interventions is that they exert selection pressure on the pathogen population, which in response might evolve adaptations to escape the intervention. A prominent example of such pathogen adaptation is the widespread occurrence of antibiotic resistant bacteria. For several interventions, strategies have been developed to reduce the risk of pathogen evolution, e.g., by combining several antiviral drugs in one therapy or by restricting the use of new antibiotics. Adaptation of pathogens to genetically selected livestock, however, has received little attention.

The main aim of this thesis is to improve our understanding of genetic selection against infectious diseases by integration of theory and methods from animal breeding and quantitative epidemiology. In Chapter 2, the potential of genetic selection to reduce the prevalence of infectious diseases is explored using simulations of an endemic infection in a population of animals with genetic variation in susceptibility to that infection. The size of the genetic variance in susceptibility was iteratively tuned such that analysis of simulated individual infection status with a linear mixed model resulted in commonly observed heritability estimates, between 0.02 and 0.1. Then, the potential response to selection was determined using sire selection. The results indicate that the common heritability estimates for infection status correspond to considerable genetic variation in susceptibility. Furthermore, in contrast to expectations from classical quantitative genetic models, eradication of the infection often occurred after one round of sire selection. This finding indicates the presence of substantial

indirect genetic effects of selection on susceptibility, resulting in a basic reproduction ratio (R_0) below 1.

The theory behind the findings in Chapter 2 is explored in Chapter 3. There, expressions for response to selection in prevalence and for the size of the indirect genetic effect of susceptibility are identified. The total breeding value for prevalence, which incorporates the indirect effect due to susceptibility and recovery, is a factor of the inverse of the endemic prevalence greater than the common breeding value for binary disease status (the direct effect). This finding indicates that the total response in prevalence increases with decreasing endemic prevalence. For example, for an endemic prevalence below a half, the total response is more than double the response expected based on the breeding values for binary disease status.

The ratio of total to direct effect found in Chapter 3 is translated beyond the field of animal breeding in Chapter 4. Here, also the mechanisms behind the large response are further explored. It turns out that for any intervention targeting infections, the inverse of the endemic prevalence provides a prediction of the lower bound of the ratio of overall to direct effect of that intervention. Thus also, for example, for vaccination. Thereby, the inverse of the endemic prevalence, either measured directly or calculated from R_0, provides a relatively easy way to predict the overall effect of an intervention once the (direct) efficacy of that intervention, as commonly estimated in vaccine trials, is known.

Chapter 5 explores the adaptation of the pathogen population to hosts that are genetically selected for resistance to infection with that pathogen, by developing a model for transmission of evolving pathogens following genetic selection of animals for increased resistance. In this model, the prevalence of the infection is reduced by genetic selection of the hosts for resistance, eventually leading to eradication of the infection ($R_0 < 1$). As long as the infection is not eradicated, the model allows escape mutants of the pathogen to emerge and spread through the host population. With this transmission model, the 'invasion window' is defined, the range in frequency of resistant hosts in which escape mutants of the pathogen can emerge and invade the host population. The size of the invasion window is affected by the strength of host resistance, the benefit of the escape mutation for the pathogen in resistant hosts and the costs of the mutation in non-resistant

hosts. Stronger host resistance, for example, reduces the width of the invasion window. To minimise opportunities for pathogens to escape, disease control through genetic selection should aim to keep (local) populations out of the invasion window. A possible strategy would be to compose herds either fully of non-resistant animals (continued wild-type infection), or fully of highly resistant animals (such that $R_0 < 1$). The results indicate that a gradual increase in resistance provides opportunity for escape mutants to invade, and thus that multi-trait selection, the common approach in animal breeding, is not a sustainable strategy when resistance to infectious disease is part of the breeding goal.

Chapter 6 investigates statistical methods and data structures required for the estimation of genetic parameters and breeding values for infectious disease susceptibility, using simulation of epidemics. For estimation, a generalised linear mixed model (GLMM) that corrects for exposure of susceptible to infectious individuals is used. The impact of sampling interval, population structure, infection characteristics and model formulation on bias and accuracy is assessed. The results show that a GLMM can produce accurate and unbiased estimates of genetic variance for susceptibility to infection, as long as the observation interval of individual infection status is short. Other factors, such as group size, have limited effect.

