Title

Mechanistic Mathematical Models for Understanding The Evolution and Epidemiology of Infectious Diseases

Author

Northrup, Graham Richard

Publication Date

2024

Peer reviewed|Thesis/dissertation

Mechanistic Mathematical Models for Understanding The Evolution and Epidemiology of Infectious Diseases

by

Graham Northrup

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computational Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mike Boots, Chair
Associate Professor Joe Lewnard
Professor in Residence John Marshall

Summer 2024

Mechanistic Mathematical Models for Understanding The Evolution and Epidemiology of Infectious Diseases

Abstract

Mechanistic Mathematical Models for Understanding The Evolution and Epidemiology of Infectious Diseases

by

Graham Northrup

Doctor of Philosophy in Computational Biology

University of California, Berkeley

Professor Mike Boots, Chair

Mathematical and statistical models of the biological world are powerful tools for working with these complex systems and distilling them down to interpretable phenomenon which we can use to develop our understanding of the dynamics at play. In this dissertation I put forward a pair of mathematical models, which advance our understanding of two crucial infectious disease systems. Firstly, I present a mathematical model of hyperparasitism, that is parasitism of a parasite, that allows us to understand guiding principles behind the evolution of these hyperparasites in nature. Among the results, I critically show that the proportion of hyperparasitized hosts that transmit both their parasite and hyperparasite has a core impact on the evolutionary outcomes of the system. This probability of co-transmitting or "hitchhiking" by the hyperparasite is central to the dynamics of a hyperparasite system. Second, I present a mathematical framework for more accurately modeling the way in which immunity from vaccination wanes over time in a host. With it I am able to derive an analytical relationship between this waning process and the ability of a population to prevent infections via population level immunity. With this relationship I show that we must measure more precisely both the waning process and the immunity remaining after the waning process, in order to create better vaccine control strategies for infectious disease. Finally, this motivates the creation of a new non-parametric estimation method for naturally acquired immunity from infection. I use a negative control design and show we can create a new odds ratio estimator for the reduction in susceptibility associated with infection and show that for certain conditions it is unbiased and is robust to individual heterogeneity in susceptibility. I then compare it to other methods for estimating this quantity, and apply it to cohort data for rotavirus infection.

To Sara, Colin, Rick, and Dawn

I truly feel like anything is possible with your support. Thank you for everything.

Contents

Acknowledgments

Thank you to Dr. Mike Boots for inspiring me to ask interesting questions and having the grace to let me grow as a scientist. Thank you to Dr. Joe Lewnard for motivating me to answer questions precisely and including me on wonderful projects and experiences during a tumultuous time. Thank you to Dr. John Marshall for convincing me to choose Berkeley for my PhD, a decision which has shaped my life immeasurably in the past six years. Thank you to the rest of my qualifying exam committee, Dr. Peter Sudmant, Dr. Perry de Valpine, and Dr. Ayesha Mahmud for charging bravely into the world of zoom qualifying exams during 2020 and for the kind feedback during the examination. Thank you to Dr. Britt Koskella for being a thoughtful and sharp colleague during presentations and conversations, and for all the meals I ate at your home. Thank you to Dr. Andy White, for being such a positive force and for always being available to help. Thank you to the members of the Boots lab who have been wonderful friends, colleagues, and mentors during this process. To Carly, Cara, Senay, and Lawrence, thank you for modeling success in academia and always being patient with a new colleague. To Laura, Sarah, and Elisa thank you for welcoming me with open arms into the community you helped create. To Nina, your brilliance, authenticity, and talent are without compare and I am honored to be your academic fraternal twin. To Sara, Claire, John, and Darien, thank you for giving me hope and sharing your enthusiasm, I know the group is in safe and capable hands. To Signe, thank you for always being willing to listen when I just had to pop my head over the wall and tell you something and for laughing at my jokes. To Chadi, thank you for the walks, the talks, the advice, and the sushi, I am so thankful you came to Berkeley when you did and for putting wind back into my scientific sails. To Xander, thank you for all the advice and for the great adventures in the Bay Area. To my CCB 2018 cohort Adam, Gonzalo, Kennedy, Ryan, and Isabel, thank you for being my first family of this crazy ride and for motivating me to keep up with your accomplishments. To Doug, for being the only outlet for my extroversion during the lockdowns of 2020. To Kate Chase for being the most helpful and reliable administrator on this campus. To Ryan, Dinakar, Kevin, Aaron, Richard, Kevin, Potts, Steven, and Arash, for not letting distance get in the way of friendship. Thank you to Dr. Sarah Cobey for taking a risk on a math major with an interest in biology and for your guidance as I learned what it meant to be a scientist. To Dr. Dmitry Kondrashov for showing me the ways which math and biology can sing together. Thank you to all the teachers I had who helped foster my curiosity, Mrs. Wells, Mr. Fry, Mr. Dudash, Gillian, Mrs. Dutcher, Mr. McCargar, and Mr. Baird. Thank you to my parents Rick and Dawn, for letting me be myself and grow and explore all that life has to offer. To my brother Colin, for being my biggest, most unwavering supporter. To my wife Dr. Sara Furukawa, for your ardent belief in me and challenging me to grow everyday. And to Roger, for the love you give without question.

Chapter 1

Evolutionary Dynamics of Hyperparasites

The work of this dissertation focuses on the use of mathematical and statistical approaches to improve the way we understand infectious disease biology systems. Infectious diseases span a great diversity of taxonomic categories, geographic regions, and effects on their hosts. As such, the methods used to approach these systems are similarly diverse with different systems and questions requiring different models and approaches. The first application of mathematical models for infectious disease epidemiology and evolution I present in this dissertation comes from the study of hyperparasites, or parasites of parasites. Their unique biology makes them an interesting study system with dynamics not seen in simple two species models. As you will see below, we can use a concise mathematical framework to describe the dynamics of these three species systems. This new model incorporates the specific biology of hyperparasites and allows for a more generalized approach than previous models. With this model I make predictions about the evolutionary outcomes of the hyperparasite.

1.1 Introduction

Evolutionary theory is well developed for simple pairwise examples of evolution such as predator/prey [1], host/mutualist [4] and in particular host/parasite [24] interactions. This theory only rarely considers interactions within communities of interacting species [131, 28]. The coevolution of hosts and pathogens is well studied both theoretically [124, 11, 24, 25] and empirically [27, 43, 73, 50, 55], providing a solid framework for understanding of the key evolutionary drivers of their life history such as virulence, resistance and recovery [35]. However, far less is known about how host-parasite interaction traits evolve when we account for the biotic community they are embedded in [5, 100, 59]. There are clearly analytical and conceptual challenges to moving beyond pairwise interactions, but one important and tractable multi-parasite interaction is hyperparasitism [104]. Hyperparasites are parasites

that infect hosts which are themselves a parasite of another host, they are widespread, and diverse in nature [104] and as such they represent an important but understudied three-way interaction. In this chapter, I develop a general evolutionary framework for hyperparasites referring throughout to the three players in this system as the host, the parasite, and the hyperparasite.

Modeling of hyperparasite systems is motivated not only by our general interest in understanding evolution beyond pairwise interactions, but also due to the number of diverse and impactful hyperparasitic systems in nature [103, 33, 135]. Hyperparasites exist across a diversity of systems and types of hosts in nature. A well-known example is the chestnut blight fungus, Cryphonectria parasitica, a parasite of chestnut trees that also plays host to the hyperparasitic CHV-1 mycovirus which has been shown to reduce the virulence in chestnut trees caused by C. parasitica [33]. There are other fungal hyperparasites such as Ampelomyces quisqualis which parasitizes powdery mildews [68] and of course a wide range of bacteriophage hyperparasites of bacterial pathogens [36]. In addition, there is considerable interest in the application of hyperparasites in biocontrol. In particular, phage therapy has already been used to treat antibiotic resistant (AMR) infections in several cases, including in particular in cystic fibrosis patients [72], and the hyperparasitic mycovirus has been successfully used to control chestnut blight [112]. More generally, it is now thought that the diversity of hyperparasites may be an important component of observed pathogen virulence patterns [103]. Understanding the ecological and evolutionary dynamics of these systems is therefore important both due to their role in natural systems and their potential use in therapeutics [142].

Current hyperparasite theory has often been focused on specific systems and therefore makes system specific assumptions about the impact of the hyperparasite on its host [128, 94] although Sandhu et al. 2021 has recently pursued more general results[123]. Here we focus on the unique characteristics of a hyperparasite. A hyperparasite decreases the fitness of its own host (a parasite), a hyperparasite has the potential to impact the effect of the parasite on its host, including changing the parasite's virulence, and a hyperparasite may increase the death rate of the parasite (hyperparasite virulence). It is important to understand that in this context hyperparasite virulence essentially clears the parasitic infection and leads to an uninfected host. This is a critical difference for hyperparasite systems compared to a parasite in a traditional host parasite system where virulence is the increased death rate of the host. In some sense hyperparasite virulence is more closely related to host recovery in traditional host parasite systems . In addition, understanding the conditions under which a hyperparasite may evolve to reduce virulence of their parasite host is of particular interest. This will elucidate the effects of hyperparasites in nature or when they are introduced as biocontrol agents. Importantly, we examine the impact of the hyperparasite's ability to be transmitted along with the parasite when it infects a new host (hitchhiking).

Holt and Hochberg (1998) proposed a hyperparasite model with a focus on how this important tritrophic interaction impacts the ecological stability of communities [62]. The model assumed the same death rates in parasitized and hyperparasitized hosts ($c_\alpha = 1$ in our general framework defined below) and no hyperparasite virulence, meaning that hyperpara-

sitized hosts do not recover to be fully susceptible. This means there is effectively no damage from hyperparasite infection to the parasite. Their models do not allow direct hyperinfection of a susceptible host from a hyperinfected host (i.e. they assume 0% hitchhiking, p=0 in our general framework) and their obligate hyperparasite model also does not allow transmission of hyperinfected parasites ($c_\beta = 0$ in our general framework). Their model predicts that hyperparasitism should select for higher parasite virulence [62], which is equivalent to the prediction of the related co-infection model [88]. Taylor et al. (1998) examine the evolution of virulence reduction in a model focused on the C. parasitica and CHV-1 system [128]. They assume 100% hitchhiking, meaning that an infection event between a susceptible host and a hyperinfected host will always transmit both pathogens to the new host, but assume reduced transmission of a hyperparasitized parasite ($c_\beta < 1$ in our general framework) and have a parameter which measures the difference in transmissibility that occurs through hitchhiking (in our general framework this is an S and H contact, and this is called vertical transmission in Taylor et al. 1998). Recently, Sandhu et al. (2021) examined the coevolution of hyperparasites and parasites with an explicit adaptive dynamical model. They predict that hyperparasites will always select for higher parasite virulence in their host, although they can still act as excellent biocontrol agents as they reduce parasite prevalence, and also show the interesting possibility of evolutionary suicide. This work assumes there is no recovery from hyperparasitized to susceptible hosts as a result of hyperparasites exploiting and killing the parasite (their 'host') and therefore, they did not examine hyperparasite virulence directly and furthermore, they assumed 100% hitchhiking. As such we lack a general understanding of the evolution of hyperparasites as existing theory has not focused on key traits such as the role of hyperparasite virulence and variation in the degrees of hyperparasite hitchhiking. We build on this previous work, by creating a general framework for the evolution of hyperparasites, incorporating virulence of the hyperparasite on the parasite, and in particular specifically examining how the assumption of hyperparasite hitchhiking affects evolutionary outcomes. While previous general evolutionary theory has concentrated on the impact of the hyperparasite on parasite evolution, we explicitly model the evolution of key hyperparasite traits including its virulence. We also include parameters which allow us to understand the extent to which a hyperparasite reducing the growth rate of the parasite within the host may also influence hyperparasite traits and evolutionary trajectories. Lastly, we emphasize that the degree of hitchhiking is a key characteristic of any particular hyperparasite system as an additional transmission pathway with fundamental implications to the evolutionary outcome.

1.2 Model Design

In order to understand the evolutionary dynamics and outcomes of hyperparasites, we build a general modelling framework consisting of three ordinary differential equations that capture the infection-status of a host population. The system of equations describes the dynamics of uninfected hosts (S), hosts infected with a parasite (I), and hyperinfected hosts

(H). Hyperinfected hosts are those that are infected by the intermediate parasite which is itself infected with the hyperparasite (see Figure 1.1). Here we consider obligate hyperparasites only, which require the parasite present in the host.

$$\frac{dS}{dt} = (b(1-qN)-d)S - \beta_I SI - c_\beta \beta_I SH + \gamma_I I + (c_\gamma \gamma_I + \alpha_H)H \tag{1.1}$$

$$\frac{dI}{dt} = \beta_I SI + (1-p)c_\beta \beta_I SH + \gamma_H H - \beta_H IH - (\alpha_I + \gamma_I + d)I \tag{1.2}$$

$$\frac{dR}{dt} = pc_\beta \beta_I SH + \beta_H IH - (c_\alpha \alpha_I + c_\gamma \gamma_I + d + \alpha_H + \gamma_H)H \tag{1.3}$$

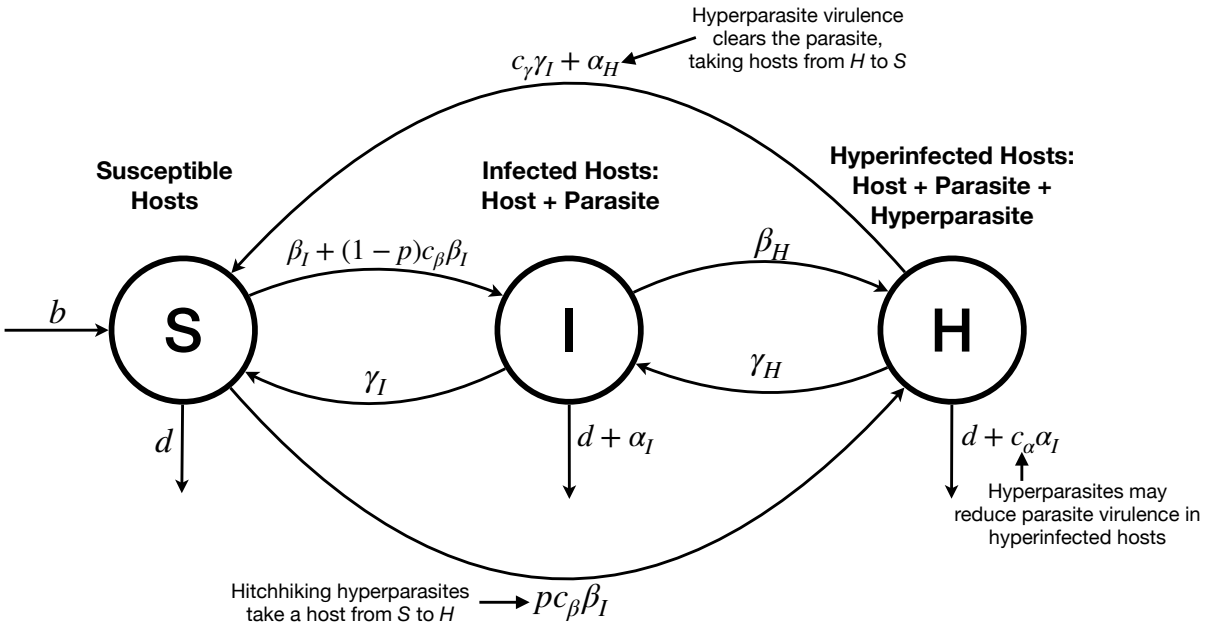

Figure 1.1: Diagram of model for hyperparasite system representing susceptible hosts, S, singly infected hosts. I, and hyperinfected hosts, H. The parameters α_I and α_H describe the virulence of the parasite and hyperparasite respectively. β_I and β_H are the transmission coefficients for the parasite and hyperparasite respectively. γ_I and γ_H are the recovery rates for the parasite and hyperparasite respectively. The parameter p is the hitchhiking probability, the probability a hyperinfected parasite brings its hyperinfection to a susceptible host. The parameters c_α ,c_β, and c_γ are the hyperparasite effects on parasite virulence, transmission, and recovery, respectively.

In this system we have representing susceptible hosts, S, singly infected hosts. I, and hyperinfected hosts, H. The parameters α_I and α_H describe the virulence of the parasite and

hyperparasite respectively.β_I and β_H are the transmission coefficients for the parasite and hyperparasite respectively.γ_I and γ_H are the recovery rates for the parasite and hyperparasite respectively. The parameter p is the hitchhiking probability, the probability a hyperinfected parasite brings its hyperinfection to a susceptible host. The parameters c_α, c_β and c_γ are the hyperparasite effects on parasite virulence, transmission, and recovery, respectively. All of these parameters default values are listed in Table 1.1

Parameters	Default Values	Interpretation
b, d, q	$1, 0.1, 0.0005$	Host demographics
$\alpha_I, \beta_I, \gamma_I$	$0.05, 0.1, 1$	Parasite life history
$\alpha_H, \beta_H, \gamma_H$	(under selection), (under selection), 1	Hyperparasite life history
$c_\alpha, c_\beta, c_\beta$	$1, 1, 1$	Hyperparasite effects on parasite
p	0.5	Hitchhiking

Table 1.1: Parameter default values and definitions

Susceptible hosts experience density dependent growth, with b as the growth rate and q as the crowding term. All hosts experience a natural death rate d, regardless of infection status. Susceptible hosts can become infected by contact with infected hosts at rate β_I. They can also become infected by hyperinfected hosts, leaving the hyperparasite behind during transmission, at rate $(1-p)c_\beta\beta_I$ reflecting the hyperparasite's ability to affect parasite fitness, c_β, and the probability, $1-p$, a susceptible host interacting with a hyperinfected host creates an infected host (rather than becoming a hyperinfected host that occurs with probability, p). Infected hosts can also become hyperinfected through contact with hyperinfected hosts at rate β_H. Infected and hyperinfected hosts can die due to parasite virulence at rate α_I and $c_\alpha\alpha_I$ respectively with c_α representing how the hyperparasite affects parasite virulence. Infected hosts recover, returning to the susceptible class, at rate γ_I while hyperinfected hosts recover and return to the infected class at rate γ_H and to the susceptible class at rate $c_\gamma\gamma_I$ (recovering from the parasite also de facto clears the hyperparasite from the host). In this same sense hyperinfected hosts can also 'recover' to the susceptible class by virtue of the hyperparasites virulence against the parasite.
An important concept in this model is that hyperparasite virulence, α_H, takes host individuals from H to S, 'killing' the parasite but not the host. This is because when the hyperparasite kills its 'host', the parasite, it functionally clears the 'base' host of its original parasite infection. In this way hyperparasite virulence, α_H, increases the recovery rate of the host in addition to natural recovery, γ_I, where the host clears the parasite. Therefore, while the standard parasite virulence, α_I, increases the mortality of its host, hyperparasite virulence α_H decreases the mortality of the base host. Hyperparasite virulence, α_H, takes individuals out of H, thus in an equivalent way to α_I it reduces the duration of the hyperparasite infection, but crucially it doesn't remove hosts from the population and instead allows them to recover. The recognition of hyperparasite virulence as a process by which hyperparasites

exploit intermediate parasites, rather that the base host (as might occur in superinfection), is a key, unique, feature of hyperparasites that has not been included in previous model studies [123].

A further key feature of our model framework is the introduction of the parameter p, to capture the propensity for exposure to hyperinfected hosts to infect the susceptible host with both the parasite and the hyperparasite. When hyperparasitized hosts come into contact with susceptible hosts, a proportion p become hyperparasitized (they contract the parasite as well as the hyperparasite) while a proportion $1-p$ lose the hyperparasite and only contract the parasite. This parameter is important as it can fundamentally change the structure of the interaction and emphasizes that there is an important distinction between hyperparasites that "hitchhike" along primary infections by the parasite and those that are not transmitted along with the parasite. The degree to which a hyperparasite is able to be transmitted with its parasite is a crucial characteristic of the hyperparasite. With $p = 0$ the hyperparasite transmission is decoupled from that of its parasite host, but once there is significant hitchhiking there is a fundamentally different relationship between the fitness of the hyperparasite and parasite.

The parameters c_α,c_β and c_γ give us considerable flexibility to parameterize this model for a wide variety of host-parasite-hyperparasite systems since they describe the effect of hyperparasites on their parasite hosts' key life histories. Different hyperparasites can affect their hosts' life histories in various ways, and this general modeling approach allows us to incorporate multiple types of effects of hyperparasites on their parasite hosts. These key traits are c_β (parasite transmission modification), c_α (parasite virulence modification), c_γ (parasite recovery modification). Each of these traits reflect the cost to the parasite of being infected with the hyperparasite which has the potential to make them less able to transmit c_β and easier to recover from c_γ and potentially less impactful on their host c_α. One simple way in which these effects can be understood is that the hyperparasite reduces the growth rate of the parasite within its host. Clearly not all hyperparasites will impact each of these traits, but in principle they are all processes where the hyperparasite reduces the parasites fitness. It is important to note that this impact on the fitness of the parasite is the key assumption that defines the hyperparasite – it is a parasite because it reduces the fitness of its host, which is also a parasite. If a hyperparasite did not impact its parasite host fitness, through one or more of these modification terms or virulence α_H, it would be more similar to a mutualist or commensal for the parasite.

An important driver of the evolution of antagonist ecological systems is the feedbacks between ecological and evolutionary dynamics [56, 125, 18, 26]. In order to understand the evolutionary implications of our model, in response to ecological feedbacks, we made use of adaptive dynamics to investigate how mutant hyperparasites with a small change in a trait of interest may or may not outcompete the wild type hyperparasite. To do this we generate a function for the fitness of a hyperparasite in the system, which we can then work with to identify the hyperparasite traits which optimize this fitness.

We introduce two tradeoff schemes, the first being a standard virulence transmission tradeoff where $\beta_H = (\alpha_H)^{(23)}$ [35]. Note that this function has accelerating costs (it is saturating)

meaning as hyperparasite exploitation of the parasite increases, the level of transmission begins to level off as it is also influenced by other factors such as contact rate. This choice allows us to make guarantees about convergence stability later, as well as conveniently comparing well to our understanding of transmission (once a parasite load is sufficiently high, creation of more parasites doesn't do much to increase onward transmission as the infected host is already infecting most or all of the susceptible hosts it contacts). We add further tradeoffs by assuming the parameters $c_\alpha, c_\beta, c_\gamma$ are functions of parasite 'host' exploitation by the hyperparasite. We assume $c_\alpha = 1 - \frac{\alpha_H}{\alpha_{Hmax}}$ and $c_\beta = 1 - \frac{\alpha_H}{\alpha_{Hmax}}$ which both decrease as host exploitation increases $c_\gamma = 1 + \frac{\alpha_H}{\alpha_{Hmax}}$ which increases as host exploitation increased. These trade off functions are all simple linear functions with respect to α_H, normalized in order to ensure they stay in the proper domains. As α_H increases the parasite will have a decreased ability to exploit the host, and as such parasite transmission and virulence will be modulated down by c_α and c_β being less than 1. Similarly host recovery will be modulated up by c_γ being larger than 1. We opted for a simple linear tradeoff here due to its mathematical simplicity and the lack of a concrete biological motivation for something more complex. In this way we can understand how effects such as the reduction in virulence of the intermediate host might be adaptive. We examine a wide range of scenarios to develop a general theory of the evolution of hyperparasites.

1.3 Results

We first investigated the existence and nature of the evolutionarily stable strategy (ESS) of virulence. By considering a mutant hyperparasite strain, with parameters α_{Hm} and β_{Hm} (initially assuming that $c_\alpha, c_\beta, c_\gamma$ are constants) we can write the mutant fitness function, r_m, as

$$r_m = pc_\beta\beta_I S + \beta_H m I - (c_\alpha\alpha_I + c_\gamma\gamma_I + d + \alpha_H m + \gamma_H) \tag{1.4}$$

where S and I are the steady state densities for the endemic resident strain (with parameters α_H and β_H). This is done by dividing a mutant version of 1.3 by H_m and rearranging the terms, which gives us a per capita growth rate of the mutant (e.g. change in H_m divided by the size of the H_m class). The higher this per capita growth rate is, the more fit a mutant will be. Using the theory of adaptive dynamics [51] evolution will occur in the positive direction of the local fitness gradient, which when we include the tradeoff function, $\beta = f(\alpha)$, is defined as

$$\frac{\partial r_m}{\partial \alpha_{Hm}} = f'(\alpha_{Hm})I - 1 \tag{1.5}$$

An evolutionarily stable (ES) point, α_H^*, occurs when the fitness gradient is zero, which is defined as

$$f'(\alpha_H^*) = \frac{1}{I^*} \tag{1.6}$$

where I^* is the steady state value of I at α_H^* and ' refers to differentiation with respect to α_H. When we choose an accelerating tradeoff function, such as $\beta_H = (\alpha_H)^{2/3}$, we can interpret 1.6 as a monotonic relationship between the equilibrium parasite density, I^*, and the evolutionarily stable hyperparasite virulence. As I^* increases, the value of $f^{'}(\alpha_H^*)$ must decrease meaning α_H^* increases. When the parameters $c_\alpha, c_\beta, c_\gamma$ are functions of α_H the mutant fitness expression is more complicated, and can be written as follows:

$$r_m = pc_{\beta m}\beta_I S + \beta_{Hm} I - (c_{\alpha m}\alpha_I + c_{\gamma m}\gamma_I + d + \alpha_{Hm} + \gamma_H) \tag{1.7}$$

This expression is more complicated in the sense that more of the parameters are a function of the mutant virulence α_{Hm}, noted with a subscript m. These functions must be differentiated when we optimize the fitness below. Then, the evolutionary stable (ES) point, α_H^*, occurs when

$$p\beta_I S^*(c_\beta^*)^{'} + f^{'}(\alpha_H^*)I^* - (c_\alpha^*)^{'}\alpha_I - (c_\gamma^*)'\gamma_I - 1 = 0 \tag{1.8}$$

where * refers to values determined at α_H^* and ' refers to differentiation with respect to α_H. We can use equations 1.6 and 1.8 to determine the value of α_H^* for different parameter combinations. To ensure that α_H^* is a continuously stable strategy, CSS, (an end point of evolution) we must also ensure that it is evolutionary stable and convergence stable (Geritz et al. 1998). We select a tradeoff that ensures that α_H^* is a CSS (Bowers et al. 2005).

A key insight is that p has a critical impact on the evolutionary stable virulence, α_H^*, for the hyperparasite (Figure 1.2A). Therefore whether the hyperparasite is able to hitchhike with the parasite at infection is critical to the evolutionary outcome. Fundamentally, as the value of p decreases, the ES virulence (α_H^*) increases. At the maximum value of $p = 1$, when the hyperparasite perfectly hitchhikes, the model selects for avirulence of the hyperparasite ($\alpha_H^* = 0$). This can be understood because as p approaches 1, more arrivals to the hyperinfected class arrive through the $SH\beta_I$ transmission term, rather than passing through the infected class first. Basically, as new parasite infections occur, they come "pre-hyperinfected" similar to a vertical transmission model. Conversely as p approaches 0, and the hyperparasites stop hitchhiking as frequently, individuals must pass through the infected category to reach the hyperinfected category, thus increasing the relative importance of higher β_H values and thus also higher levels of virulence at the ESS. This result is key to understanding how the fundamental nature of hyperparasite transmission will affect the ESS, as the interplay between p and β_H can exert significant control over the system. In particular, hyperparasites will be selected to reduce their own transmission to very low levels when they can hitchhike and become mutualistic with their parasites – leading to commensalism/mutualism.

We examine the model results when we implement a broader tradeoff to include the parameters $c_\alpha, c_\beta, c_\gamma$ all as functions of host exploitation (Figure 1.2). We implement all three tradeoff functions of hyperparasite virulence to create the most extreme case of hyperparasite infection impairing the parasites natural life history. Results indicate that p still exerts strong control over the ES virulence of the hyperparasite (Figure 1.2A). Again with $p = 1$ the hyperparasite is selected towards hypermutualism and not impacting the parasite fitness. We also examined the importance of the parasite's life history parameters, to understand

how this may influence the evolutionary properties of the hyperparasite). In particular, increasing values of α_I corresponds to an increase in α_H^* (Figure 1.2B). Meaning if the parasite is deadly, its hyperparasite will evolve to become more virulent in turn. Similarly, when increasing the host recovery rate from the parasite, γ_I, this causes an increase in hyperparasite virulence (Figure 1.2C). As host recovery rate or parasite virulence increases, the effective lifespan of the hyperparasite (duration of parasite infection) decreases. This then selects for more virulent, faster replicating hyperparasites. When hyperparasite evolution can change parasite life history traits, the hyperparasite evolves to lower levels of virulence compared to when hyperparasite has neutral effects on parasite traits (compare the dotted and solid line in Figure 1.2). This is because an increase in hyperparasite virulence now reduces parasite fitness through multiple modes and therefore reduces the need for the hyperparasite to compensate with higher levels of transmission and associated virulence.

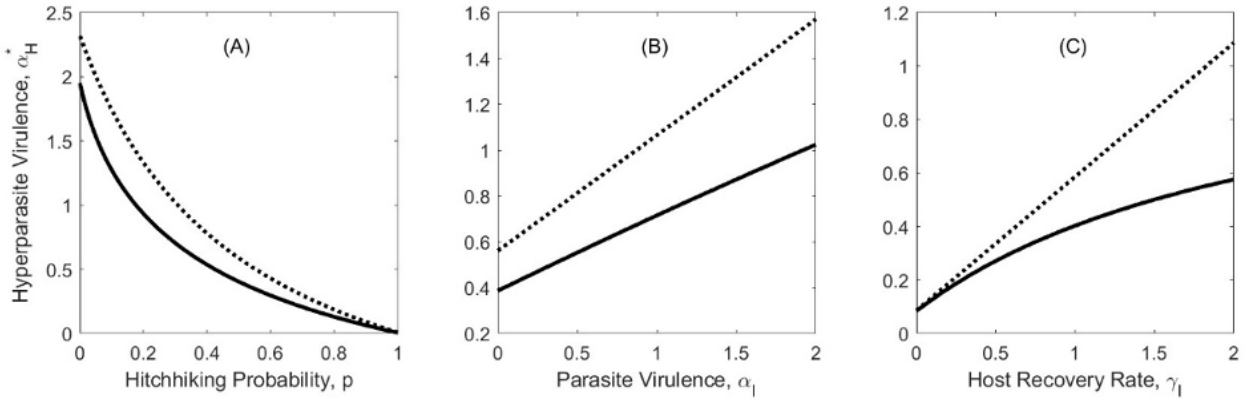

Figure 1.2: Plots of the evolved level of hyperparasite virulence, α_H^*, against (A) hitching hiking probability, p, (B) parasite virulence, α_I and (C) host recovery rate, γ_I. When not varied in the figures the parameter values are $b = 1, d = 0.1, q = 0.0005, \alpha_I = 0.05, \beta_I = 0.01, \gamma_I = 1, p = 0.5, \gamma_H = 0.01$ and we assume a tradeoff $\beta_H = \alpha_H^{(2/3)}$. The dotted line represents the case where $c_\alpha = 1, c_\beta = 1, c_\gamma = 1$ which is consistent with a hyperparasite that has neutral effects on the life history traits of the parasite. The solid line represents the case where $c_\alpha = 1 - (\alpha_H/\alpha_{max}), c_\beta = 1 - (\alpha_H/\alpha_{max}), c_\gamma = 1 + (\alpha_H/\alpha_{max})$, where $\alpha_{max} = 5$, in which hyperparasite virulence also effects the interaction between the hyperparasite and the parasite.

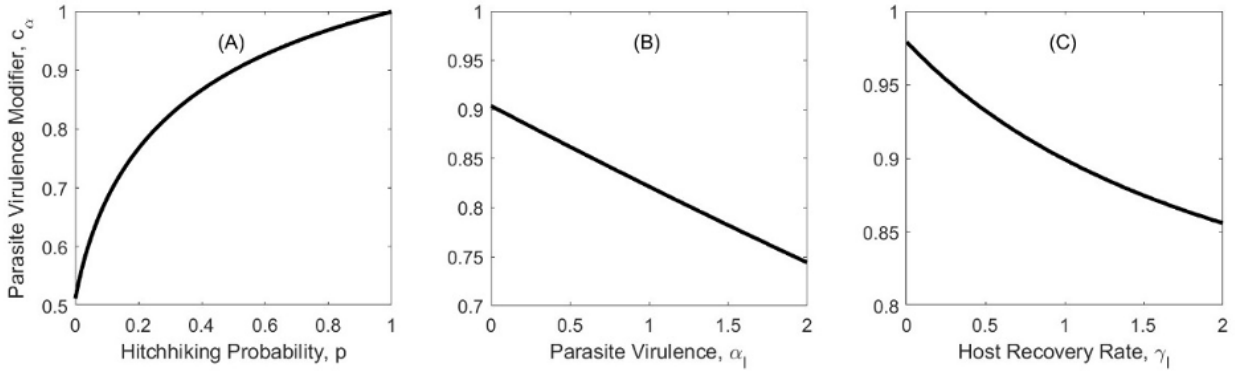

Figure 1.3: Plots of evolutionarily stable parasite virulence modifier, $c_\alpha(\alpha_H^*)$, against (A) hitching hiking probability, p, (B) parasite virulence, α_I and (C) host recovery rate, γ_I. When not varied in the figures the parameter values are $b = 1, d = 0.1, q = 0.0005, \alpha_I = 0.05, \beta_I = 0.01, \gamma_I = 1, p = 0.5, \gamma_H = 0.01$ and we assume a tradeoff $\beta_H = \alpha_H^{(2/3)}$. We assume that $c_\alpha = 1 - (\alpha_H/\alpha_{max}), c_\beta = 1 - (\alpha_H/\alpha_{max}), c_\gamma = 1 + (\alpha_H/\alpha_{max})$, where $\alpha_{max} = 5$, in which hyperparasite virulence also effects the interaction between the hyperparasite and the parasite.

We can also explore the consequences of hyperparasite trait evolution on the parasite's life history through lens of c_α, the effect of the hyperparasite on parasite virulence α_I (Figure 1.3). This is of interest in the context of parasite virulence evolution as well as possible biocontrol uses for certain hyperparasites. As shown in Figure 1.2B and 1.3B we can see that increasing the virulence of their host will select for hyperparasites with higher virulence and therefore this causes a stronger reduction in parasite virulence. This is because the hyperparasite will need to exploit its host more, a reduction in the value of c_α under the tradeoff scheme. A similar effect is seen when increasing host recovery rate, γ_I. The opposite effect is seen when increasing p, where c_α values will increase as the ESS shifts accordingly with decreased host exploitation in line with the decreasing virulence and transmission previously discussed, according to the hypothesized tradeoff where c_α grows as host exploitation decreases.
Finally, we examine how host traits impact the selection on hyperparasite traits. In particular longer-lived hosts selected for less virulent hyperparasites (Figure 1.4). This is consistent with our understanding of virulence selection as increasing the host death rate would reduce the duration of hyperinfection. This in turn provides selection pressure for more rapid growth and reproduction of the hyperparasite, the source of virulence under our assumptions.

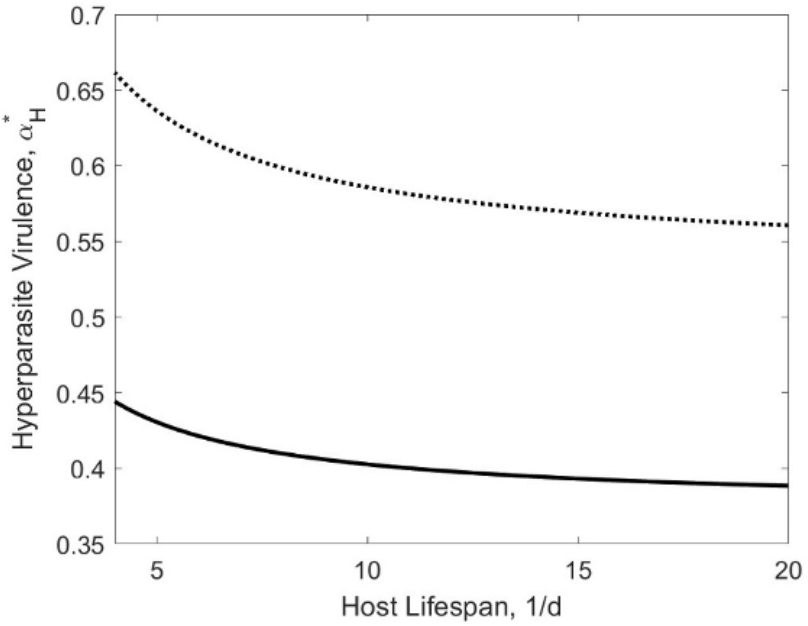

Figure 1.4: Plot of evolutionarily stable hyperparasite virulence, α_H^*, against host lifespan defined as $1/d$. When not varied in the figures the parameter values are $b = 1, d = 0.1, q = 0.0005, \alpha_I = 0.05, \beta_I = 0.01, \gamma_I = 1, p = 0.5, \gamma_H = 0.01$ and we assume a tradeoff $\beta_H = \alpha_H^{(2/3)}$. The dotted line represents the case where $c_\alpha = 1, c_\beta = 1, c_\gamma = 1$ which is consistent with a hyperparasite that has neutral effects on the life history traits of the parasite. The solid line represents the case where $c_\alpha = 1 - (\alpha_H/\alpha_{max}), c_\beta = 1 - (\alpha_H/\alpha_{max}), c_\gamma = 1 + (\alpha_H/\alpha_{max})$, where $\alpha_{max} = 5$, in which hyperparasite virulence also effects the interaction between the hyperparasite and the parasite.

1.4 Discussion

We have presented a general model of hyperparasitic interactions that can be applied to a range of specific systems and gives us broad insights into the evolutionary dynamics of hyperparasites. The model makes explicit the unique evolutionary dynamics resulting from the interaction between two pathogens, one playing host to another, and sharing a base host. Understanding the evolution of the hyperparasite, and specifically the evolution of the effect of the hyperparasite on the parasite's traits, is key to understanding these systems more generally. We are only just beginning to uncover the true diversity of hyperparasites in natural populations and to date, little is known about the ecological and evolutionary consequences of hyperparasitism in nature. Our model can guide empirical work by identifying key life-history traits and hypotheses to be tested.

We have shown that the virulence of hyperparasites (and accordingly, their effects on the par-

asite's (their host) natural history parameters) are sensitive to changes in parasite and host parameters as well as the key characteristics of the hyperparasitic interaction. Importantly, the extent to which hyperparasites are transmitted with their parasites – hitchhiking – is critical to the evolutionary dynamics in the hyperparasite. Our key result is that increased proportions of hitchhiking, p, by a hyperparasite causes selection for lower levels of hyperparasite virulence (α_H), all the way to when $p = 1$ (hitchhiking every time) when avirulence is selected for. This makes intuitive sense when considering the virulence transmission tradeoff implemented in the model. We expect both virulence and transmission to be functions of host exploitation, with lower virulence and higher transmission being favorable (hence the tradeoff, as virulence decreases, so does transmission). However, when a hyperparasite is able to hitchhike, this leads to a different transmission pathway for the hyperparasite that is not subject to the virulence transmission tradeoff and therefore 'cost free'. This intuition about an alternate transmission pathway is similar to what we see in models of vertical transmission. Effectively as hitchhiking increases, the fitness of the parasite and the hyperparasite are more tightly linked with both invested in parasite transmission. As hitchhiking increases the hyperparasite is less and less constrained by its own classical virulence-transmission trade off, ultimately resulting in the avirulence shown (which the hyperparasite can tolerate as it is still transmitting to new hosts via hitchhiking). It is worth noting that the hyperparasite could still be incurring costs that we aren't considering in this model even with lots of hitchhiking, such as the negative impact of the hyperparasite on its parasite host may take place during life-history stages that take place between transmission seasons [132]. Here we do not examine hitchhiking as an evolvable trait as we believe it is most likely a consequence of the nature of a particular system (i.e. type of parasite, hyperparasite, specific biology), but our models emphasize that it is absolutely the critical trait determining the evolution of hyperparasite systems. The proportion of hyperparasite infections that hitchhike at parasite transmission events should therefore be a key trait that is estimated in the field when studying a natural hyperparasite system.

We examined the impact of the evolution of the hyperparasite on its parasitic host assuming that there are tradeoffs imposed on hyperparasite traits as a function of exploitation. This is in line with the classic assumption of the tradeoff theory of parasite virulence and transmission [11, 6]. The conceptual basis of these tradeoffs is that the growth of the hyperparasite in the parasite causes harm that could reduce the growth rate of the parasite in the host and therefore reducing its virulence (decreasing c_α to between 0 and 1) transmission (decreasing c_β to between 0 and 1) and making it easier for the host to clear (increasing c_γ above 1). This is the same conceptual basis to the classic tradeoff assumption for which there is mounting empirical evidence in general host parasite systems [2, 19]. In particular, we focused on values of c_α, the parasites reduction in virulence as a result of hyperinfection. The idea that hyperparasites can reduce virulence in their hosts has been the subject of considerable attention in specific systems [33, 91] but here we examine the general conditions under which it is likely to be found. Selection for this reduction in parasite virulence by a hyperparasite increases in strength as parasite virulence on its host increases and as such, we expect this evolution in hyperparasites of highly virulent parasites. This is also selected for

when hyperparasites cannot hitchhike, as there is selection for higher levels of hyperparasite virulence and transmission in this case (and therefore also values of c_α that are smaller). Parasites with chronic infections (low recovery rates) in our model also select for a similar reduction in hyperparasite virulence. Hyperparasitic biocontrol agents aimed at reducing the virulence of the parasites would therefore be most likely to be found for non-hitchhiking, chronic parasites. In natural communities, a virulence reducing hyperparasite could promote host-parasite coexistence by attenuating parasite virulence [144].

The traits of the parasites are also important in selecting their hyperparasites. In particular, we show that in chronic parasites, where the host is less able to clear infection, will select for decreased hyperparasite virulence. When the duration of hyperparasite infection is longer (due to slow host recovery), hyperparasites do not need to prioritize fast replicating, high virulence strategies. Using similar logic, parasites with higher virulence select for hyperparasites of higher virulence as they prioritize transmission given the shorter lifetime (infectious period) of their parasite host.

We have also shown how host traits can also influence selection pressures on the hyperparasite. Despite not being the hyperparasite‘s direct host, the parasite-hyperparasite interactions occur in the context of the base host. We show that if this host is particularly long lived this may result in selection for lower levels of virulence in the hyperparasite. It is therefore important to not only consider the parasite and hyperparasite traits as important for selection but to look at the tripartite system in its entirety. It follows that a knowledge of the host life history can improve our ability to make predictions of the nature of the parasite-hyperparasite interaction. These predictions may allow the identification of systems which may be susceptible to hyperparasitic invasion or candidates for biocontrol of a parasite [112].

Although there is limited data, we are able to compare our predictions to what we see in hyperparasitic systems in nature. It is important to note that we cannot say anything about precise values of these parameters as the exact tradeoff functions or parameter values are not well known. But our models emphasize that determining whether a hyperparasite is a hitchhiking hyperparasite allows us to make several predictions about what we should see about the other parameter values of interest.