In the general discussion (Chapter 7), the implications of the results in this thesis for the animal breeding industry are discussed. The discussion elaborates on the difference in perspective between animal breeding and epidemiology, the indirect genetic effects that occur in the transmission of infectious diseases, the availability of genetic variation in infection traits, the traits that could be selected for to reduce the impact of infectious diseases, the phenotypes needed to select for those traits, how these phenotypes can be collected, and requirements for implementation. The conclusions are that selection for increased tolerance of infection provides no safe alternative to selection for reduced susceptibility, because pathogen adaptation can also occur because of tolerant hosts. Also, developments in automated phenotyping give opportunity to obtain the longitudinal infection data needed for accurate estimation of susceptibility and potentially also infectivity. Where testing is possible to obtain phenotypes, transmission experiments should be performed instead of challenge tests. Finally, for effective and sustainable

selection against infectious diseases, breeding programs should adopt the epidemiological approach and aim for eradication of infections, for example stepwise with one local population at a time, instead of gradual reduction in the population-wide infection prevalence. Such a shift requires control on the genetic composition of commercial livestock herds and might therefore not be equally feasible in all livestock species.

Samenvatting

Infectieziekten hebben niet alleen een groot effect op de gezondheid en het welzijn van landbouwhuisdieren, zij zetten ook de winstgevendheid van de veehouderijsector onder druk. Om infectieziekten in landbouwhuisdieren tegen te gaan wordt een verscheidenheid aan maatregelen toegepast, waaronder genetische selectie. Algemeen wordt echter aangenomen dat genetische selectie slechts beperkt in staat is om de prevalentie van een infectieziekte te verminderen, vanwege de lage erfelijkheidsgraad van infectieziektekenmerken. Hierbij valt de kanttekening te plaatsen dat de kwantitatief genetische modellen die gebruikt worden om deze erfelijkheidsgraad te schatten volledig voorbijgaan aan de transmissie van infectieziekten. Juist de effecten van deze transmissie op modelvoorspellingen leiden tot grote verandering in het denken over het effect van genetische selectie tegen infectieziekten.

Een ander, meer algemeen, probleem is dat maatregelen tegen infectieziekten selectiedruk uitoefenen op populaties van ziekteverwekkers. Deze selectiedruk kan leiden tot de evolutie van resistentie tegen de maatregel. Het algemeen voorkomen van antibioticaresistente bacteriën is hiervan een voor de hand liggend voorbeeld. Voor sommige infectieziektemaatregelen zijn strategieën bedacht om dergelijke aanpassing van de ziekteverwekker te voorkomen, bijvoorbeeld door een aantal antivirale middelen met verschil in werking te combineren of door het gebruik van nieuwe antibiotica zo veel mogelijk te beperken. Voor evolutie van resistente ziekteverwekkers tegen genetisch geselecteerde landbouwhuisdieren is tot nu toe echter weinig aandacht geweest.

Het doel van deze dissertatie is om genetische selectie tegen infectieziekten beter te begrijpen. Hiervoor worden theorie en modellen uit het fokkerijveld en het epidemiologisch veld gecombineerd. Hoofdstuk 2 onderzoekt het theoretisch potentieel van genetische selectie om de prevalentie van een infectieziekte te verminderen. Hiervoor wordt een endemische infectieziekte gesimuleerd in een populatie van dieren die genetisch verschillen in hun gevoeligheid voor besmetting met de infectie. De simulaties laten zien dat een grote variantie in gevoeligheid ten grondslag ligt aan de typische erfelijkheidsgraden voor ziektestatus. Onverwacht, want compleet anders dan de voorspelling van het algemeen kwantitatief genetisch model, leidde één ronde van vaderselectie meestal direct tot uitroeiing van de infectie. Dit komt doordat dieren die niet geïnfecteerd raken

ook geen andere dieren kunnen besmetten, wat uiteindelijk resulteert in een reproductiegetal (R-waarde) kleiner dan 1.

In hoofdstuk 3 wordt de theorie achter de resultaten van hoofdstuk 2 uitgediept. Hiervoor worden vergelijkingen afgeleid voor selectierespons in prevalentie en voor de grootte van het indirecte genetische effect. Deze afleidingen laten zien dat totale selectierespons in prevalentie toeneemt als de prevalentie van de infectie afneemt. Als de endemische prevalentie bijvoorbeeld lager is dan een half (0.5), is de totale selectierespons meer dan twee keer de selectierespons die men verwacht op basis van de fokwaarde voor binaire infectiestatus.

Deze ratio van totaal tegen direct effect wordt in hoofdstuk 4 in een algemenere context gebracht. Ook worden in dit hoofdstuk de onderliggende mechanismen verder onderzocht. Het blijkt dat de inverse van de endemische prevalentie de algemene ondergrens van de ratio van totaal tegen direct effect van een infectiemaatregel voorspelt. Dit geldt voor iedere maatregel gericht tegen infectie, dus ook voor bijvoorbeeld vaccinatie. De inverse van de endemische prevalentie, ofwel direct gemeten ofwel afgeleid vanuit een geschatte R_0, is daarmee een toegankelijke manier om het totale effect van een infectieziektemaatregel te voorspellen zodra het directe effect van die maatregel bekend is.