Interest in C. parasitica and CHV-1 stems primarily from the observation that CHV-1 can reduce the virulence of C. parasitica in its host the chestnut tree [33]. Our model would predict that this will occur in CHV-1 in a system where p is close to 1, and if C. parasitica is highly virulent. There has been previous work investigating how not all strains of CHV-1 may reduce virulence in C. parasitica and can be the result of be a combination of biological factors [91]. Further it has been shown that in some cases, virulence reducing strains of CHV-1 can be outcompeted by other strains [29]. While there is strong evidence to suggest there is variation amongst CHV-1 strains in transmission [39, 40], there have been no studies done on variation of hitchhiking ability between strains. What is clear when examining the C. parasitica and CHV-1 system is that C. parasitica can be highly virulent in natural settings [106], and it has been shown that naturally occurring CHV-1 may have been a factor in differences in outcome at the population level [30]. As discussed previously our model shows that the high virulence of C. parasitica provides pressure to drive the system towards

the strategy employed by CHV-1 in natural settings. Overall, it is not surprising to find that the effect of hyperparasites is not perfectly consistent across space and time, as these interactions may be strongly mediated by the local environment [146].
Other prominent examples of hyperparasitic systems suggest that the conclusions derived from this model hold in multiple systems. For example, fungal hyperparasites tend to have low p and therefore select for higher values of α_H [8], including members of the Ampelomyces genus [132, 103]. This is similar to lytic bacteriophage systems which we would also expect p to be low but observed virulence can be extremely high [36, 72]. Given that they are obligate killers, meaning they must kill their host cell in order to release the burst of viral particles, it is unclear if this high virulence has anything to do with the low hitchhiking observed but the relationship between hitchhiking and virulence is certainly present. These observations taken with the CHV-1 system make a compelling case for the relationship between hitchhiking and hyperparasite virulence as described in our model. The hyperparasite effects on parasite life history traits are less clearly matched to our model predictions, and it is unclear if this may be due to naïve assumptions on behalf of the model, specific ecologies or biological mechanisms involved etc. For example, lysogenic phages in contrast are often transmitted with their bacterial hosts and we would therefore predict very different impacts of the lysogenic compared with lytic phage. Adapting our modelling framework to explore the range of phage-bacteria systems is likely to reveal further nuances of coevolutionary dynamics. There is a great diversity of hyperparasites in nature, but relatively little is known about the key parameters we have identified as important for understanding host-parasite-hyperparasite evolution. This chapter provides motivation for studies to estimate more of these parameters in natural systems. In particular there is a critical need to measure the probability of hitchhiking, which our models have shown to have fundamental effect on resulting selection.
In summary, we have presented a general model of the evolution of hyperparasites using adaptive dynamics. We showed that the ability of the hyperparasite to hitchhike with a primary infection can have dramatic effects of the ES levels of virulence and transmission of the hyperparasite. We also show how the life history traits of the intermediate parasite can exert effects over selection on the hyperparasite showing how hyperparasite systems can have very different evolutionary behavior despite having a similar, tri-species, hierarchical structure. Ultimately these results can inform the conditions under which we might expect reductions in virulence by a hyperparasite, with implications for biocontrol.

Acknowledgments and additional information

This work has been published in the Journal of Theoretical Biology [99], please refer to the paper for additional information and for code to reproduce the figures seen in this chapter. I would like to thank Dr. Andy White, Dr. Steven Parratt, Dr. Carly Rozins, Dr. Anna-Liisa Laine, and Dr. Mike Boots for their input and contributions which helped improve this work.

Chapter 2

Shape of Waning Immunity

As seen in chapter one, mathematical models can be a powerful tool for understanding the evolutionary outcomes of highly nonlinear systems such as those seen in infectious disease biology. In this chapter however, I focus on the ecological or epidemiological insights that mechanistic models can provide. Taken together these chapters can paint a clear picture of the variety of insights into these systems that can be taken from straightforward mathematical modeling approaches.

As mentioned, one of the key reasons to be interested in hyperparasites and hyperparasitic systems is their potential use as biocontrol agents, that is, biological tools to control infectious disease transmission. While there is growing interesting in their use in human infectious disease control, vaccines are currently a much more common and powerful tool for control infectious disease burden in human populations. In this chapter I create a mathematical framework for understanding how the precise way in which a host loses their immunity after a vaccination can have large population level implications for disease control.

2.1 Introduction

The COVID-19 pandemic highlighted the importance of waning immunity on epidemiological dynamics and public health interventions (*e.g.* [71, 41, 69, 120]) not least in the context of proposed nonpharmaceutical interventions that hinge on individual immune status, such as 'shield immunity' [143, 81]. The watershed moment during the COVID-19 pandemic was the development and deployment of safe vaccines [111, 20, 139] and while current formulations of COVID-19 vaccines only transiently prevent infection, effective transmission-blocking vaccines can result in pathogen control (see *e.g.* [9, 10, 122]). However, for SARS-CoV-2 and other pathogens, protective immunity from vaccinations is known to wane over time [130, 46]. There have been important recent efforts to generate pan-coronavirus/pan-sarbecovirus vaccines that would provide broad immunological protection (see *e.g.* [85, 102, 22, 31]). and similarly, the development of 'Universal Influenza Vaccines (UIVs)' has been an important focus (see *e.g.* [116, 17, 13, 14, 108, 16]).. Such broadly-protective vaccines will be

crucial to prevent the next pandemic and given the current situation with H5N1 influenza, it is important for us to better understand the epidemiological dynamics of vaccination in the context of waning immunity (see *e.g.* [66]).

Clearly, whether community immunity prevents pathogen establishment (or whether elimination is achieved if vaccination begins after establishment) crucially hinges on the strength and duration of vaccinal immunity. Furthermore, it may depend on the way in which immunity is lost through time, *i.e.* the shape of waning immunity. Thus, an important question is how immunity wanes, in addition to the rate at which it is lost. For example, does it matter if immunity remains relatively high and then falls off rapidly (an accelerating loss of immunity), or if there an initial loss followed by a slow decline (a decelerating loss)? It seems clear that the duration and shape of waning immunity are likely to be important to epidemiological dynamics and control through vaccination.

Classic susceptible-infected-recovered-susceptible (SIRS) models capture waning immunity by using a single waning rate or immunity period, after which the host is susceptible fully again [21]. As such they assume individuals are either fully susceptible or fully immune, but it is clear that individuals may never return to full susceptibility, even if they lose their (initial) immunity against all infection. To capture this effect, Morris et al. [95] developed a model with buffered susceptibility, where individuals wane to a different, but also susceptible, class. Recent work has leveraged simple mathematical models with buffered susceptibility to investigate the impacts of immune uncertainties on potential future SARS-CoV-2 trajectories [120, 119, 141, 121] and have showed that the characteristics of buffered immunity can lead to a large range of future immuno-epidemiological outcomes. In other work, El Khalifi and Britton [44, 45] have examined gradual waning (*i.e.* regaining progressively stronger partial susceptibility during waning) in a variety of contexts. To model gradual waning in an SIRS model, they considered individuals flowing through a series of partial immune (and thus partially susceptible) classes until returning to complete susceptibility, and they compared linear and exponential waning with classical assumptions [44]. As a calibration method, they assumed that the cumulative immunity was constant. In subsequent work [45], these authors further developed this framework to include heterogeneity. While El Khalifi and Britton [44, 45] showed the effects of gradual waning in SIRS frameworks, and Saad-Roy, Wagner et al. [120] examined the case where individuals wane to a partially susceptible class (*i.e.* 'buffered susceptibility), the effects of the shape of immunity with buffered susceptibility remains unknown.

In this chapter, I develop a framework to examine the effects of the shape of waning vaccinal immunity on pathogen control. We use a buffered susceptible approach and consider multiple susceptible classes, each with their own relative susceptibility to infection, with the last (fully-waned) state (potentially) preserving some immunity (*i.e.* having lower susceptibility to infection than a never-exposed individual, see [95, 120, 122, 141, 121]). Thus, our mathematical model consists of a series of partially susceptible classes, which extends (and generalizes) previous analyses with two 'susceptible' classes (see [120]). We capture waning immunity with different shapes and examine how this impacts epidemiological dynamics and in particular pathogen control through vaccination. Furthermore, while El Khalifi and Brit-

ton [44, 45] calibrated their model based on cumulative immunity, we assume here that the duration until individuals are fully-waned is the same, and vary the relative susceptibility as they progress through multiple partially immune states. Thus, the cumulative immunity for particular schemes may vary, and future work should examine the effects of keeping this constant in a buffered susceptibility framework.

2.2 Model framework

We extend the model of Saad-Roy, Wagner et al. [120] to include multiple partially susceptible classes. However, to focus on the waning process, we ignore characteristics of infections in priorly-immune individuals (which was modelled in [120]). Thus, in our model, S_0 is the class of completely naive hosts, S_1 describes the class of vaccinated individuals where immunity has fully waned so that the relative susceptibility to infection is ε_0 (e.g. $\varepsilon_0 = 1$ describes a vaccine where immunity is eventually completely lost). For $j = 2, ..., n$, S_j describes vaccinated hosts that flow from S_{j+1}, where S_n consists of individuals that have been most recently vaccinated. Finally, I represents the infected class of hosts. We denote μ as the birth/death rate of the host, ν as the vaccination rate of the host population (*i.e.* we assume that vaccination occurs at random), ω as the waning rate of the immunity (*i.e.* $\frac{1}{\omega}$ is the average time to get from S_n to S_1), β as the transmission rate, and γ as the recovery rate. Furthermore, $\varepsilon(j)$ denotes the relative susceptibility of individuals in S_j, *i.e.* this represents the degree of waning that has occurred. Thus, the function $\varepsilon(j)$ is the 'shape of waning immunity'. The model equations are therefore

$$\frac{dS_0}{dt} = \mu - (\mu + \nu)S_0 - \beta S_0 I, \tag{2.1}$$

$$\frac{dS_1}{dt} = m\omega S_2 - (\mu + \nu)S_1 - \beta\varepsilon(1)S_1 I, \tag{2.2}$$

$$\frac{dS_i}{dt} = m\omega S_{i+1} - (\mu + \nu + m\omega)S_i - \beta\varepsilon(i)S_i I \qquad (1 < i < n), \tag{2.3}$$

$$\frac{dS_n}{dt} = \gamma I + \nu \sum_{j=0}^{m} S_j - (\mu + m\omega)S_n, \tag{2.4}$$

$$\frac{dI}{dt} = \beta S_0 I + \beta \sum_{j=1}^{m} \varepsilon(j)S_j I - (\mu + \gamma)I, \tag{2.5}$$

where $m = n - 1$.

Derivation of equilibrium values

In this chapter, we examine the effect of the shape of waning immunity on potential pathogen invasion. Thus, we focus on the disease free equilibrium (P_0, where $I_0 = 0$), and

the vaccination required so that it is stable. At the disease free equilibrium, $I = 0$. Since the other differential equations are also set to 0 as well, solving these gives S_0^* immediately

$$S_0^* = \frac{\mu}{\mu + \nu} \tag{2.6}$$

To get S_n^*, we start with the $\frac{dS_n}{dt}$ equation where it is set equal to 0 and $I = 0$.

$$0 = \nu \sum_{j=0}^{m} S_j - (\mu + m\omega)S_n \tag{2.7}$$

If we add and subtract νS_n, and use the fact that because $I = 0$ the remaining classes S_0 to S_n must sum to 1 we get the following

$$0 = \nu \sum_{j=0}^{m} S_j - (\mu + m\omega)S_n \tag{2.8}$$

$$0 = \nu \sum_{j=0}^{m} S_j - (\mu + m\omega)S_n + \nu S_n - \nu S_n \tag{2.9}$$

$$0 = \nu \sum_{j=0}^{n} S_j - (\mu + \nu + m\omega+)S_n \tag{2.10}$$

$$0 = \nu - (\mu + \nu + m\omega+)S_n \tag{2.11}$$

$$S_n^* = \frac{\nu}{\mu + \nu + m\omega} \tag{2.12}$$

Next we can see there is a recurrence relation for the susceptible classes between 1 and n

$$S_i^* = \frac{m\omega}{\mu + \nu + m\omega} S_{i+1}^* \tag{2.13}$$

Then we can use this recurrence relation to get an expression for S_i^* as a function of S_n which we already know

$$S_i^* = (\frac{m\omega}{\mu + \nu + m\omega})^{n-i} S_n^* \tag{2.14}$$

$$S_i^* = (\frac{m\omega}{\mu + \nu + m\omega}) \frac{\nu}{\mu + \nu + m\omega} \tag{2.15}$$

We can get S_1^* by remembering the classes S_0 to S_n sum to 1 at the disease free equilibrium

$$S_1^* = 1 - S_0 - \sum_{i=2}^{n} S_i^* \tag{2.16}$$

$$\tag{2.17}$$

The sum here is a geometric sum so we can use a substitution trick to make progress

$$S_1^* = 1 - \frac{\mu}{\mu+\nu} - \frac{\nu}{\mu+\nu+m\omega}\left(\frac{1-(\frac{m\omega}{\mu+\nu+m\omega})^m}{1-\frac{m\omega}{\mu+\nu+m\omega}}\right) \tag{2.18}$$

$$= 1 - \frac{\mu}{\mu+\nu} - \frac{\nu}{\mu+\nu+m\omega}\left(\frac{1-(\frac{m\omega}{\mu+\nu+m\omega})^m}{\frac{\mu+\nu}{\mu+\nu+m\omega}}\right) \tag{2.19}$$

$$= 1 - \frac{\mu}{\mu+\nu} - \frac{\nu}{\mu+\nu+m\omega}\frac{\mu+\nu+m\omega}{\mu+\nu}\left(1-(\frac{m\omega}{\mu+\nu+m\omega})^m\right) \tag{2.20}$$

$$= 1 - \frac{\mu}{\mu+\nu} - \frac{\nu}{\mu+\nu}\left(1-(\frac{m\omega}{\mu+\nu+m\omega})^m\right) \tag{2.21}$$

$$= 1 - \frac{\mu}{\mu+\nu} - \frac{\nu}{\mu+\nu} + \frac{\nu}{\mu+\nu}(\frac{m\omega}{\mu+\nu+m\omega})^m \tag{2.22}$$

$$S_1^* = \frac{\nu}{\mu+\nu}(\frac{m\omega}{\mu+\nu+m\omega})^m \tag{2.23}$$

Which gives us the complete set of equilibrium values

$$\begin{aligned}
S_0^* &= \frac{\mu}{\mu+\nu} \\
S_1^* &= \frac{\nu}{\mu+\nu}(\frac{m\omega}{\mu+\nu+m\omega})^m \\
S_i^* &= (\frac{m\omega}{\mu+\nu+m\omega})^{n-i}\frac{\nu}{\mu+\nu+m\omega} \qquad (1<i<n) \\
S_n^* &= \frac{\nu}{\mu+\nu+m\omega}
\end{aligned}$$

$\mathcal{R}_{inv}$ and Forms of waning

In the absence of vaccination, the basic reproduction number is that of the classic SIR model, *i.e.*

$$\mathcal{R}_0 = \frac{\beta}{\mu+\alpha+\gamma} \tag{2.24}$$

With vaccination, *i.e.* $\nu > 0$, the control reproduction number $\mathcal{R}_C$ can be determined using the linearization of the $\frac{dI}{dt}$ equation at the disease-free equilibrium (see [134] for a generalized approach), *i.e.*,

$$\mathcal{R}_C = \frac{\partial}{\partial I}\frac{dI}{dt}\bigg|_{P_0} = \mathcal{R}_0\left(S_0^* + \sum_{j=1}^{m}\varepsilon(j)S_j^*\right). \tag{2.25}$$

Thus, when $\frac{1}{S_0^*+\sum_{i=1}^{m}\varepsilon(i)S_i^*}$ is less than $\mathcal{R}_0$, the pathogen can invade a population that is undergoing vaccination at rate ν, with the shape of immunity $\varepsilon(i)$ and rate of waning immunity ω. We denote the value of $\mathcal{R}_0$ required to invade a population being vaccinated under such a scheme to be $\mathcal{R}_{inv}$ or the "R invasion". Thus, we can compute this quantity

$\mathcal{R}_{inv}$ for any formulation of $\varepsilon(i)$. Clearly $\mathcal{R}_{inv}$ depends on the particular shape of waning. To illustrate a variety of cases, we use two different waning schemes. Both of which, facilitate the waning from fully immune individuals, $\varepsilon(n) = 0$, to fully waned individuals $\varepsilon(1) = \varepsilon_0$. The first is a simple exponential-like decelerating or accelerating function where more waning happens early or late in the waning period (Figure 2.1B), *i.e.*

$$\varepsilon(i) = \varepsilon_0 \left(\frac{n-i}{n-1} \right)^p . \tag{2.26}$$

The second scheme we use is a Hill function, to model 'threshold'-like waning (Figure 2.1C-D), *i.e.*,

$$\varepsilon(i) = \varepsilon_0 \frac{(n-i)^p}{k^p + (n-i)^p} \left(\frac{k^p + (n-1)^p}{(n-1)^p} \right) . \tag{2.27}$$

Note that the factor $\left(\frac{k^p + (n-1)^p}{(n-1)^p} \right)$ is so that $\varepsilon(1) = \varepsilon_0$.

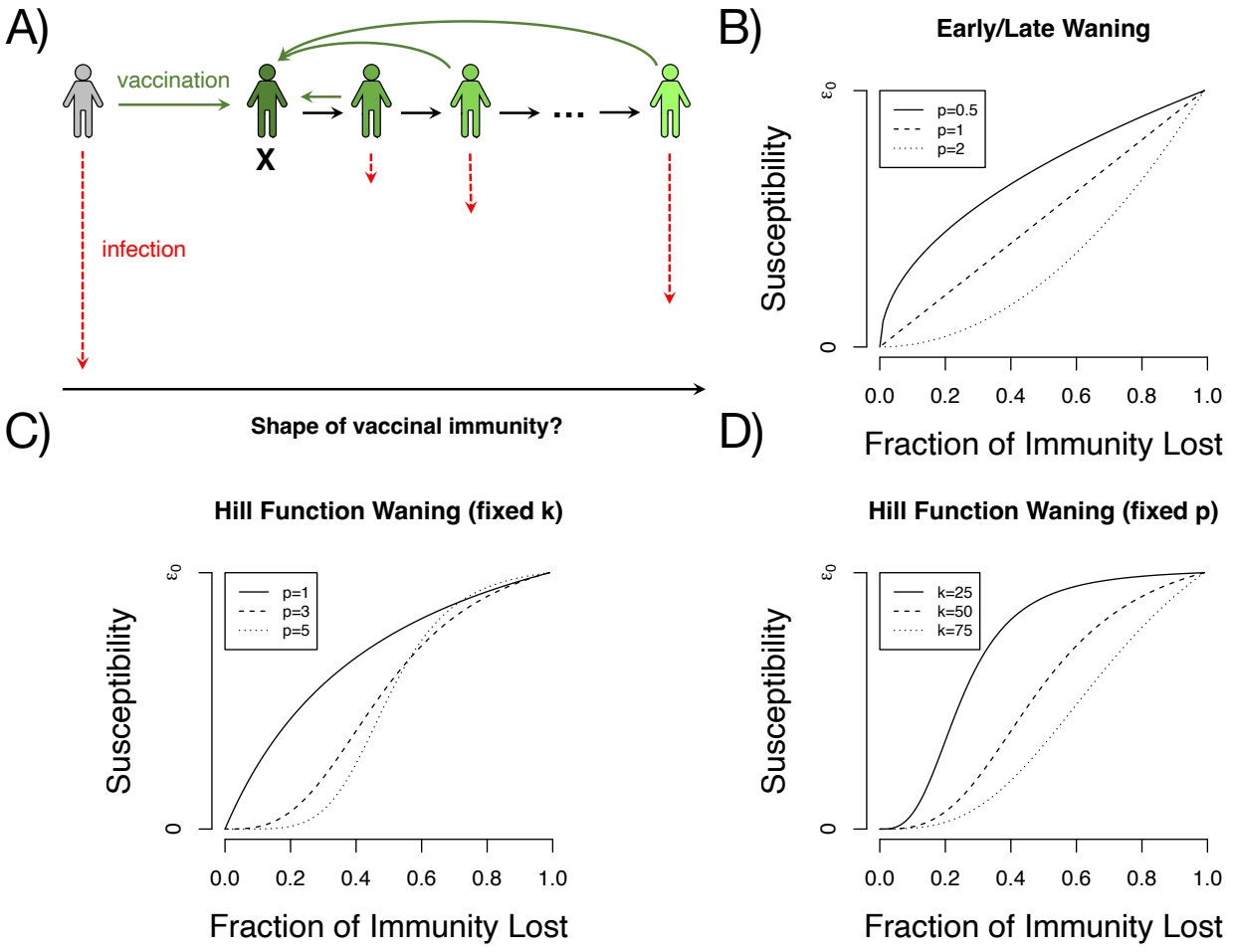

Figure 2.1: A) shows a schematic of population vaccination and immunity waning process. Subplot B-D shows possible variation in the multiplier of ε_0. B) shows accelerating/decelerating waning when $p = 0.5, 1, 2$. C) Shows a hill-like function for fixed $k = 50$ and $p = 1, 3, 5$. D) Shows the hill-like function fixed for $p = 3$ and $k = 25, 50, 75$

2.3 Results and Discussion

We begin by visualizing $\mathcal{R}_{inv}$ over a variety of vaccination rates, waning rates, and over various waning schemes. In Figure 2.2, we illustrate the dependence of $\mathcal{R}_{inv}$ on the duration of immunity and vaccination rate. For a particular parameter pair, we see that changing the waning scheme can dramatically alter how the protection of a population. For example, a comparison of the early and late waning schemes (Figure 2.2, comparing the top row and bottom row respectively) for the same vaccination rate and duration of immunity, reveals

that the $\mathcal{R}_{inv}$ can be more than four times greater if waning happens later. This is a massive difference, and has important implications for population level pathogen prevention.

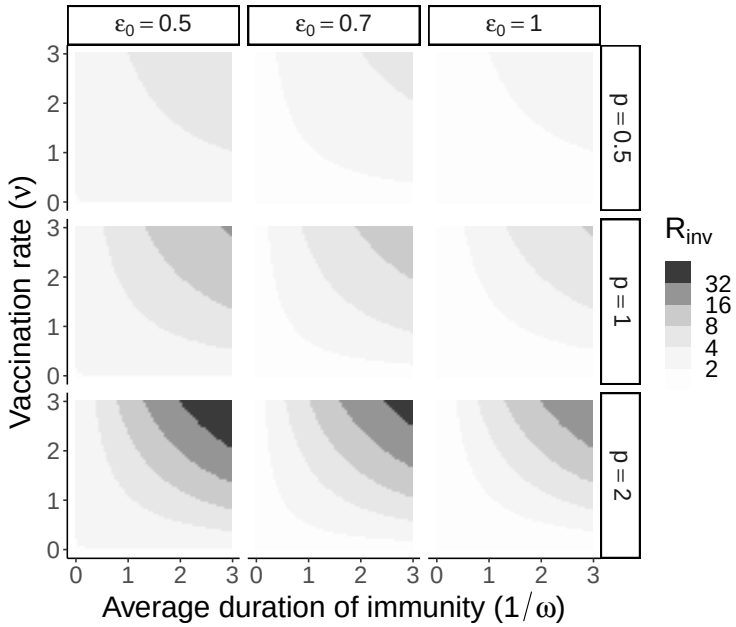

Figure 2.2: $\mathcal{R}_{inv}$ calculated for the combination of duration of immunity, vaccination rate, immunity remaining after waning, and the shape parameter of the waning function.

So far, we have assumed that the shape of waning immunity is exponential-like, and thus is either concave-up or concave-down. In reality, waning is likely more complex, with changes in concavity in the relative susceptibility to infection. To explore this, we use a Hill-like function to describe the shape of waning immunity. In Figure 2.3, we plot $\mathcal{R}_{inv}$ (as in Figure 2.2), but we now varying the two waning shape parameters of the Hill-like function. In particular, we find that as p increases there is some increase in $\mathcal{R}_{inv}$. However, the effect size is much larger if the relatively susceptibility after waning (ε_0) is changed. This is due in part to each form of the Hill-like waning function ultimately waning to ε_0 regardless of shape, so fully waned individuals are still contributing to population level protection more if ε_0 is smaller. As such, these two parameters combine to explain the protection experienced by the population. In Figure 2.3B, we adjust k instead, showing it can also have a significant impact on the resulting $\mathcal{R}_{inv}$. As expected, we find that when individuals lose their immunity earlier in the waning period (when k is smaller), the population is overall more susceptible, and thus the $\mathcal{R}_{inv}$ is lower (Note that this is akin to our findings in 2.2).

In both situations that we have examined, it is clear that ε_0 and the shape of vaccinal waning can have a major effect on the resulting values of $\mathcal{R}_{inv}$. Importantly, as more immunity is preserved after the waning period (*i.e.*, as ε_0 is decreased), the potential for progressively

larger *inv* values becomes possible. Additionally, waning curves that preserve immunity early in the waning process provide higher levels of population protection. Therefore, our results illustrate the importance of including both the shape of waning and this endpoint level of immunity in epidemiological models. In particular, such granularity enables more precise predictions about the $\mathcal{R}_{inv}$ for a particular population. Furthermore, increases in vaccination rates can alleviate pessimistic outcomes. Thus, quantifying the shape of waning vaccinal immunity is paramount, especially for emerging broadly-protective vaccines.

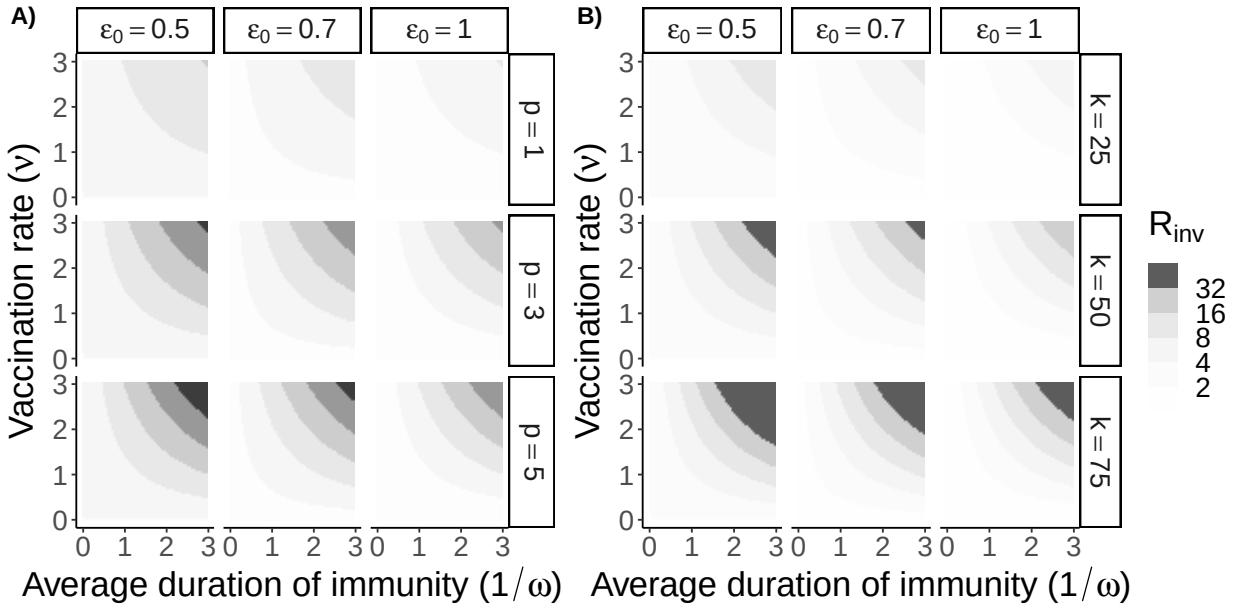

Figure 2.3: $\mathcal{R}_{inv}$ calculated for the combination of duration of immunity, vaccination rate, immunity remaining after waning, and the shape parameter of the waning function.

In Figure 2.4, we illustrate the fold change in $\mathcal{R}_{inv}$ for a particular shape parameter and relative susceptibility after waning for both shape functions. These changes are with respect to a baseline assumption about the shape of the waning function, where $\varepsilon_0 = 1$ and $p = 1$ (e.g. simple linear waning all the way to fully susceptible, see Figure 2.1B) for the simple accelerating/decelerating waning function or $\varepsilon_0 = 1$, $p = 3$, and $k = 50$ (*e.g.* a parameter set where p and k both can exert control over the shape of the curve), see Figure 2.1C-D) for the Hill-like function type waning.

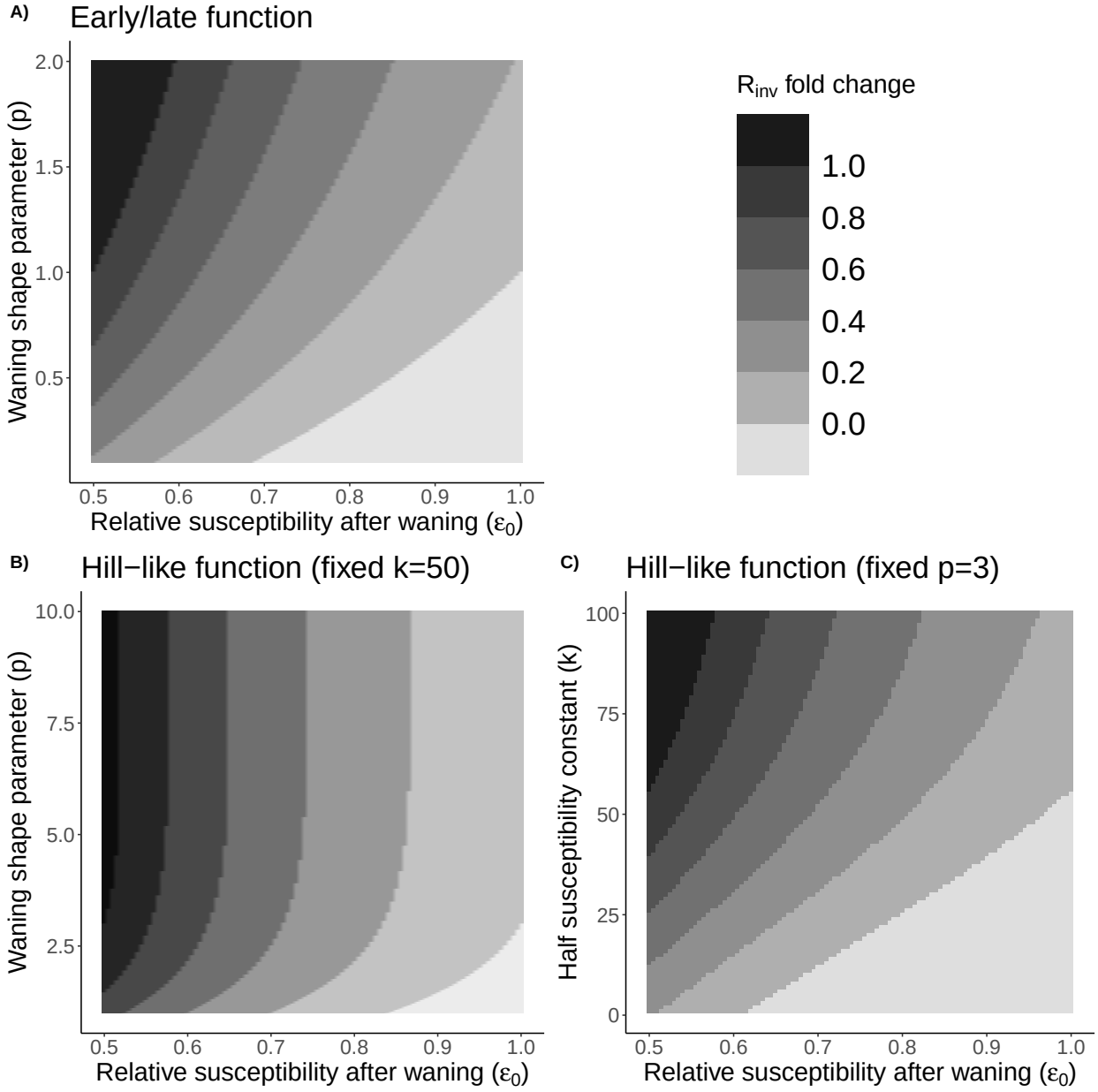

Figure 2.4: Shows the $\mathcal{R}_{inv}$ fold change with respect to a baseline waning shape where ε_0 for the A) accelerating/decelerating waning function where the baseline is $p = 1$, or the B) hill-like function for fixed $k = n/2$ or for C) fixed $p = 3$ where the baseline is $k = 50$ and $p = 3$

We can again here see the importance of understanding how waning happens in an individual. The change in $\mathcal{R}_{inv}$ across the range of even this modest set of choices can be

more than double, or even result in a decrease in the $\mathcal{R}_{inv}$ for a given choice compared to the baseline. When we look at specifically threshold type waning, we see that the relative susceptibility after waning (ε_0) exerts much more control over this change than p. But when we allow k to vary we can see a very large range of outcomes for $\mathcal{R}_{inv}$.

The shape of the waning process clearly has large implications for the population level dynamics, as even for the same exact vaccination strategy and duration of immunity we can see populations can be protected against pathogens with $\mathcal{R}_0$ up to four times another population with a different waning process. In particular, the amount of immunity that is lost early in the waning period and the relative susceptibility after waning (ε_0) matter significantly to population level susceptibility. As more immunity is preserved later into and after the waning period, a population is comparatively much more protected. This improved protection is true for multiple for forms of waning functions, showing that what matters is the host susceptibility rather than any particular functional form.

2.4 Caveats and Future Work

To examine the effect of the shape of waning vaccinal immunity on community immunity, our modelling framework has many simplifying assumptions, and we have omitted other important characteristic. Furthemore, and perhaps most importantly, we have ignored epidemiological dynamics and focused on community immunity. However, in the absence of such protection, pathogen invasion would be possible, and it would be particularly salient to explore the resulting medium- and long-term dynamics in our model. Classical work has shown that sustained epidemiological oscillations can occur if there are more than two fully immune compartments in an SIRS-like model [61]. Thus, our model may also exhibit such periodicity, and investigating this possibility would be important to guide vaccination policies for ongoing epidemics.

Second, our model assumes that individuals are vaccinated randomly. However, non-random vaccination, *e.g.*, vaccinating individuals who have gone the longest without receiving a vaccination, could cause increases in $\mathcal{R}_{inv}$. Furthermore, heterogeneities in vaccine uptake, or in age structure [74] or transmission [75, 7], could also affect $\mathcal{R}_{inv}$. For example, May and Anderson [87] examined the effect of heterogeneities in space on optimal vaccination. Furthermore, both the duration and the shape of waning vaccinal immunity may themselves be age-dependent. Thus, the interplay of these factors with the shape of waning vaccinal immunity are important areas for future research.

Third, we have focused on transmission-blocking immunity, but clinical severity of infections as vaccinal immunity wanes may be epidemiologically important, and may crucially shape policy. Relatedly, we have ignored the economics of vaccination in our model. However, these will likely hinge on clinical severity, and can affect decision-making regarding vaccine development and deployment[70, 17]. Thus, extending our model to include clinical immunity in addition to economic considerations would be fruitful avenues to explore.

Additionally, social dynamics and adherence to (pharmaceutical or nonpharmaceutical) interventions can influence pathogen dynamics, and vice-versa (see *e.g.* [23, 133, 118]). Furthermore, vaccine hesitancy and refusal may shape population susceptibility [140, 42]. For example, the synergistic effects of multiple family members being vaccinated together or social groups refusing vaccines can be important factors for pathogen control via community immunity.

Finally, while social dynamics could potentially complicate vaccination strategies, our model could also be adapted to incorporate other biological features, such as maternal immunity of explicit vaccinal cross-immunity. For example, maternal antibodies present in infants could be key for elimination of certain pathogens, and these wane over time [58]. Additionally, our framework could be adapted to examine strain-specific cross-protection, which is key for understanding how UIVs may be employed [116, 17, 13, 14, 108, 16].

2.5 Conclusion

In the midst of the current circulation of H5N1 in cattle, the general pandemic risk of influenza viruses has been underlined. In parallel, current circulation of SARS-CoV-2, and the continued emergence of new variants, highlights the importance of community immunity. To prevent pathogen invasion, such protection hinges on the characteristics of vaccinal immunity. These include duration, strength (*i.e.* relative susceptibility after waning is complete), and, importantly, how vaccinal immunity is lost. Thus, understanding the shape of waning vaccinal immunity is central to control efforts.

In this chapter, I developed a modeling framework to investigate the interplay between the shape of waning vaccinal immunity with buffered susceptibility. For general vaccination and waning schemes, we analytically computed the $\mathcal{R}_0$ pathogens would require for successful invasion. We then examined multiple potential waning functions, and we showed that the relative susceptibility of a fully-waned host is a crucial determinant of invasion potential. Furthermore, we illustrated that a simple change regarding how much immunity wanes early after infection versus later in the waning period could result in more than four times the population-level protection.

Overall, this work highlights the importance of a potential Global Immunological Observatory [90, 89, 92] and large cohort studies [117] to determine the shape of vaccinal immunity for a number of emerging and circulating pathogens. Furthermore, these data would be particularly relevant in the context of the development and deployment of pan-coronavirus/pan-sarbecovirus vaccines, in addition to broadly-protective influenza vaccines, and could inform policy on repeat vaccinations.

Acknowledgments and additional information

I would like to thank Dr. Mike Boots and Dr. Chadi Saad-Roy for their input and contributions which helped improve this work.

Chapter 3

Inference of Naturally Acquired Immunity

While chapters one and two both present new mathematical frameworks for understanding evolutionary and epidemiological questions related to infectious disease, it is important to know what to do with these new insights. Specifically, chapter two motivates an important question; how do we identify infectious diseases which may make good vaccine candidates? Clearly it is important that a pathogen elicits an adaptive immune response, which we want to identify from data. In this final chapter, I present a new method of identifying the level of immunity conferred by a natural infection which makes use of a negative control design that allows the estimation to be robust to individual differences in susceptibility. Taken together with chapter two, these chapters provide a clearer understanding of naturally acquired and vaccine conferred immune landscape.

3.1 Introduction

Host adaptive immune responses may protect against infection or disease when a pathogen is repeatedly encountered. The hazard ratio of infection or disease, given previous infection, is typically sought to estimate the strength of protective immunity. However, variation in individual exposure or susceptibility to infection may introduce frailty bias, whereby a tendency for infections to recur among individuals with greater risk confounds the causal association between previous infection and susceptibility. We introduce a self-matched "case-only" inference method to control for unmeasured individual heterogeneity, making use of negative-control endpoints not attributable to the pathogen of interest. To control for confounding, this method compares event times for endpoints due to the pathogen of interest and negative-control endpoints during counterfactual risk periods, defined according to individuals' infection history. We derive a standard Mantel-Haenszel (matched) odds ratio conveying the effect of prior infection on time to recurrence. We compare performance of this approach to several proportional hazards modeling frameworks, and estimate statistical power of the

proposed strategy under various conditions. In an example application, we use the proposed method to re-estimate naturally-acquired protection against rotavirus gastroenteritis using data from previously-published cohort studies. This self-matched negative-control design may present a flexible alternative to existing approaches for analyzing naturally-acquired immunity, as well as other exposures affecting the distribution of recurrent event times.

Host adaptive immune responses often protect against infection or disease when a pathogen is repeatedly encountered. Vaccines aim to exploit this mechanism of protection by exposing hosts to an attenuated infection, or to immunizing subunits of a pathogen. As such, evidence of protective naturally-acquired immunity provides strong rationale for vaccine development[80]. Quantitative estimates of the strength of naturally-acquired protection also inform the interpretation of epidemiologic data, for instance providing a baseline against which vaccine performance can be evaluated[105]. These estimates are further sought to parameterize mathematical models of pathogen transmission[60].

Naturally-acquired immunity is often estimated via the hazard ratio of infection or disease, comparing counterfactual periods representing person-time at risk in the presence and absence of prior infection[137, 52, 115, 67, 12, 64, 113]. Thus, inference centers on the distribution of recurrent event times. Unmeasured heterogeneity in individuals' hazards of infection or disease presents a challenge in such analyses, originally termed a problem of "varying liabilities" by Greenwood and Yule[57] and subsequently addressed as "accident-proneness"[15] or "frailty".[136] The tendency for events to recur among certain individuals must be accounted for in statistical analyses[54]. In studies of naturally-acquired immunity, recurrence of infection or disease among individuals with the greatest susceptibility or exposure to a pathogen, irrespective of previous infection, may bias estimates of naturally-acquired protection[76].

This consideration may have relevance to several diseases against which immune responses are thought to generate imperfect protection. Tuberculosis presents a notable example, where despite evidence of protective cell-mediated and humoral immunity[3], several epidemiologic studies have reported higher rates of new-onset infection or disease among persons previously treated successfully for active tuberculosis, as compared to those without history of tuberculosis[138, 86, 32, 53]. Similar conflict about the consequences of prior infection has arisen in epidemiologic studies of gonorrhea[48, 110]. In recent analyses of a multi-site pediatric cohort study addressing enteric disease, previous infection predicted higher rates of recurrent infection or disease associated with several pathogens, including Shigella spp., Campylobacter spp., and various diarrheagenic Escherichia coli strains[114]. Evidence supporting the feasibility of protective vaccines against many of these pathogens suggests a need to revisit the impacts of naturally-acquired immunity[127, 107, 82]. Similar causal inference challenges arise in the relationship between chronic inflammation and repeated infection in conditions such as cystic fibrosis[38, 109], otitis media[37, 63], and environmental enteric dysfunction[65].

Formalizing unmeasured heterogeneity as a problem of confounding suggests potential strategies to identify naturally-acquired protection. Terming Y1 and Y2 as primary and recurrent infection or disease outcomes, respectively, and U as the constellation of unmeasured

individual factors influencing exposure or susceptibility to a pathogen of interest, a directed acyclic graph (Figure) reveals that Y1← U → Y2 may introduce bias into estimation of the causal relationship of interest, Y1→Y2, the effect of primary infection on recurrence. Conditioning on unmeasured individual factors by comparing observations during counterfactual risk periods from the same individual (Y1← U → Y2) permits unbiased inference of the effect of Y1. This intuition provides the basis for numerous self-matched designs (e.g. case-crossover, case-time control, and self-controlled case series), which have garnered increasing interest in epidemiology[97].

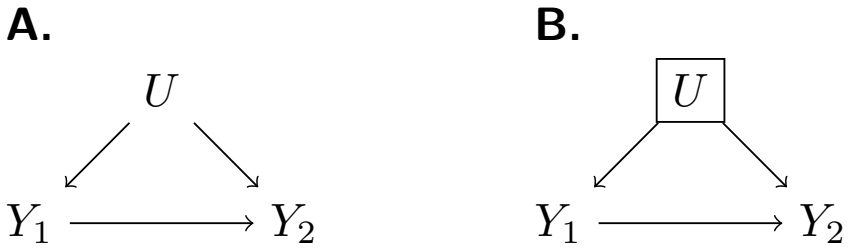

Figure 3.1: Directed acyclic graph addressing unmeasured confounding. We illustrate a causal framework wherein the effect of previous infection on time to subsequent infection (Y1→Y2) is of interest for analysis. One or more unmeasured confounding factors (U) creates a backdoor path (A) which can be blocked by conditioning on U (B).

In this chapter, we present an adaptation of these methods harnessing data from "negative control" events to permit causal inference in the presence of heterogeneous individual frailty. We derive a matched (Mantel-Haenszel) odds ratio (ORMH)[34, 84] estimator for the hazard ratio of infection or disease, given previous infection. We conduct simulations to compare this approach against alternative methods based on proportional hazards models common in the analysis of longitudinal data, and to assess statistical power under varying conditions. Last, we use the proposed method to reassess protective effects of rotavirus infection in data from previously-published birth-cohort studies[137, 52].