In hoofdstuk 5 wordt de evolutie van resistentie in een populatie van ziekteverwekkers tegen gastheren die genetisch geselecteerd zijn voor verminderde gevoeligheid voor die ziekteverwekkers onderzocht door middel van een transmissiemodel dat deze verschillende populaties en hun evolutie beschrijft. In het model daalt de prevalentie van de infectie ten gevolge van genetische selectie van de gastheer (het dier), uiteindelijk leidend tot uitroeiing van de infectie ($R_0 < 1$). Zolang de infectie nog niet is uitgeroeid, kunnen mutanten van de ziekteverwekker ontstaan en mogelijk spreiden in de gastheerpopulatie. Met dit transmissiemodel wordt een zogenaamd invasievenster afgeleid. Dit is het venster in frequentie van genetisch geselecteerde gastheren waarin er kans is op ontstaan en spreiden van resistente ziekteverwekkers. De grootte van het invasievenster wordt bepaald door de afname in gevoeligheid van de gastheer, het voordeel van de mutant t.o.v. de oorspronkelijke ziekteverwekker in de minder gevoelige gastheer en het negatieve effect (de kosten) van de mutatie bij infectie van een volledig gevoelige gastheer. Een grotere afname van

gastheergevoeligheid leidt bijvoorbeeld tot een kleiner invasievenster. Om resistentieontwikkeling zo veel mogelijk tegen te gaan, moeten (lokale) gastheerpopulaties zo veel mogelijk buiten het invasievenster worden gehouden. Dit zou bijvoorbeeld kunnen door een koppel dieren samen te stellen op basis van hun genetische weerstand tegen infectie, zodat ofwel het wildtype van de ziekteverwekker dominant blijft, ofwel de infectie uitgeroeid wordt zodat mutanten niet kunnen ontstaan. De resultaten laten ook zien dat resistentiemutanten alle kans hebben om zich te verspreiden bij een langzame afname in gevoeligheid. Een strategie zoals gebruikelijk in de fokkerij, met selectie op veel verschillende genetische kenmerken en langzame verandering per kenmerk, leidt dus waarschijnlijk tot resistentiemutanten wanneer een infectieziekte onderdeel uitmaakt van het fokdoel.

Hoofdstuk 6 onderzoekt statische methodes en datastructuren die nodig zijn om genetische parameters en fokwaardes voor gevoeligheid voor een infectieziekte te schatten, daarbij gebruik makend van simulaties van een epidemische infectieziekte. Voor schatting wordt een statistisch model gebruikt dat corrigeert voor de blootstelling van gevoelige aan infectieuze individuen. Hiermee wordt het effect van observatie-interval, populatiestructuur, infectieziekteparameters en modelformulering op onzuiverheid en nauwkeurigheid van de schatter geëvalueerd. De resultaten laten zien dat een lang tijdsinterval tussen observaties leidt tot slechtere schattingen, terwijl andere factoren, zoals groepsgrootte, weinig uitmaken.

In de algemene discussie worden de implicaties van de resultaten van deze dissertatie voor de fokkerijindustrie besproken. Hierin wordt uitgeweid over het verschil in benadering van infectieziekteproblematiek tussen fokkerij en epidemiologie, de indirecte genetische effecten die optreden in de transmissie van infectieziekten, de aanwezigheid van genetische variatie in infectieziektekenmerken, op welke kenmerken geselecteerd kan worden om infectieziekten in de veehouderij te verminderen, de fenotypische waarnemingen die nodig zijn om op deze kenmerken te kunnen selecteren, hoe deze fenotypes kunnen worden gemeten en vereisten voor praktische implementatie. De belangrijkste conclusies uit deze discussie zijn dat selectie voor verhoogde tolerantie van infectie geen veilig alternatief is voor selectie voor lagere

gevoeligheid, omdat ziekteverwekkers zich ook kunnen aanpassen aan tolerante gastheren. Verder leiden ontwikkelingen op het gebied van sensoren, beeldherkenning en automatische intelligentie tot meer mogelijkheden om frequent te observeren, wat hoognodig is voor precieze schatting van fokwaardes voor gevoeligheid en mogelijk ook infectieusiteit. Waar mogelijk moeten transmissie-experimenten in plaats van 'challenge'-experimenten uitgevoerd worden om aan fenotypes te komen. Als laatste wordt beargumenteerd dat voor een effectieve en duurzame selectie tegen infectieziekten, fokprogramma's de epidemiologische benadering dienen te kiezen en uitroeiing van infectie na te streven, bijvoorbeeld door stap voor stap lokale populaties vrij te maken, in plaats van zich te richten op langzame afname van prevalentie in de algehele populatie. Een dergelijke verandering vereist echter een zekere mate van controle op de genetische kuddesamenstelling van commercieel gehouden landbouwhuisdieren en is daarom mogelijk niet toepasbaar bij ieder diersoort.