3.2 Approach

Self-matched negative control design

Parameter	Definition
λ_{Pi}	Rate at which individual i experiences the "outcome of interest", without naturally-acquired immunity
λ_{Ni}	Rate at which individual i experiences a negative-control outcome
θ	Hazard ratio for the outcome of interest, owing to naturally-acquired protection
$\frac{\beta_{Pi}(1)}{\beta_{Pi}(0)}$	Hazard ratio for the outcome of interest after primary infection, for individual i, due to all (confounding) factors other than naturally-acquired protection
$\frac{\beta_{Ni}(1)}{\beta_{Ni}(0)}$	Hazard ratio for the negative control outcome after primary infection, for individual i

Table 3.1: Parameters and Definitions

Consider an outcome such as acquisition of a pathogen of interest, or onset of disease due to this pathogen (Table 3.1). The proposed design only includes individuals who experience recurrent episodes of this outcome of interest (case-only). Define Y_i and X_i as variables indicating outcome and exposure status for individual i at each observation, with $Y_i = 1$ indicating infection or disease with the pathogen of interest and $Y_i = 0$ indicating a negative-control outcome. Consideration of negative-control observations is of interest for studies involving event-based data capture (e.g. episodes of acute illness), and provides a basis for a competing risks estimation framework as we detail below. Last, let $X_i = 1$ indicate an individual has previously experienced infection with the pathogen of interest, and with $X_i = 0$ indicating the individual has no history of infection with the pathogen of interest.

	Outcome of interest ($Y_i = 1$)	Negative control outcome ($Y_i = 0$)
Previously uninfected ($X_i = 0$)	$A_i \sim exp(\lambda_{Pi})$	$B_i \sim exp(\lambda_{Ni})$
Previously infected ($X_i = 1$)	$C_i \sim exp(\theta\lambda_{Pi})$	$D_i \sim exp(\lambda_{Ni})$

Table 3.2: Contingency table for event time distributions, given prior infection.

Define A_i to D_i as random variables indicating event times for observations of $Y_i = 1$ and $Y_i = 0$, conditioned on X_i, according to the contingency structure presented in Table

3.2. A_i and B_i are the time to first occurrence of the outcome of interest and the negative control outcome, respectively, for an individual with no history of infection ($X_i = 0$). C_i and D_i are the time to the first occurrence of the outcome of interest and the negative control outcome, respectively, following infection with the pathogen of interest (such that $X_i = 1$; Figure 3.2). Here we note that B_i and D_i are censored if $A_i < B_i$ and $C_i < D_i$, respectively.

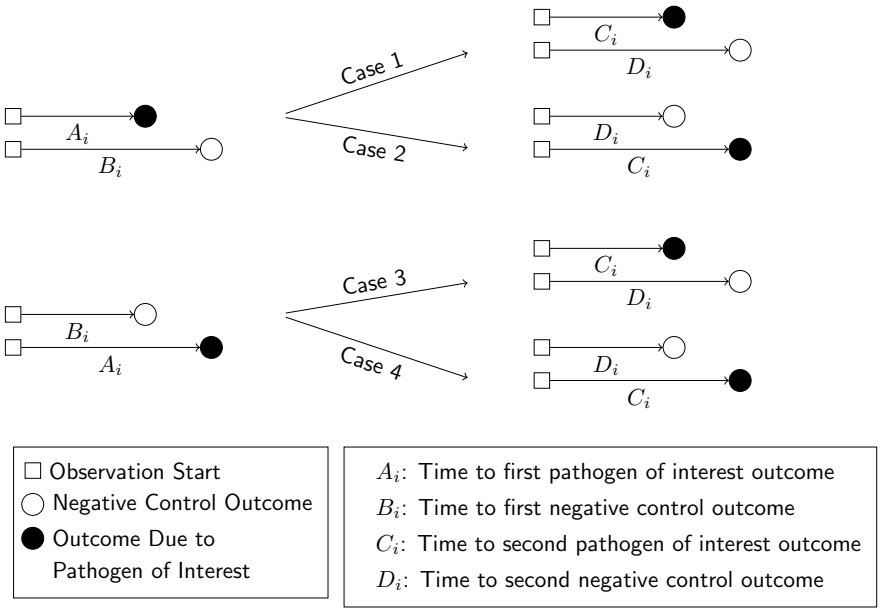

Figure 3.2: Schematic presentation of potential outcomes. We illustrate potential outcomes in terms of the sequence of events A_i-D_i for a given individual. In cases 1 and 2, we observe $A_i < B_i$ (truncating observation of a negative-control event while $X_i = 0$). In cases 3 and 4, we observe $B_i < A_i$, with the negative-control outcome preceding infection with the pathogen of interest. We illustrate the corresponding potential outcomes for C_i and D_i, when $X_i = 1$, in the right-hand side of the figure.

Event time distributions

Define the hazard for the outcome of interest (P) for individual i as λ_{Pi} , and define θ as the hazard ratio of this outcome given previous infection (Table 3.1). Assuming events occur independently and continuously during the follow up after conditioning on each individuals' hazard and infection history, event times are exponentially distributed at the individual level. We note this circumstance gives rise to a non-exponential event time distribution at the population level, whose variance is augmented (relative to the pure exponential case) by heterogeneity in individual frailty. The probability of the outcome by time t, for a previously-uninfected individual, is $1 - exp(-\lambda_{Pi}t)$, while the probability of the outcome by time t, had

the same individual counterfactually been previously infected, is $1-exp(-\theta\lambda_{Pi}t)$. We discuss alternative individual event time distributions in a later section.

Consider that data are collected from each individual for endpoints besides the primary outcome of interest. Among these, suppose a negative control outcome (N) occurs at a rate λ_{Ni} for individual i. This rate should be unaffected by individuals' prior exposure to the pathogen of interest, according to the definition of a negative control in this context[79]. Under the same assumptions, the probability of experiencing the negative control outcome by time t, for individual i, is $1 - exp(-\lambda_{Ni}t$.

Estimating the effect of naturally-acquired immunity

For an individual with no history of previous infection, consider the outcome of interest and the negative-control outcome to be competing risks. The events E_i to H_i may be defined to indicate the relative ordering of event times A_i to D_i according to the contingency structure presented in Table 3.3, for each individual i. Specifically, take $E_i = A_i \leq B_i$ and $G_i = C_i \leq D_i$ to indicate the outcome of interest precedes the negative-control outcome during the periods with $X_i = 0$ and $X_i = 1$, respectively. Define $F_i = B_i < A_i$ and $H_i = D_i < C_i$ as complements to E_i and G_i. Assuming event times are exponentially distributed after conditioning on each individual's unique hazard, we may start the clock for event times A_i and B_i from cohort entry or an alternative milestone (e.g. birth or exposure onset); for C_i and D_i we consider time from the most recent infection, although any subsequent milestone is likewise appropriate.

	Outcome of interest precedes negative control	Negative control outcome precedes outcome of interest
Previously uninfected ($X_i = 0$)	$E_i = A_i < B_i$	$F_i = B_i < A_i$
Previously infected ($X_i = 1$)	$G_i = C_i < D_i$	$H_i = D_i < C_i$

Table 3.3: Contingency table for competing risks, given prior infection.

To derive the probabilities $Pr(E_i), Pr(F_i), Pr(G_i), Pr(H_i)$, consider two competing, independent event times τ_j and τ_k occuring at rates r_j and r_k, we may obtain the probability for τ_j to precede τ_k from the improper integral

$$Pr(\tau_j < \tau_k) = \int_0^\infty Pr(\tau_k > t)Pr(\tau_j = t)dt \tag{3.1}$$

Then with the assumption of exponentially-distributed event times,

$$Pr(\tau_j < \tau_k) = \int_0^\infty exp(-r_k t)r_j exp(-r_j t)dt = \frac{r_j}{r_j + r_k} \tag{3.2}$$

Using this we can now write the probabilities for our events E_i through H_i:

$$Pr(E_i) = Pr(A_i \leq B_i) = \frac{\lambda_{Pi}}{\lambda_{Pi} + \lambda_{Ni}} \tag{3.3}$$

$$Pr(F_i) = Pr(B_i < A_i) = \frac{\lambda_{Ni}}{\lambda_{Pi} + \lambda_{Ni}} \tag{3.4}$$

$$Pr(G_i) = Pr(C_i \leq D_i) = \frac{\theta\lambda_{Pi}}{\theta\lambda_{Pi} + \lambda_{Ni}} \tag{3.5}$$

$$Pr(H_i) = Pr(D_i < C_i) = \frac{\lambda_{Ni}}{\theta\lambda_{Pi} + \lambda_{Ni}} \tag{3.6}$$

Consider the Mantel-Haenszel odds ratio[34, 84] constructed from the competing risks of $Y_i = 1$ and $Y_i = 0$, given X_i, matching observations from each individual i:

$$OR_{MH} = \frac{\sum_i I(F_i)I(G_i)}{\sum_i I(E_i)I(H_i)} \tag{3.7}$$

such that the expectation of the odds ratio is

$$E(OR_{MH}) = \frac{\sum_i Pr(F_i)Pr(G_i)}{\sum_i Pr(E_i)Pr(H_i)} \tag{3.8}$$

Using the above derivations of $Pr(E_i)$ through $Pr(H_i)$,

$$E(OR_{MH}) = \frac{\sum_i (\frac{\lambda_{Ni}}{\lambda_{Pi}+\lambda_{Ni}} \cdot \frac{\theta\lambda_{Pi}}{\theta\lambda_{Pi}+\lambda_{Ni}})}{\sum_i (\frac{\lambda_{Pi}}{\lambda_{Pi}+\lambda_{Ni}} \cdot \frac{\lambda_{Ni}}{\theta\lambda_{Pi}+\lambda_{Ni})}} \tag{3.9}$$

$$= \theta \frac{\sum_i (\frac{\lambda_{Ni}}{\lambda_{Pi}+\lambda_{Ni}} \cdot \frac{\lambda_{Pi}}{\theta\lambda_{Pi}+\lambda_{Ni}})}{\sum_i (\frac{\lambda_{Pi}}{\lambda_{Pi}+\lambda_{Ni}} \cdot \frac{\lambda_{Ni}}{\theta\lambda_{Pi}+\lambda_{Ni})}} \tag{3.10}$$

$$E(OR_{MH}) = \theta \tag{3.11}$$

In the event that observations end at some finite time δ, which is more in line with real world studies, equation 3.11 still holds. To see this consider the variation on 3.1 where we integrate to δ rather than infinity:

$$Pr(\tau_j < \tau_k) = \int_0^\delta Pr(\tau_k > t)Pr(\tau_j = t)dt = \frac{r_j(1 - exp[-\delta(r_k + r_j)])}{r_k + r_j} \tag{3.12}$$

This then gives us new expressions

$$Pr(E_i) = \frac{\lambda_{Pi}(1 - exp[-\delta(\lambda_{Pi} + \lambda_{Ni})])}{\lambda_{Pi} + \lambda_{Ni}} \tag{3.13}$$

$$Pr(F_i) = \frac{\lambda_{Ni}(1 - exp[-\delta(\lambda_{Pi} + \lambda_{Ni})])}{\lambda_{Pi} + \lambda_{Ni}} \tag{3.14}$$

$$Pr(G_i) = \frac{\theta\lambda_{Pi}(1 - exp[-\delta(\theta\lambda_{Pi} + \lambda_{Ni})])}{\theta\lambda_{Pi} + \lambda_{Ni}} \tag{3.15}$$

$$Pr(H_i) = \frac{\lambda_{Ni}(1 - exp[-\delta(\theta\lambda_{Pi} + \lambda_{Ni})])}{\theta\lambda_{Pi} + \lambda_{Ni}} \tag{3.16}$$

where the additional terms still cancel out in the matched odds ratio formulation:

$$E(OR_{MH}) = \frac{\sum_i Pr(F_i)Pr(G_i)}{\sum_i Pr(E_i)Pr(H_i)} \tag{3.17}$$

$$= \frac{\sum_i (\frac{\lambda_{Ni}(1-exp[-\delta(\lambda_{Pi}+\lambda_{Ni})])}{\lambda_{Pi}+\lambda_{Ni}} \cdot \frac{\theta\lambda_{Pi}(1-exp[-\delta(\theta\lambda_{Pi}+\lambda_{Ni})])}{\theta\lambda_{Pi}+\lambda_{Ni}})}{\sum_i (\frac{\lambda_{Pi}(1-exp[-\delta(\lambda_{Pi}+\lambda_{Ni})])}{\lambda_{Pi}+\lambda_{Ni}} \cdot \frac{\lambda_{Ni}(1-exp[-\delta(\theta\lambda_{Pi}+\lambda_{Ni})])}{\theta\lambda_{Pi}+\lambda_{Ni})}} \tag{3.18}$$

$$E(OR_{MH}) = \theta \tag{3.19}$$

Thus, in both the idealized case and under truncated observations, the ratio of the matched odds for the outcome of interest to precede a negative-control outcome, given individuals' unique hazards and history of prior infection, provides an unbiased estimate of the effect of previous infection on time to recurrence of the outcome of interest.

3.3 Comparison to cohort design using proportional hazards analysis

Simulation study

We conducted a simulation study across various underlying distributions of λ_{Pi} and λ_{Ni} to test for bias of point estimates under the proposed approach and under alternative methods often used in the analysis of cohort study data (without consideration of negative controls). As comparisons, we considered several proportional hazards models which could be applied to time-to-event data for recurrent observations of the outcome of interest. We considered four approaches to control for differences in hazard among individuals:

1. "Naïve" proportional hazards model without inclusion of additional terms to account for differences in event times among individuals. We define the hazard ratio estimated via fitting this model as θ_{Naive}.

2. Proportional hazards model accounting for variation in individual frailty via "random effects". Fitting this model estimates the hazard ratio $\hat{\theta}_{RE}$ for the effect of previous infection, as well as $\hat{\sigma}^2$ representing the estimated variance in (log) individual-specific event rates, assumed to represent independent draws from a Normal distribution with mean 0[101, 129].

3. Proportional hazards model including Gamma-distributed frailty terms[136]. Fitting this model estimates the hazard ratio $\hat{\theta}_{Frailty}$ for the effect of previous infection, along with the parameters of the underlying Gamma distribution describing individual-specific frailties.

4. Proportional hazards model with "fixed effects" for individual subjects. Fitting this model estimates a hazard ratio $\hat{\theta}_{FE}$ for the effect of previous infection and estimates subject-specific rates of infection (via individual-specific intercepts) which have no pre-specified distributional assumption.

We defined $\hat{\theta}_{MH} = OR_{MH}$ for the proposed analysis strategy of a self-matched, negative control design and considered various distributions for λ_{Pi}:

1. Truncated Normal distribution (with a pre-specified lower bound at $a = 0$);

2. Truncated Cauchy distribution (with a pre-specified lower bound at $a = 0$);

3. Uniform distribution;

4. Gamma distribution;

5. Mixtures of Gamma distributions.

We considered multiple parameterizations of each of these distributions (Table 3.4), holding the mean rate (or location parameter of the Cauchy distribution) constant at one infection per year across all simulations to determine effects of inter-individual heterogeneity on estimates of θ. We illustrate the distributions in Figure 3.3. Considering cohorts of 500 individuals, we drew λ_{Pi} values at random and sampled exponentially-distributed event times of first and second infections for each individual, truncating observations at five years. We repeated simulations 500 times for each $\theta \in \{0.01, 0.02, ..., 0.99\}$, drawing λ_{Pi} values independently for each simulation. We used the simulated datasets to estimate $\hat{\theta}_{Naive}, \hat{\theta}_{RE}, \hat{\theta}_{FE}, and \hat{\theta}_{Frailty}$ taking the average of estimates obtained across all 500 iterations to obtain a single point estimate for each parameterization.

To compute $\hat{\theta}_{MH}$, we drew hazards (λ_{Pi}) and event times for negative control observations from each subject, assuming event times were exponentially-distributed with respect to the sampled rate parameters. To standardize comparisons of $\hat{\theta}_{MH}$ under differing distributions of λ_{Pi}, we defined $\lambda_{Ni} = 1 \forall i$ under each simulation.

To investigate how the different modeling frameworks performed in capturing the distribution of individual-specific hazards, we saved estimates of individual-specific fixed effects,

random effects, and frailties alongside estimates of $\hat{\theta}$. We fitted a single density kernel to the distribution of individual-specific estimates across 10 simulated cohorts for each true value of θ and underlying distribution of λ_{Pi}.

Distribution	Parameters	I	II	III	IV	V
Truncated Normal	Mean μ	$\mu = 1$	$\mu = 1$	$\mu = 1$	$\mu = 1$	$\mu = 1$
	Variance σ^2	$\sigma^2 = 1/4$	$\sigma^2 = 1/2$	$\sigma^2 = 1$	$\sigma^2 = 2$	$\sigma^2 = 4$
	Lower bound a	$a = 0$	$a = 0$	$a = 0$	$a = 0$	$a = 0$
	Upper bound b	$b = \infty$	$b = \infty$	$b = \infty$	$b = \infty$	$b = \infty$
Truncated Normal	Location x_0	$x_0 = 1$	$x_0 = 1$	$x_0 = 1$	$x_0 = 1$	$x_0 = 1$
	Scale γ	$\gamma = 1/8$	$\gamma = 1/4$	$\gamma = 1$	$\gamma = 4$	$\gamma = 8$
	Lower bound a	$a = 0$	$a = 0$	$a = 0$	$a = 0$	$a = 0$
	Upper bound b	$b = \infty$	$b = \infty$	$b = \infty$	$b = \infty$	$b = \infty$
Uniform	Lower bound a	$a = 7/8$	$a = 3/4$	$a = 1/2$	$a = 1/4$	$a = 0$
	Upper bound b	$b = 9/8$	$b = 5/4$	$b = 3/2$	$b = 7/4$	$b = 2$
Gamma	Shape k	$k = 8$	$k = 4$	$k = 1$	$k = 1/4$	$k = 1/8$
	Scale θ	$\theta = 1/8$	$\theta = 1/4$	$\theta = 1$	$\theta = 4$	$\theta = 8$
Gamma mixture (i)	Shape 1 k_1	$k_1 = 1/8$	$k_1 = 1/8$	$k_1 = 1/8$	$k_1 = 1/8$	$k_1 = 1/8$
	Scale 1 θ_1	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$
	Shape 2 k_2	$k_2 = 3/8$	$k_2 = 15/8$	$k_2 = 15/16$	$k_2 = 3/32$	$k_2 = 3/320$
	Scale 2 θ_2	$\theta_2 = 1/5$	$\theta_2 = 1$	$\theta_2 = 2$	$\theta_2 = 20$	$\theta_2 = 200$
Gamma mixture (ii)	Shape 1 k_1	$k_1 = 1/2$	$k_1 = 1/2$	$k_1 = 1/2$	$k_1 = 1/2$	$k_1 = 1/2$
	Scale 1 θ_1	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$
	Shape 2 k_2	$k_2 = 3/10$	$k_2 = 3/2$	$k_2 = 3$	$k_2 = 3/10$	$k_2 = 3/10$
	Scale 2 θ_2	$\theta_2 = 1/5$	$\theta_2 = 1$	$\theta_2 = 2$	$\theta_2 = 20$	$\theta_2 = 200$

Table 3.4: Event rate distributions applied to simulation study. Parameterizations are listed in order of increasing variance from I to V. Gamma mixture distributions are weighted to be half from each distribution.

Results

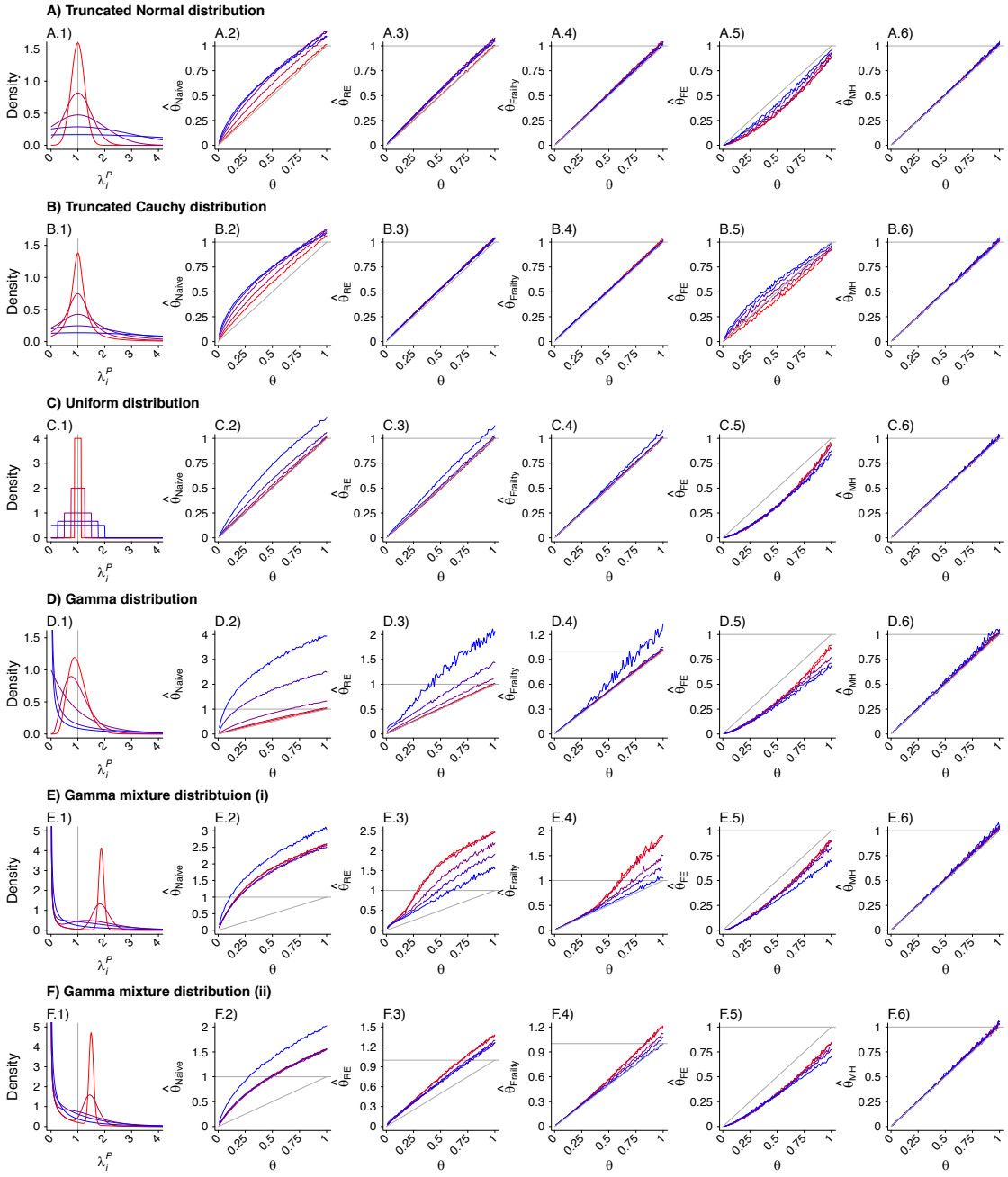

Figure 3.3: Caption continued on next page

In figure 3.3 we show simulated distributions and hazard ratio estimates under "naïve" inference approaches and under the proposed approach of self-matched inference with negative controls. Panels are organized to present, in each row, the assumed distribution (column 1), the estimate $\hat{\theta}_{Naive}$ based on a Cox proportional hazards model without any correction for inter-individual heterogeneity (column 2), proportional hazards models employing various frailty frameworks (columns 3-5), and the estimate $\hat{\theta}_{MH}$ based on the proposed approach (column 6). One-to-one lines plotted in grey in columns 2-6 indicate where estimates would recover the true value, i.e. $\hat{\theta} = \theta$. Horizontal grey lines plotted at $\hat{\theta} = 1$ indicate where estimates exceed 1, indicating directionally-misspecified estimates of the causal effect of interest. Values are plotted on a red-to-blue color ramp corresponding to the parameterizations I-V, respectively, in order of least (I; red) to greatest (V; blue) variance as detailed in Table 3.4. A) Truncated Normal distribution; B) Truncated Cauchy distribution; C) Uniform distribution; D) Gamma distribution; E) Mixture of Gamma distributions (i) with means at 0.125 and 1.875; and F) Mixture of Gamma distributions (ii) with means at 0.5 and 1.5.

We plot distributions and estimates under each approach in Figure 3.3 as described above. The naïve hazards ratio tended to overestimate θ, leading to under-estimation of the degree of protection $(1 - \theta)$. Bias was minimized as θ approached zero, consistent with a scenario of strong protective immunity. Values of $\hat{\theta}_{Naive}$ often exceeded 1 in scenarios where $\theta < 1$; in practice, such an estimate would lead to inference that prior infection increases susceptibility to infection or disease due to the pathogen of interest, when in fact prior infection is protective. For all distributions considered, bias in $\hat{\theta}_{Naive}$ was greatest under parameterizations yielding the highest between-individual variance in λ_{Pi}.

Alternative methods performed variably under the differing conditions (Figure 3.3). Lower degrees of bias were evident in $\hat{\theta}_{MH}$ as compared to estimates generated under the other methods assessed. Gamma frailty models and random effects models tended to yield less-biased estimates of θ than $\hat{\theta}_{Naive}$. However, the same direction of bias (resulting in under-estimation of the reduction in susceptibility, or $\hat{\theta} > \theta$) was evident with all three of these approaches. Bias was worst when λ_{Pi} values were drawn from Gamma or Gamma mixture distributions, and tended to increase under distributions with greater variance in λ_{Pi}, or greater irregularity in the case of Gamma mixture distributions. In contrast, fixed-effects models estimating multipliers on hazards for each individual tended to under-estimate θ under most distributions of λ_{Pi}, although both $\hat{\theta}_{FE} > \theta$ and $\hat{\theta}_{FE} < \theta$ were apparent in simulations using the truncated Cauchy distribution for λ_{Pi}. For the truncated Normal distribution, bias in $\hat{\theta}_{FE}$ decreased with greater variance in λ_{Pi}, whereas for the Uniform, Gamma, and Gamma mixture distributions, bias increased with greater variance in λ_{Pi}.

Biased estimation of θ occur in connection with a failure to accurately recover the underlying individual-specific frailty distributions. For each modeling approach, the extent of this misspecification in individual frailties varied over values of θ and distributions of λ_{Pi}. This misspecification is quantified in figures 3.4, 3.5, and 3.6

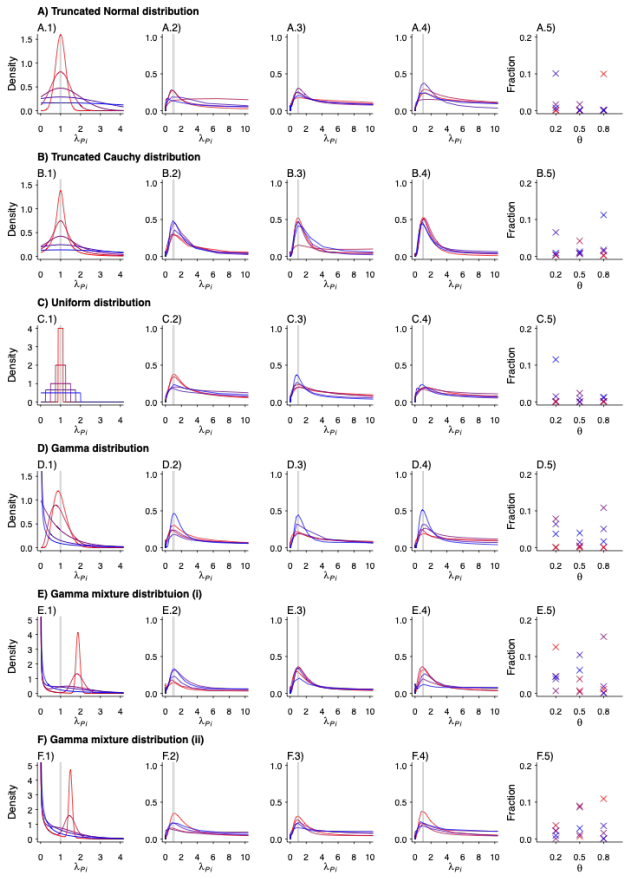

Figure 3.4: Estimated Density Kernels For Fixed Effects Model. In each row, column 1 is a reproduction of the densities used to produce the individual λ_{Pi}. Columns 2-4 are the estimated density kernels for $\theta = 0.2$, $\theta = 0.5$, $\theta = 0.8$, respectively. Column 5 shows the proportion of estimates which were over 1000 for each set of parameters. Values are plotted on a red-to-blue color ramp corresponding to the parameterizations I-V, respectively, in order of least (I; red) to greatest (V; blue) variance as detailed in Table 3.4. A) Truncated Normal distribution; B) Truncated Cauchy distribution; C) Uniform distribution; D) Gamma distribution; E) Mixture of Gamma distributions (i) with means at 0.125 and 1.875; and F) Mixture of Gamma distributions (ii) with means at 0.5 and 1.5.

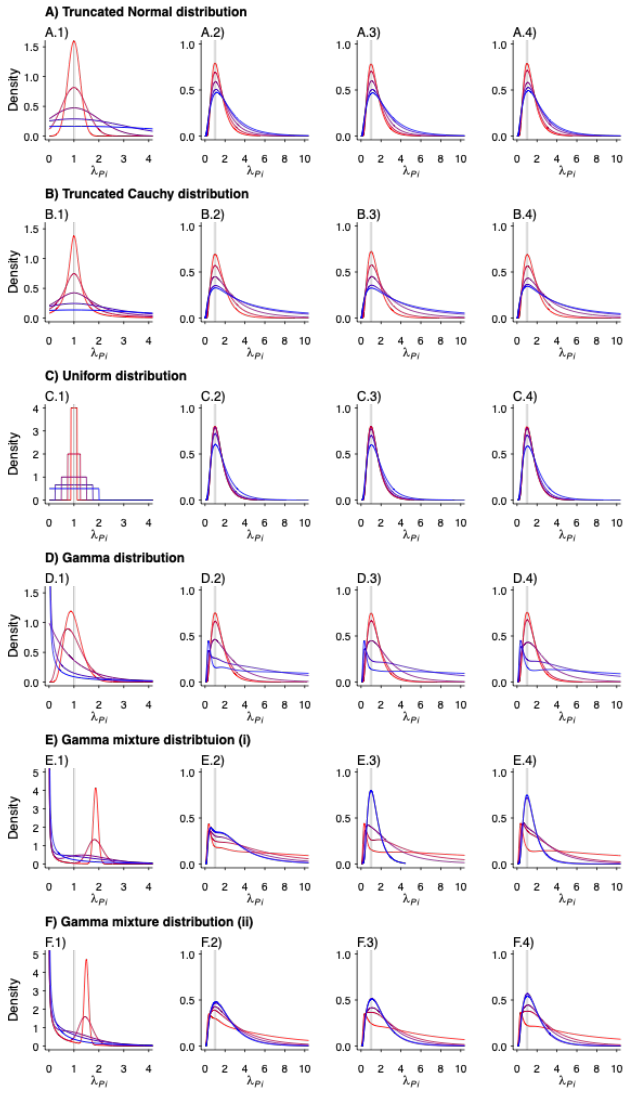

Figure 3.5: Estimated Density Kernels For Random Effects Model. In each row, column 1 is a reproduction of the densities used to produce the individual λ_{Pi}. Columns 2-4 are the estimated density kernels for $\theta = 0.2$, $\theta = 0.5$, $\theta = 0.8$, respectively. Column 5 shows the proportion of estimates which were over 1000 for each set of parameters. Values are plotted on a red-to-blue color ramp corresponding to the parameterizations I-V, respectively, in order of least (I; red) to greatest (V; blue) variance as detailed in Table 3.4. A) Truncated Normal distribution; B) Truncated Cauchy distribution; C) Uniform distribution; D) Gamma distribution; E) Mixture of Gamma distributions (i) with means at 0.125 and 1.875; and F) Mixture of Gamma distributions (ii) with means at 0.5 and 1.5.

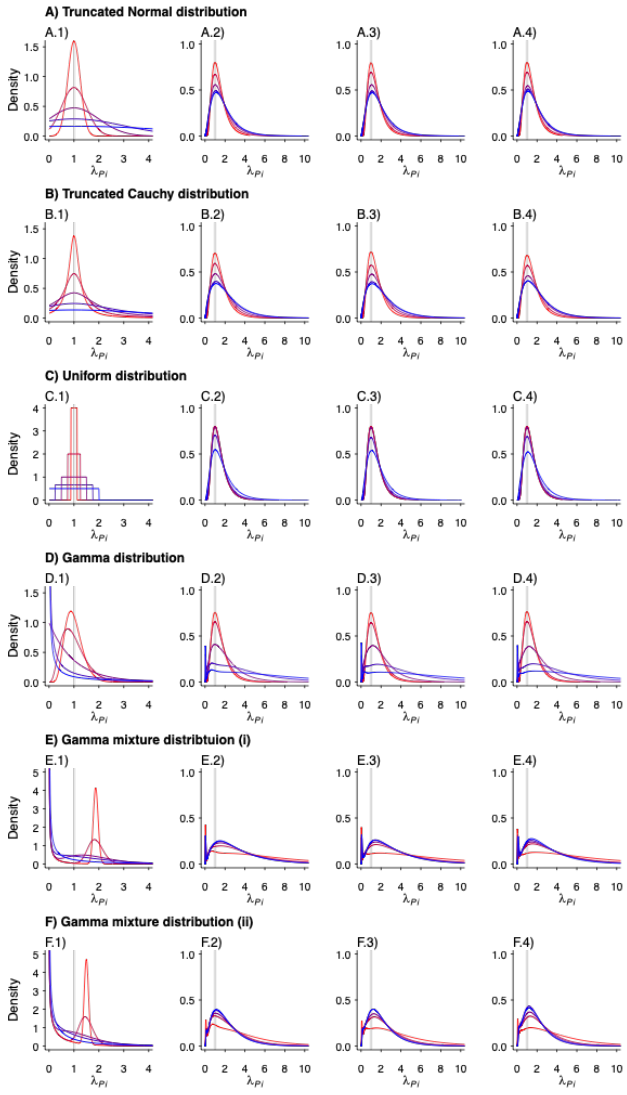

Figure 3.6: Estimated Density Kernels For Gamma Frailty Model. In each row, column 1 is a reproduction of the densities used to produce the individual λ_{Pi}. Columns 2-4 are the estimated density kernels for $\theta = 0.2$, $\theta = 0.5$, $\theta = 0.8$, respectively. Column 5 shows the proportion of estimates which were over 1000 for each set of parameters. Values are plotted on a red-to-blue color ramp corresponding to the parameterizations I-V, respectively, in order of least (I; red) to greatest (V; blue) variance as detailed in Table 3.4. A) Truncated Normal distribution; B) Truncated Cauchy distribution; C) Uniform distribution; D) Gamma distribution; E) Mixture of Gamma distributions (i) with means at 0.125 and 1.875; and F) Mixture of Gamma distributions (ii) with means at 0.5 and 1.5.

3.4 Sample size considerations

Simulation study

To inform applications of the proposed method, we next assessed statistical power under differing conditions. A test statistic (ξ_{MH}) has previously been identified for OR_{MH} under the null hypothesis of no difference in risk given exposure[83]. For the contingency structure (Table 3) formulated from the terms E_i to H_i (by which we define OR_{MH}), this statistic can be written generally as

$$\xi_{MH} = \frac{(\sum_i^N (E_i - \frac{(E_i+F_i)(E_i+G_i)}{2}))^2}{\sum_i^N \frac{(E_i+F_i)(E_i+G_i)(F_i+H_i)(G_i+H_i)}{4}} \tag{3.20}$$

which can then be simplified according to $E_i + F_i = 1, G_i + H_i = 1$, and $F_i + H_i = 2 - E_i - G_i$. Thus,

$$\xi_{MH} = \frac{(\sum_i^N (E_i - G_i))^2}{\sum_i^N (E_i + G_i)(2 - E_i - G_i)} \tag{3.21}$$

which is expected to follow a χ^2 distribution with one degree of freedom under the null hypothesis. We calculated values of ξ_{MH} obtained for cohorts of varying sizes under differing parameterizations of θ, λ_{Pi}, and λ_{Ni}. For values of $\theta \in \{0.1, 0.2, ..., 0.9\}$, we sampled individual event times A_i to D_i for a population of 100,000 individuals whom we subsequently partitioned (without replacement) into 2000 hypothetical study cohorts each of size N=25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, and 1000. For these analyses, we considered λ_{Pi} values drawn from truncated Normal, Gamma, and Gamma mixture distributions, under the parameterizations of each of these distributions with greatest and least variance listed in Table 3.4. We determined statistical power via the proportion of simulated cohorts for which the upper bound of a 95% confidence interval around OR_{MH} would be expected to correctly exclude the null value, i.e. $Pr(\xi_{MH} > \xi_{MH}^{97.5\%} = 5.02)$.

To assess how correlation between λ_{Pi}, and λ_{Ni} could affect the statistical power of estimates, we conducted simulations under two sets of assumptions. Under the first, we considered $\lambda_{Ni} = 1$ for all i (equal to the expected value of λ_{Pi} under all parameterizations), so that $\lambda_{Ni} \perp \lambda_{Pi}$; under the second, we defined λ_{Ni}=λ_{Pi}, under the assumption that individuals with greater risk of the outcome of interest would also experience higher incidence of the negative control condition. These conditions bound power estimates, corresponding to assuming no correlation and perfect correlation between λ_{Ni} and λ_{Pi}, respectively.

Results

We present results of the power analyses in Figure 4. Analyses with as few as 50 subjects had roughly 80% power or greater to estimate $\theta = 0.1$ (corresponding to 90% protection) under all conditions explored; analyses with 500 subjects had 80% power or greater to

estimate $\theta \leq 0.5$ (corresponding to 50% protection or greater) under all conditions. No scenarios revealed 80% or greater power for estimation of $\theta \geq 0.8$ (corresponding to less than 20% protection), even with 1000 subjects; statistical power for estimation of $\theta = 0.9$ was 10% or lower under nearly all conditions explored. This is a limitation of the method; however, detecting θ slightly less than 1 presents weak motivation to investigate a pathogen for vaccine development as there is little naturally acquired immunity.

For simulations with $\lambda_{Ni} \perp \lambda_{Pi}$, statistical power was weaker under parameterizations resulting in greater variance in λ_{Pi}. In contrast, for simulations with $\lambda_{Ni} = \lambda_{Pi}$, differences in statistical power were less strongly apparent with increasing variance in λ_{Pi}. Taken together, these findings suggest statistical power is maximized when negative control endpoints are chosen which tend to occur more commonly among individuals who are at greatest risk of the outcome of interest.

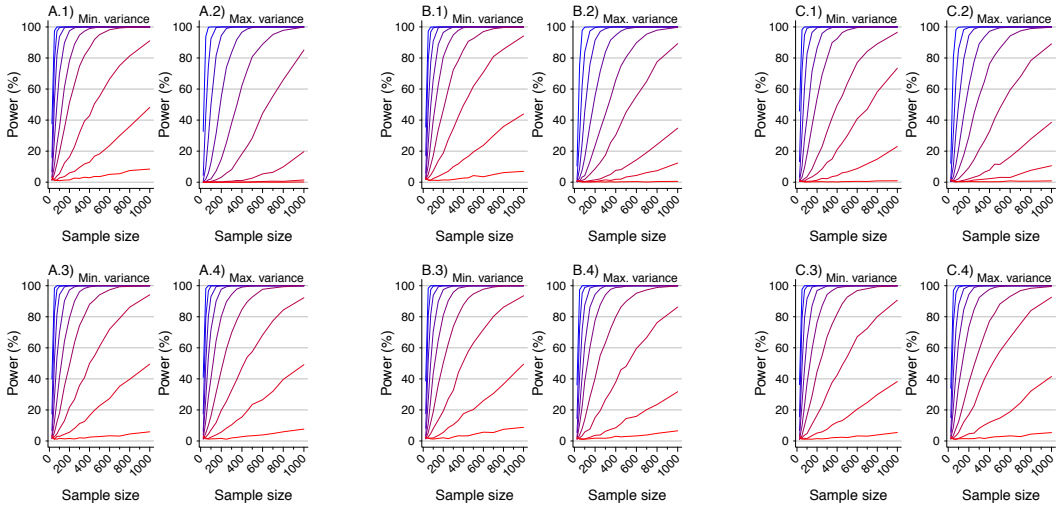

Figure 3.7: Statistical power for simulated analyses using the proposed approach of self-matched inference via negative controls. Each panel presents the statistical power for rejecting the null hypothesis with two-sided p¡0.05 under varying conditions. Lines plotted in red to blue correspond to decreasing values of θ: 0.9 (red), 0.8, 0.7, ..., 0.1 (blue), corresponding to increasing protection from 10% to 90%. Plots are presented in groups of 4 panels, each corresponding to analyses with values drawn from the following distributions: A) Truncated Normal distribution; B) Gamma distribution; C) Mixture of Gamma distributions with means at 0.125 and 1.875 (as detailed in Table 3.4). Panels in the top row (A.1, A.2, B.1, B.2, C.1, C.2) represent analyses in which no correlation is assumed between rates of the outcome of interest and negative control outcome ($\lambda_{Ni} = 1 \forall i$). Panels in the bottom row (A.3, A.4, B.3, B.4, C.3, C.4) represent analyses in which the correlation between rates of the outcome of interest and the negative control outcome are is maximized ($\lambda_{Ni} = \lambda_{Pi}$). Within each grouping, panels on the left-hand side (A.1, A.3, B.1, B.3, C.1, C.3) correspond to distributions with the least variance in individual rates of the outcome of interest (λ_{Pi}; i.e., parameterization I in Table 3.4). Panels on the right-hand side within each grouping (A.2, A.4, B.2, B.4, C.2, C.4) correspond to distributions with the greatest variance in individual rates of the outcome of interest (λ_{Pi}; i.e., parameterization V in Table 3.4).

3.5 Potential Introduction of Bias

At a design level, self-matched inference reduces or eliminates the potential for bias due to time-invariant factors that individually influence risk[145]. However, complications arise when individuals' risk of experiencing these endpoints differs over time.

Consider continuously varying hazards $\lambda_{Pi}(t)$ and $\lambda_{Ni}(t)$. When variation over time is identical for both disease of interest and negative control events (e.g. due to shared

seasonal patterns, or exposure to interventions affecting both conditions equally), $E(OR_{MH})$ is unaffected. Formally, we express this scenario as

$$\lambda_{Pi}(t) = \lambda_{Pi} f(t) \tag{3.22}$$

$$\lambda_{Ni}(t) = \lambda_{Ni} f(t) \tag{3.23}$$

Dividing the observation period into small windows of length dt, where the hazard is approximately constant (i.e. the probability of a negative control event happening is $\lambda_{Ni}(t)dt + o(dt)$), we consider the windows to be multinomial trials, where

$$P(A_i \leq B_i) = \frac{\lambda_{Pi}(t)dt}{\lambda_{Pi}(t)dt + \lambda_{Ni}(t)dt} = \frac{\lambda_{Pi}(t)}{\lambda_{Pi}(t) + \lambda_{Ni}(t)} \tag{3.24}$$

This circumstance motivates the selection of negative-control outcomes which share seasonal patterns with the outcome of interest. We illustrate bias that may occur when hazards do not vary synchronously in Figure 3.8, identifying the greatest bias under conditions of inverted seasonal patterns in relation to one another.