About the author

Andries Hulst was born on the 14th of July 1993 in Hoorn, The Netherlands. After secondary school, he completed a BSc in Veterinary Medicine at Utrecht University. Not considering himself a future vet, he switched to the MSc Animal Sciences at Wageningen University, with specialisation in Genetics and Biodiversity. His MSc thesis 'Genetic epidemiological analysis of endemic diseases', also under supervision of Dr Piter Bijma and Prof. Mart de Jong, was awarded the UFW-KLV Thesis Award in the domain Life Sciences in 2020. He continued his work on the integration of quantitative epidemiology and animal breeding in his PhD-project, of which the results are described in this thesis.

Dankwoord/Acknowledgements

This thesis would never have been as it is without the help and input from many people, too many too mention. So first of all, thanks to all colleagues from ABG and QVE for their support, feedback, interesting discussion groups, coffee breaks, company on conferences and for making the hours in the office a bit pleasant.

Een speciaal woord van dank voor Mart en Piter. Allereerst om mij überhaupt zo gek te krijgen om een afstudeervak te gaan doen over de genetica van infectieziekten. Dit bleek en blijkt een fascinerend en ingewikkeld genoeg onderwerp om niet alleen maar een half jaar mee bezig te zijn. Ook nu, vier jaar en een PhD verder, liggen er misschien wel meer vragen dan toen ik begon. Het vertrouwen in mijn capaciteiten en de grote mate van vrijheid in het project bleken een grote luxe en hebben geresulteerd in een mooi en interessant proefschrift. Ons wekelijks overleg ontspoorde soms misschien een klein beetje, maar het staat buiten kijf dat dit altijd interessante en leerzame uren waren. Jullie toewijding, interesse, inhoudelijke kennis en vermogen om dit over te brengen zijn van zeer grote waarde geweest voor mijn PhD en bij de totstandkoming van dit proefschrift.

Andrea, Ricardo and the other members of Andrea's group at the Roslin Institute, I deeply appreciate our ongoing collaboration and really enjoyed my visit to Roslin. It has been very motivating to meet with people that work on the same topic and speak the same language. Ricardo, your work on the GLMM-part has been of great value, I enjoy our discussions and your, sometimes stubborn, different perspectives. Andrea, your warm and friendly attitude made me feel very welcome in your group. Your knowledge of and ideas on the genetics of infectious diseases have been very valuable for my PhD, the last chapters of this thesis and has provided (and I'm sure will provide) me the so much needed external perspective on the topic.

Als laatste mijn allerliefste Wendy, schop onder mijn kont wanneer ik die nodig had, heel veel broodjes kaas en appelstroop of hagelslag en kopjes koffie die uit het niets op mijn bureau staan tijdens lange Teams meetings. Zonder dat alles was het zeker niet gelukt.

Training and Supervision Plan (TSP)

A. The Basic Package	**year**
WIAS Introduction Day	2019
Scientific Integrity (WGS)	2020
Ethics and Animal Sciences (WGS)	2020
WIAS Introduction course on personal effectiveness for your PhD	2021
B. Disciplinary Competences	
Writing WIAS project proposal	2020
ABG Course Genotype by environment interaction, uniformity and resilience (WIAS)	2020
Review WIAS PhD proposal	2022
Advanced Statistics course Design of Experiments	2023
C. Professional Competences	
Writing in English for Publication (Babel, Utrecht)	2019-2020
Presenting with Impact (In'to)	2020
Supervising BSc & MSc thesis students (ESC)	2022
WIAS Course - The Final Touch: Writing the General Introduction and Discussion	2022
Organisation of Quantitative Genetics Discussion Group (ABG)	2021-2023
D. Societal Relevance	
WIAS course Societal Impact of your Research	2021-2022
Writing news item about publication of 1st paper for ABG newsletter	2021

E. Presentation Skills

71st annual meeting of EAAP, online (oral)	2020
ModAH 2021, online (oral)	2021
72nd annual meeting of EAAP, Davos, Switzerland (oral)	2021
WIAS Annual Conference, Lunteren (oral)	2022
WCGALP 2022, Rotterdam (oral)	2022
WIAS Annual Conference, Ede (oral)	2023
74th annual meeting of EAAP, Lyon, France (oral)	2023

F. Teaching competences

Assisting practicals MIDA (MSc)	2020-2023
Assisting practicals/case study ABG (BSc)	2020-2021
Assisting practicals/case study GIL (MSc)	2020
Supervision group IAS practical project (BSc)	2022
Supervising 3 MSc theses students	2021 - 2022

Education and Training Total (minimum 30 credits)* **30 ECTS**

*One ECTS credit equals a study load of approximately 28 hours

The research described in this thesis was financially supported by Wageningen University.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Ron Zijlmans, www.ron.nu

Printed by proefschriftmaken.nl on FSC-certified paper