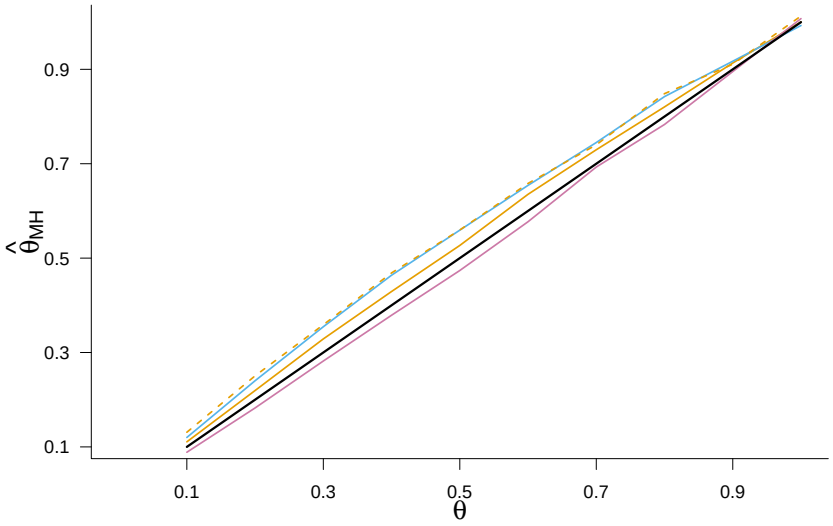

Figure 3.8: Seasonal variation introduces bias. We plot the estimated values of $\hat{\theta}$ for a range of values of θ, under various seasonal conditions. Seasonal disease of interest but not negative control is plotted in pink. INverse seasonality of the two events is plotted in blue, with the larger seasonal effect shown using a dotted line. The orange line shows a case when the two seasonal effects have different periods. To implement seasonality, the base hazard was multiplied by a periodic seasonal effect function with a period of one year

Broadly, we may interpret these findings to support the selection of negative-control outcomes that resemble the outcome of interest in their association with other time-varying confounders such as individuals' age, health status, and sociodemographic exposures.

We may also relax the assumption that event times are exponentially-distributed, conditional on individual hazards and exposures. Here we are centrally concerned with expressions for $P(A_i < B_i)$ under alternative distributions, as $P(A_i < B_i) = \frac{\lambda_{Pi}}{\lambda_{Pi}+\lambda_{Ni}}$is key to cancelling out the individual effects. This probability for gamma-distributed (k, θ) events is the regularized incomplete beta function $IB(\frac{\theta_A}{\theta_A+\theta_B})$ while for Weibull-distributed event times it is $1/(1+(\frac{\lambda_{Ni}}{\lambda_{Pi}})k)$ where k is the shared shape parameter of the distributions. Discrepancies arise in both cases when the distributions are parameterized differently (Figures 3.9 and 3.10).

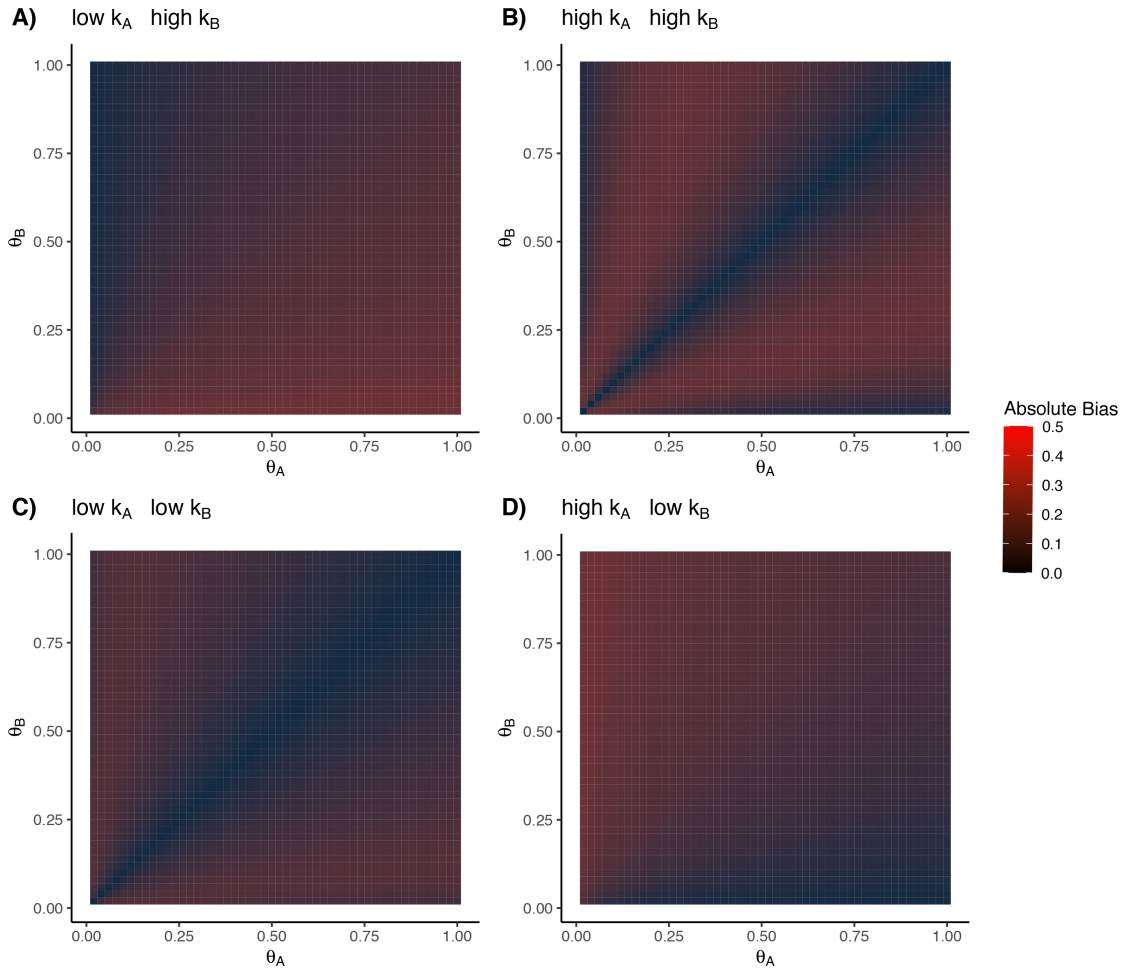

Figure 3.9: Bias of Gamma Distributed Event Time Assumption. Here we plot the absolute difference between $P(A_i < B_i)$ for the exponential event times and gamma event times. Larger discrepancies are shown as more red, and greater concordance as more black. We show four pairs of shape where k_A and k_B are either low (equal to 0.5) or high (equal to 3) in the four panels. Each cell is filled according to the absolute value of the difference between $\frac{k_A\theta_A}{k_A\theta_A+k_B\theta_B}$ and the regularized incomplete beta function $IB(\frac{\theta_A}{\theta_A+\theta_B}, k_B, k_A)$.

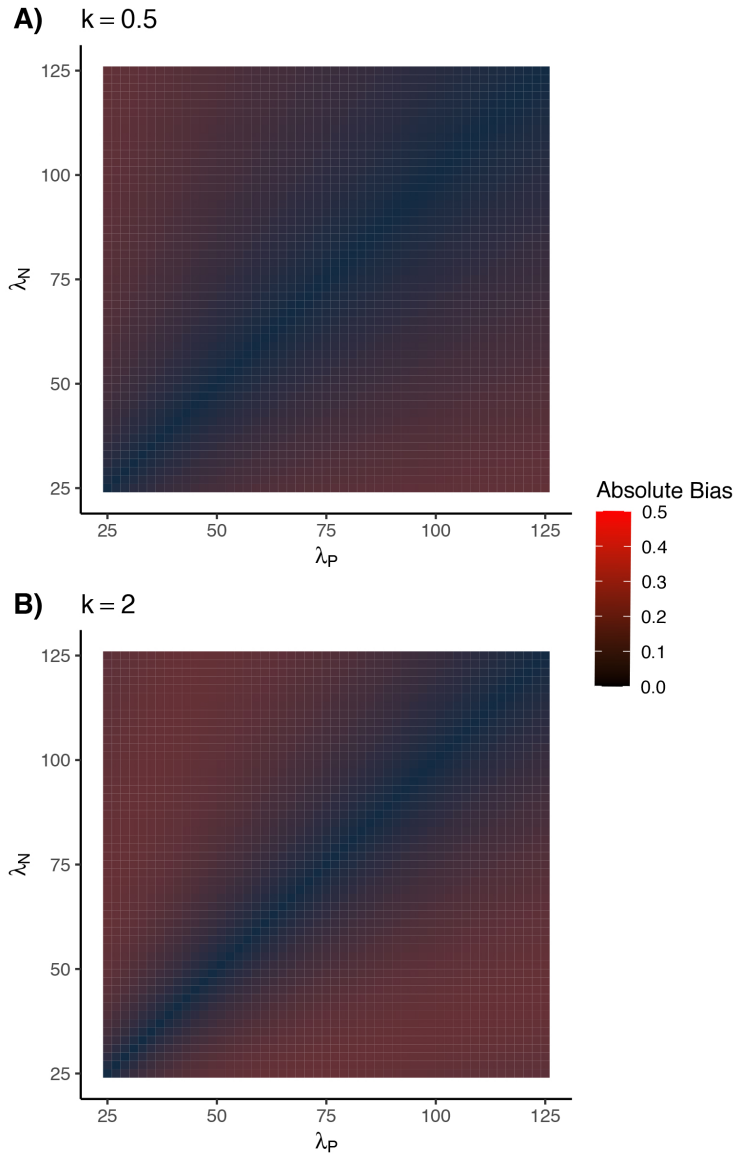

Figure 3.10: Bias of Weibull Distributed Event Time Assumption. Here we plot the absolute difference between $P(A_i < B_i)$ for the exponential event times and Weibull event times. Larger discrepancies are shown as more red, and greater concordance as more black. In A) the shape of the Weibull distributions $k = 0.5$, and in B) $k = 2$. Cells are filled according to the absolute difference between $\frac{\lambda_P}{\lambda_P+\lambda_N}$ and $\frac{1}{1+\frac{\lambda_N}{\lambda_P}^k}$

3.6 Application to Rotavirus Birth Cohort Data

Last, we applied the proposed method to real-world data collected in two birth-cohort studies of rotavirus infection and disease among 200 children in Mexico City, Mexico and 373 children in Vellore, India. These datasets have been described extensively in primary study publications[137, 52] and subsequent re-analyses [76, 78]. In these studies, pregnant mothers were enrolled prior to childbirth, and children were followed from birth to ages 2 years (in Mexico City) and 3 years (in Vellore). Investigators aimed to identify all rotavirus infections through routine testing of asymptomatic stool specimens (collected by field workers at regular home visits) for rotavirus, and by monitoring children for anti-rotavirus seroconversion over serial blood draws at scheduled intervals. Active surveillance was undertaken for all cases of gastroenteritis among children to characterize symptoms and test diarrheal stool specimens for rotavirus.

Initial analyses of the datasets gave differing conclusions about the strength of protection against rotavirus gastroenteritis (RVGE). Based on Cox proportional hazards models that did not account for variation in individual frailty, children in Mexico City were estimated to have experienced 77% (95% confidence interval: 60-88%), 83% (64-92%), and 92% (44-99%) lower rates of RVGE following one, two, and three previous infections, respectively, as compared to zero infections[137]. In contrast, children in Vellore, where the rate of rotavirus acquisition was higher, were estimated to have experienced 43% (24-56%), 71% (59-80%), and 81% (69-88%) lower rates of RVGE after one, two, and three previous infections, as compared to zero infections, based on a parametric Exponential survival model allowing Gamma frailty terms[52]. Subsequent analyses of the datasets revealed substantial variation in rates of rotavirus infection and risk of RVGE among individual children, as well as a potential for confounding due to declining risk of RVGE when infections were acquired at older ages, irrespective of previous infection[78]. In contrast, model-based analyses accounting for the independent effects of age and previous infection on children's susceptibility to RVGE estimated that children experienced 33% (23-41%), 50% (42-57%), and 64% (55-70%) lower rates of RVGE after one, two, and three previous infections, respectively, as compared to zero infections[76].

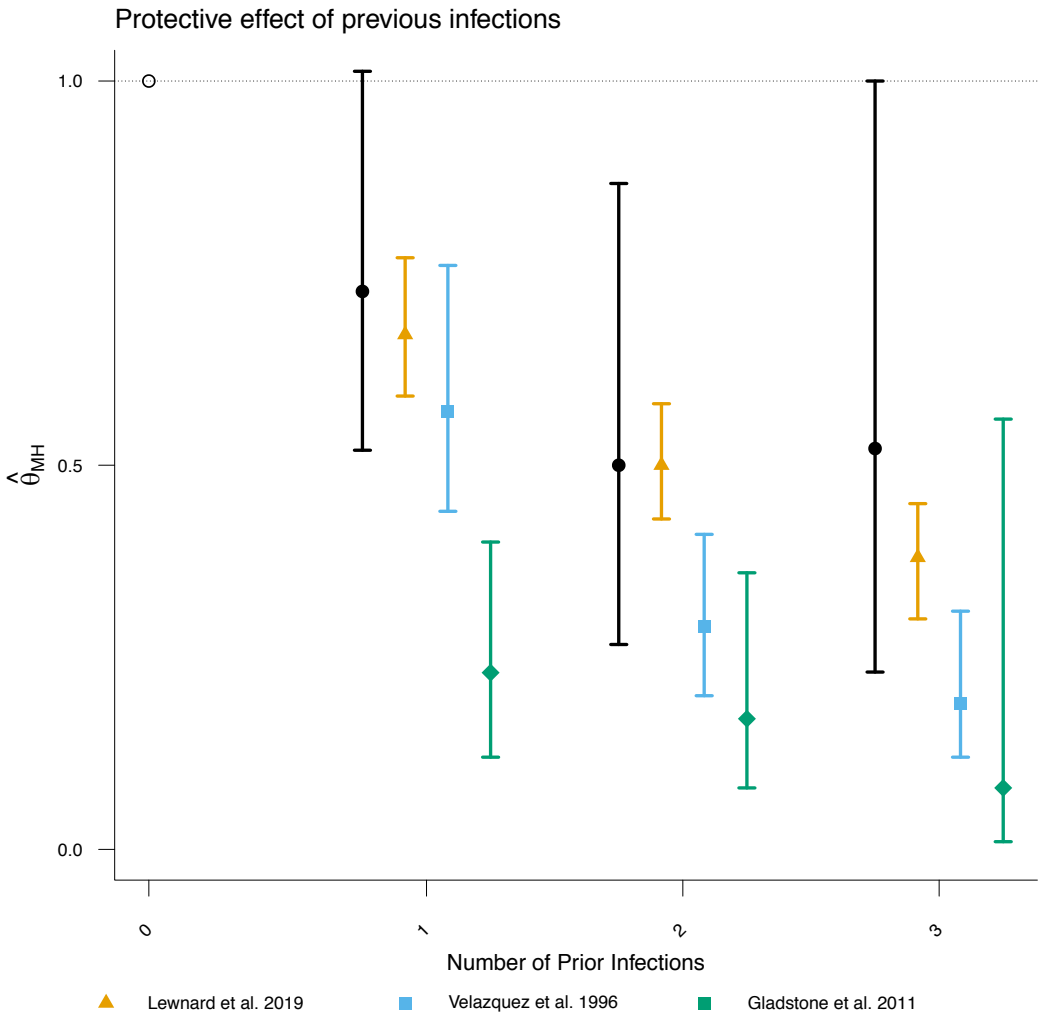

Figure 3.11: Estimated protection against rotavirus gastroenteritis associated with previous infection. We plot point estimates and 95% confidence intervals (lines) for estimates of the hazard ratio of rotavirus gastroenteritis associated with having previously experienced one, two, and three previous infections, versus zero previous infections, estimated via re-analysis of the Mexico City and Vellore rotavirus birth cohort studies[137, 52], Analyses include rotavirus-negative diarrhea occurrences as a negative control endpoint.

We used the proposed self-matched negative control design to re-estimate naturally-acquired protection against RVGE in the cohort datasets. Here, RVGE episodes (acute, new-onset diarrhea with rotavirus detected in the stool) are the outcome of interest and

acute, new-onset diarrhea episodes without rotavirus detection as the negative control. We compared the times of RVGE and rotavirus-negative diarrhea episodes from each child beginning from birth, and thereafter following detection of the first, second, and third rotavirus infection (generating confidence intervals via resampling of individual children). This yielded estimates of 27% (–1-48%), 50% (13-73%), and 48% (0-77%) lower rates of RVGE following one, two, and three previous infections, as compared to zero infections (Figure 5). Notwithstanding lower statistical power for the proposed method, these estimates are in agreement with previous findings15 suggesting lower strength of naturally-acquired protection than what was estimated in initial analyses of the birth cohort studies[137, 52].

3.7 Discussion

We propose a novel self-matched negative control method for estimating the hazard ratio of time to infection or disease due to a pathogen of interest, given previous infection. Analytically and via simulation, we show this method recovers unbiased estimates under a range of conditions, including when individual hazards are drawn from highly irregular or skewed distributions. We find these irregular or skewed distributions may lead to bias under proportional hazards models with commonly-used frailty estimation frameworks. Desirably, the proposed approach requires no parametric assumptions other than events occurring continuously and independently after conditioning on their underlying, individual-specific rates of occurrence. Beyond infectious disease natural history studies, this approach may have value for assessing the effects of other exposures on recurrent event times.

Our findings provide several practical insights for real-world longitudinal cohort studies. Collecting data on multiple endpoints affords the opportunity to leverage negative-control observations to support causal inference. For studies applying the proposed approach, negative-control endpoints affected by the same risk factors or exposures as the outcome of interest are desirable both to reduce potential risks of confounding due to time-varying factors, and to maximize statistical power based on the correlation between event rates for the outcome of interest and negative-control outcome, λ_{Pi} and λ_{Ni}. "Test-negative" control conditions which resemble the outcome of interest, but are not attributable to the same pathogen[77, 126], may provide a compelling choice, particularly if their occurrence is predicted by similar risk factors. For instance, shared risk factors are well-documented for rotavirus-positive and rotavirus-negative diarrhea[76, 65]. Considering respiratory illness, multiple etiologic viruses may share similar seasonal transmission patterns[93], routes of transmission via high-risk contact[96], and associations of disease progression or severity with host comorbidities[47]. For sexually-transmitted infections, particular risk behaviors[49] differing among individuals or over time could alter risk of any infection, rather than infection with the pathogen of interest alone[107]. In the context of real-world cohort studies, test-negative control conditions which are clinically similar to the outcome of interest would likely result in a study visit or other recorded interaction with similar probability. This further supports consideration of inference methods making use of test-positive and test-negative occurrences of a particular

clinical syndrome.

In summary, self-matched inference via negative controls may provide a flexible strategy to circumvent bias introduced by variation in individual frailty for analyses of naturally-acquired immunity. Applications to other exposures affecting the distribution of recurrent event times merit consideration, given the possible limitations we identify in existing analysis frameworks.

Acknowledgments and additional information

This work has been published in Epidemiology [98], please refer to the paper for additional information and for code to reproduce the figures seen in this chapter. I would like to thank Dr. Lei Quan, Dr. Katia Bruxvoort, Dr. Florian Marx, Dr. Lilith Whittles, and Dr. Joseph Lewnard for their input and contributions which helped improve this work.

Chapter 4

Concluding thoughts

This dissertation combines three chapters of work united by their ability to shed light on complicated infectious disease processes with the use of straightforward mechanistic modeling and statistical techniques. Differential equation modeling and odds ratio frameworks are not new approaches in the field, but this work shows that there are still many applications of them that have gone lacking. It is crucial to continue to strive to understand our world through such lenses as new approaches may be perhaps more complex but obfuscate the key details that allow us to truly understand the core ideas of certain systems.

Chapter one details a general model of hyperparasitism that relaxes several key assumptions made in prior models. With this simple model, an adaptive dynamics approach can be taken to look more closely at the evolutionary trajectories of the hyperparasite. By using this approach, we are able to show that the probability of a hyperparasite hitchhiking is a key parameter in this system. The value of the hitchhiking parameter dramatically alters any predictions about evolutionary outcomes in hyperparasites. This model has the flexibility to be used in multiple systems and makes a strong case for continued study of these trispecies systems. In particular, it shows that the measurement of the hitchhiking probability is an essential part of growing the understanding of a hyperparasitic system.

Chapter two also presents a differential equation model but this time focusing on the ecology or epidemiology of the disease in question, namely one being controlled by a vaccine. This model makes clear that both the trajectory of immunity during the waning process, and the resulting level of susceptibility after vaccinal immunity has waned are core to any understanding of population level protection provided by the vaccine. With a simple model we can reveal these simple facets of vaccinal immunity to be driving the resulting protection experienced by the population. This new conceptual model is ripe for further advances and adaptations to understand the interaction between vaccination and a variety of factors.

Chapter three further builds the understanding of partial immunity by created a new technique for measuring the amount of immunity conferred by a natural infection, using a nonparametric negative control design. By designing a new method that is robust to individual differences in susceptibility, this new method can be used in a variety of settings and even in cases where individuals are coming from potentially two unknown subpopulations.

Chapter three also shows this method is unbiased providing for certain conditions. Accurate estimates of protection are vital for identifying vaccine candidate pathogens, and for making accurate estimates about the epidemiology of the disease and burden in a population.

In summary the approaches presented, two differential equation models and a statistical estimation method, grow our understanding of infectious disease dynamics. We can apply these frameworks to simplify complicated systems, distilling them down into lessons that can be learned and applied to future study of hyperparasitism, waning vaccinal immunity, partial naturally acquired immunity, or perhaps other related questions in the future.

Bibliography

[1] Peter A. Abrams. "The evolution of predator-prey interactions: theory and evidence". en. In: *Annual Review of Ecology and Systematics* 31 (2000), pp. 79–105.

[2] Miguel A Acevedo et al. "Virulence-driven trade-offs in disease transmission: A meta-analysis". In: *Evolution* 73.4 (2019), pp. 636–647.

[3] Jacqueline M. Achkar and Arturo Casadevall. *Antibody-mediated immunity against tuberculosis: Implications for vaccine development*. 2013.

[4] Erol Akcay. *Evolutionary models of mutualism*. en. 2015.

[5] Samuel Alizon and Minus van Baalen. "Multiple infections, immune dynamics and virulence evolution". en. In: *American Naturalist* 172:E150-E158 (2008).

[6] Samuel Alizon et al. "Virulence evolution and the tradeoff hypothesis: history, current state of affairs and the future". en. In: *Journal of Evolutionary Biology* 22 (2009), pp. 245–259.

[7] Benjamin M Althouse et al. "Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2". In: *arXiv preprint arXiv:2005.13689* (2020).

[8] Sandra B Andersen et al. "Disease dynamics in a specialized parasite of ant societies". In: *PloS one* 7.5 (2012), e36352.

[9] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1991.

[10] Roy M Anderson et al. "Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination". In: *The Lancet* 396.10263 (2020), pp. 1614–1616.

[11] Roy M. Anderson and Robert M. May. "Coevolution of hosts and parasites". en. In: *Parasitology* 85 (1982), pp. 411–426.

[12] Jason R. Andrews et al. "Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis". In: *Clinical Infectious Diseases* (2012).

[13] Sarah F Andrews et al. "Is it possible to develop a "universal" influenza virus vaccine? Immunogenetic considerations underlying B-cell biology in the development of a pan-subtype influenza A vaccine targeting the hemagglutinin stem". In: *Cold Spring Harbor Perspectives in Biology* 10.7 (2018), a029413.

[14] Sarah F Andrews et al. "Preferential induction of cross-group influenza A hemagglutinin stem–specific memory B cells after H7N9 immunization in humans". In: *Science immunology* 2.13 (2017), eaan2676.

[15] Adrian Garth Arbous and JE Kerrich. "Accident statistics and the concept of accident-proneness". In: *Biometrics* 7.4 (1951), pp. 340–432.

[16] N Arinaminpathy et al. "Population implications of the deployment of novel universal vaccines against epidemic and pandemic influenza". In: *Journal of the Royal Society Interface* 17.164 (2020), p. 20190879.

[17] Nimalan Arinaminpathy et al. "A global system for the next generation of vaccines". In: *Science* 376.6592 (2022), pp. 462–464.

[18] Ben Ashby et al. "Understanding the role of eco-evolutionary feedbacks in host-parasite coevolution". en. In: *Journal of Theoretical Biology* 464 (2019), pp. 115–125.

[19] Katherine E Atkins et al. "Modelling Marek's disease virus (MDV) infection: parameter estimates for mortality rate and infectiousness". In: *BMC veterinary research* 7 (2011), pp. 1–12.

[20] Lindsey R Baden et al. "Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine". In: *New England journal of medicine* 384.5 (2021), pp. 403–416.

[21] Norman TJ Bailey. *The mathematical theory of infectious diseases and its applications.* 2nd ediition. Charles Griffin, 1975.

[22] Sarah M Bartsch et al. "The potential epidemiologic, clinical, and economic value of a universal coronavirus vaccine: a modelling study". In: *EClinicalMedicine* 68 (2024).

[23] Chris T Bauch and David JD Earn. "Vaccination and the theory of games". In: *Proceedings of the National Academy of Sciences* 101.36 (2004), pp. 13391–13394.

[24] Alex Best, Andy White, and Mike Boots. "The Implications of Coevolutionary Dynamics to Host-Parasite Interactions". en. In: *The American Naturalist* 173 (2009), pp. 779–791.

[25] Mike Boots et al. "How Specificity and Epidemiology Drive the Coevolution of Static Trait Diversity in Hosts and Parasites". en. In: *Evolution* 68 (2014), pp. 1594–1606.

[26] Mike Boots et al. "The role of ecological feedbacks in the evolution of host defence: what does theory tell us?" en. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (2009), pp. 27–36.

[27] Michael A. Brockhurst and Britt Koskella. "Experimental coevolution of species interactions". en. In: *Trends in Ecology and Evolution* 28 (2013), pp. 367–375.

[28] ED Brodie III and BJ Ridenhour. "Reciprocal selection at the phenotypic interface of coevolution". In: *Integrative and Comparative Biology* 43.3 (2003), pp. 408–418.

[29] Sarah F. Bryner and Daniel Rigling. "Hypovirus Virulence and Vegetative Incompatibility in Populations of the Chestnut Blight Fungus". en. In: *Phytopathology* 102 (2012), pp. 1161–1167.

[30] Sarah F. Bryner, Daniel Rigling, and Patrick C. Brunner. "Invasion history and demographic pattern of Cryphonectria hypovirus 1 across European populations of the chestnut blight fungus". en. In: *Ecology and Evolution* 2 (2012), pp. 3227–3241.

[31] S Cankat, MU Demael, and L Swadling. "In search of a pan-coronavirus vaccine: next-generation vaccine design and immune mechanisms". In: *Cellular and molecular immunology* 21.2 (2024), pp. 103–118.

[32] Chen Yuan Chiang and Lee W. Riley. *Exogenous reinfection in tuberculosis*. 2005.

[33] G.H. Choi and D.L. Nuss. "Hypovirulence of Chestnut Blight Fungus Conferred by an Infectious Viral cDNA". en. In: *Science* 257 (1992), pp. 800–803.

[34] William G Cochran. "Some methods for strengthening the common χ 2 tests". In: *Biometrics* 10.4 (1954), pp. 417–451.

[35] C.E. Cressler et al. "The adaptive evolution of virulence: a review of theoretical predictions and empirical tests". en. In: *Parasitology* 143 (2016), pp. 915–930.

[36] Roberto Cruz-Flores et al. "Hyperparasitism by the bacteriophage (Caudovirales) infecting Candidatus Xenohaliotis californiensis (Rickettsiales-like prokaryote) parasite of wild abalone Haliotis fulgens and Haliotis corrugata from the Peninsula of Baja California, Mexico". In: *Journal of invertebrate pathology* 140 (2016), pp. 58–67.

[37] Ron Dagan et al. "Prevention of early episodes of otitis media by pneumococcal vaccines might reduce progression to complex disease". In: *The Lancet infectious diseases* 16.4 (2016), pp. 480–492.

[38] Carolyn J. Dakin et al. "Inflammation, infection, and pulmonary function in infants and young children with cystic fibrosis". In: *American Journal of Respiratory and Critical Care Medicine* (2002).

[39] QingChao Deng et al. "The horizontal transmission of Cryphonectria hypovirus 1 (CHV1) is affected by virus strains". In: *Chinese science bulletin* 54.17 (2009), pp. 3053–3060.

[40] Ping Ding et al. "Transmission of Cryphonectria hypovirus to protect chestnut trees from chestnut blight disease". In: *Biological control* 40.1 (2007), pp. 9–14.

[41] Ensheng Dong, Hongru Du, and Lauren Gardner. "An interactive web-based dashboard to track COVID-19 in real time". In: *The Lancet infectious diseases* 20.5 (2020), pp. 533–534.

[42] Eve Dubé et al. "Vaccine hesitancy: an overview". In: *Human vaccines and immunotherapeutics* 9.8 (2013), pp. 1763–1773.

[43] Dieter Ebert. "Host-parasite coevolution: Insights from the Daphnia parasite model system". en. In: *Current Opinion in Microbiology* 11 (2008), pp. 290–301.

[44] Mohamed El Khalifi and Tom Britton. "Extending susceptible-infectious-recovered-susceptible epidemics to allow for gradual waning of immunity". In: *Journal of the Royal Society Interface* 20.206 (2023), p. 20230042.

[45] Mohamed El Khalifi and Tom Britton. "SIRS epidemics with individual heterogeneity of immunity waning". In: *Journal of Theoretical Biology* 587 (2024), p. 111815.

[46] Jill M Ferdinands et al. "Intraseason waning of influenza vaccine protection: evidence from the US influenza vaccine effectiveness network, 2011–2012 through 2014–2015". In: *Clinical Infectious Diseases* 64.5 (2017), pp. 544–550.

[47] Julia Fitzner et al. "Revision of clinical case definitions: Influenza-like illness and severe acute respiratory infection". In: *Bulletin of the World Health Organization* (2018).

[48] Kimberley K. Fox et al. "Longitudinal evaluation of serovar-specific immunity to Neisseria gonorrhoeae". In: *American Journal of Epidemiology* (1999).

[49] Maria F. Gallo et al. "Self-reported condom use is associated with reduced risk of chlamydia, gonorrhea, and trichomoniasis". In: *Sexually Transmitted Diseases* (2007).

[50] Sylvain Gandon et al. "Host-parasite coevolution and patterns of adaptation across time and space". en. In: *Journal of Evolutionary Biology* 21 (2008), pp. 1861–1866.

[51] Stefan AH Geritz et al. "Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree". In: *Evolutionary ecology* 12 (1998), pp. 35–57.

[52] Beryl P. Gladstone et al. "Protective Effect of Natural Rotavirus Infection in an Indian Birth Cohort". In: *New England Journal of Medicine* 365.4 (July 2011), pp. 337–346.

[53] JudithÂ R. Glynn et al. "High Rates of Recurrence in HIVâInfected and HIVâUninfected Patients with Tuberculosis". In: *The Journal of Infectious Diseases* (2010).

[54] Robert J Glynn and Julie E Buring. "Ways of measuring rates of recurrent events". In: *Bmj* 312.7027 (1996), pp. 364–367.

[55] Pedro Gomez and Angus Buckling. "Bacteria-Phage Antagonistic Coevolution in Soil". en. In: *Science* 332 (2011), pp. 106–109.

[56] Lynn Govaert et al. "Eco-evolutionary feedbacks—Theoretical models and perspectives". In: *Functional Ecology* 33.1 (2019), pp. 13–30.

[57] Major Greenwood and G Udny Yule. "An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents". In: *Journal of the Royal statistical society* 83.2 (1920), pp. 255–279.

[58] Fiona M Guerra et al. "Waning of measles maternal antibody in infants in measles elimination settings–a systematic literature review". In: *Vaccine* 36.10 (2018), pp. 1248–1255.

[59] Alex R. Hall et al. "Measuring Coevolutionary Dynamics in Species-Rich Communities". en. In: *Trends in Ecology and Evolution* 35 (2020), pp. 539–550.

[60] Hans Heesterbeek et al. *Modeling infectious disease dynamics in the complex landscape of global health*. Mar. 2015.

[61] Herbert W Hethcote, Harlan W Stech, and Pauline Van Den Driessche. "Nonlinear oscillations in epidemic models". In: *SIAM Journal on Applied Mathematics* 40.1 (1981), pp. 1–9.

[62] Robert D. Holt and Michael E. Hochberg. "The Coexistence of Competing Parasites. Part II-Hyperparasitism and Food Chain Dynamics". en. In: *Journal of Theoretical Biology* 193 (1998), pp. 485–495.

[63] V M Howie, J H Ploussard, and J Sloyer. "The "otitis-prone" condition." In: *American journal of diseases of children (1960)* 129.6 (1975), pp. 676–8.

[64] Leah C Katzelnick et al. "Antibody-dependent enhancement of severe dengue disease in humans". In: *Science* 358.6365 (2017), pp. 929–932.

[65] Gerald T. Keusch et al. "Environmental enteric dysfunction: Pathogenesis, diagnosis, and clinical consequences". In: *Clinical Infectious Diseases* 59 (2014), S207–S212.

[66] Surender Khurana et al. "Licensed H5N1 vaccines generate cross-neutralizing antibodies against highly pathogenic H5N1 clade 2.3. 4.4 b influenza virus". In: *Nature Medicine* (2024), pp. 1–1.

[67] Aaron A King et al. "Inapparent infections and cholera dynamics". In: *Nature* 454.7206 (2008), pp. 877–880.

[68] Levente Kiss et al. "Biology and biocontrol potential of Ampelomyces mycoparasites, natural antagonists of powdery mildew fungi". In: *Biocontrol Science and Technology* 14.7 (2004), pp. 635–651.

[69] Stephen M Kissler et al. "Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period". In: *Science* 368.6493 (2020), pp. 860–868.

[70] Petra Klepac, Ramanan Laxminarayan, and Bryan T Grenfell. "Synthesizing epidemiological and economic optima for control of immunizing infections". In: *Proceedings of the National Academy of Sciences* 108.34 (2011), pp. 14366–14370.

[71] Katia Koelle et al. "The changing epidemiology of SARS-CoV-2". In: *Science* 375.6585 (2022), pp. 1116–1121.

[72] Kaitlyn E. Kortright et al. "Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria". en. In: *Cell Host and Microbe* 25 (2019), pp. 219–232.

[73] Anna-Liisa Laine. "Evolution of host resistance: looking for coevolutionary hotspots at small spatial scales". en. In: *Proceedings of the Royal Society B: Biological Sciences* 273 (2006), pp. 267–273.

[74] Jennie S Lavine, Ottar N Bjornstad, and Rustom Antia. "Immunological characteristics govern the transition of COVID-19 to endemicity". In: *Science* 371.6530 (2021), pp. 741–745.

[75] Ramanan Laxminarayan et al. "Epidemiology and transmission dynamics of COVID-19 in two Indian states". In: *Science* 370.6517 (2020), pp. 691–697.

[76] Joseph A Lewnard et al. "Heterogeneous susceptibility to rotavirus infection and gastroenteritis in two birth cohort studies: Parameter estimation and epidemiological implications". In: *PLoS computational biology* 15.7 (2019), e1007014.

[77] Joseph A Lewnard et al. "Measurement of Vaccine Direct Effects under the Test-Negative Design". In: *American Journal of Epidemiology* (Aug. 2018).

[78] Joseph A Lewnard et al. "Naturally acquired immunity against rotavirus infection and gastroenteritis in children: paired reanalyses of birth cohort studies". In: *The Journal of infectious diseases* 216.3 (2017), pp. 317–326.

[79] Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. "Negative Controls: A tool for detecting confounding and bias in observational studies". In: *Epidemiology* (2010).

[80] Ben Lopman and Gagandeep Kang. *In praise of birth cohorts: Norovirus infection, disease, and immunity*. 2014.

[81] Adriana Lucia-Sanz et al. "Modeling shield immunity to reduce COVID-19 transmission in long-term care facilities". In: *Annals of Epidemiology* 77 (2023), pp. 44–52.

[82] Sachin Mani, Thomas Wierzba, and Richard I. Walker. "Status of vaccine research and development for Shigella". In: *Vaccine* (2016).

[83] Nathan Mantel. "Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure". In: *Journal of the American Statistical Association* 58.303 (1963), pp. 690–700.

[84] Nathan Mantel and William Haenszel. "Statistical aspects of the analysis of data from retrospective studies of disease". In: *Journal of the national cancer institute* 22.4 (1959), pp. 719–748.

[85] Tianyang Mao et al. "Unadjuvanted intranasal spike vaccine elicits protective mucosal immunity against sarbecoviruses". In: *Science* 378.6622 (2022), eabo2523.

[86] Florian M Marx et al. "The temporal dynamics of relapse and reinfection tuberculosis after successful treatment: a retrospective cohort study". In: *Clinical infectious diseases* 58.12 (2014), pp. 1676–1683.

[87] Robert M May and Roy M Anderson. "Spatial heterogeneity and the design of immunization programs". In: *Mathematical Biosciences* 72.1 (1984), pp. 83–111.

[88] Robert M. May and Martin A. Nowak. "Coinfection and the evolution of parasite virulence". en. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences*. Vol. 261. 1995, pp. 209–215.

[89] C Jessica E Metcalf et al. "Opportunities and challenges of a World Serum Bank–Authors' reply". In: *The Lancet* 389.10066 (2017), p. 252.

[90] C Jessica E Metcalf et al. "Use of serological surveys to generate key insights into the changing global landscape of infectious disease". In: *The Lancet* 388.10045 (2016), pp. 728–730.

[91] Michael G Milgroom and Paolo Cortesi. "Biological control of chestnut blight with hypovirulence: a critical analysis". In: *Annu. Rev. Phytopathol.* 42.1 (2004), pp. 311–338.

[92] Michael J Mina et al. "A Global lmmunological Observatory to meet a time of pandemics". In: *Elife* 9 (2020), e58989.

[93] A. S. Monto and K. M. Sullivan. "Acute respiratory illness in the community. Frequency of illness and the agents involved". In: *Epidemiology and Infection* (1993).

[94] Andrew Y. Morozov, Cecile Robin, and Alain Franc. "A simple model for the dynamics of a host-parasite-hyperparasite interaction". en. In: *Journal of Theoretical Biology* 249 (2007), pp. 246–253.

[95] Sinead E Morris et al. "Demographic buffering: titrating the effects of birth rate and imperfect immunity on epidemic dynamics". In: *Journal of the Royal Society Interface* 12.104 (2015), p. 20141245.

[96] Jo??l Mossong et al. "Social contacts and mixing patterns relevant to the spread of infectious diseases". In: *PLoS Medicine* 5.3 (2008), pp. 0381–0391.

[97] Elizabeth Mostofsky, Brent A. Coull, and Murray A. Mittleman. "Analysis of Observational Self-matched Data to Examine Acute Triggers of Outcome Events with Abrupt Onset". In: *Epidemiology* 29.6 (2018), pp. 804–816.

[98] Graham R Northrup et al. "Inference of naturally acquired immunity using a self-matched negative-control design". In: *Epidemiology* 32.2 (2021), pp. 168–178.

[99] Graham R Northrup et al. "The evolutionary dynamics of hyperparasites". In: *Journal of Theoretical Biology* 582 (2024), p. 111741.

[100] Erik E. Osnas and Andrew P. Dobson. "Evolution of Virulence in Heterogeneous Host Communities Under Multiple Tradeoffs". no. In: *Evolution* 66 (2012), pp. 391–401.

[101] V. Shane Pankratz, Mariza De Andrade, and Terry M. Therneau. "Random-effects cox proportional hazards model: General variance components methods for time-to-event data". In: *Genetic Epidemiology* (2005).

[102] Young-Jun Park et al. "Antibody-mediated broad sarbecovirus neutralization through ACE2 molecular mimicry". In: *Science* 375.6579 (2022), pp. 449–454.

[103] Steven R. Parratt and Anna-Liisa Laine. "Pathogen dynamics under both bottom-up host resistance and top-down hyperparasite attack". en. In: *Journal of Applied Ecology* 55 (2018), pp. 2976–2985.

[104] Steven R. Parratt and Anna-Liisa Laine. "The role of hyperparasitism in microbial pathogen ecology and evolution". en. In: *ISME J* 10 (2016), pp. 1815–1822.

[105] Marcela F. Pasetti and Myron M. Levine. *Insights from natural infection-derived immunity to cholera instruct vaccine efforts*. Jan. 2012.

[106] Tobin L Peever et al. "Variation in tolerance and virulence in the chestnut blight fungus-hypovirus interaction". In: *Applied and environmental microbiology* 66.11 (2000), pp. 4863–4869.

[107] Helen Petousis-Harris et al. "Effectiveness of a group B outer membrane vesicle meningococcal vaccine against gonorrhoea in New Zealand: a retrospective case-control study". In: *The Lancet* 390.10102 (2017), pp. 1603–1610.

[108] Natalie Pica and Peter Palese. "Toward a Universal Influenza Virus Vaccine: Prospects and Challenges". In: *Annual Review of Medicine* 64.1 (2013), pp. 189–202.

[109] Naveen Pillarisetti et al. "Infection, inflammation,and lung function decline in infants with cystic fibrosis". In: *American Journal of Respiratory and Critical Care Medicine* (2011).

[110] F. A. Plummer et al. "Epidemiologic evidence for the development of serovar-specific immunity after gonococcal infection". In: *Journal of Clinical Investigation* (1989).

[111] Fernando P Polack et al. "Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine". In: *New England journal of medicine* 383.27 (2020), pp. 2603–2615.

[112] Daniel Rigling and Simone Prospero. "Cryphonectria parasitica, the causal agent of chestnut blight: invasion history, population biology and disease control". en. In: *Molecular Plant Pathology* 19 (2017), pp. 7–20.

[113] Isabel Rodriguez-Barraquer et al. "Quantifying Heterogeneous Malaria Exposure and Clinical Protection in a Cohort of Ugandan Children". In: *Journal of Infectious Diseases*. 2016.

[114] Elizabeth T Rogawski McQuade et al. "Protection From Natural Immunity Against Enteric Infections and Etiology-Specific Diarrhea in a Longitudinal Birth Cohort". In: *The Journal of Infectious Diseases* (2020).

[115] Saba Rouhani et al. "Norovirus infection and acquired immunity in 8 countries: Results from the MAL-ED study". In: *Clinical Infectious Diseases* 62.10 (2016), pp. 1210–1217.

[116] Chadi M Saad-Roy, Adrian B McDermott, and Bryan T Grenfell. "Dynamic perspectives on the search for a universal influenza vaccine". In: *The Journal of Infectious Diseases* 219.Supplement_1 (2019), S46–S56.

[117] Chadi M Saad-Roy, C Jessica E Metcalf, and Bryan T Grenfell. "Immuno-epidemiology and the predictability of viral evolution". In: *Science* 376.6598 (2022), pp. 1161–1162.

[118] Chadi M Saad-Roy and Arne Traulsen. "Dynamics in a behavioral–epidemiological model for individual adherence to a nonpharmaceutical intervention". In: *Proceedings of the National Academy of Sciences* 120.44 (2023), e2311584120.

[119] Chadi M Saad-Roy et al. "Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes". In: *Science* 372.6540 (2021), pp. 363–370.

[120] Chadi M Saad-Roy et al. "Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years". In: *Science* 370.6518 (2020), pp. 811–818.

[121] Chadi M Saad-Roy et al. "Medium-term scenarios of COVID-19 as a function of immune uncertainties and chronic disease". In: *Journal of the Royal Society Interface* 20.205 (2023), p. 20230247.

[122] Chadi M Saad-Roy et al. "Trajectory of individual immunity and vaccination required for SARS-CoV-2 community immunity: a conceptual investigation". In: *Journal of The Royal Society Interface* 18.175 (2021), p. 20200683.

[123] Simran K Sandhu et al. "Revisiting the role of hyperparasitism in the evolution of virulence". In: *The American Naturalist* 197.2 (2021), pp. 216–235.

[124] Akira Sasaki. "Host-parasite coevolution in a multilocus gene-for-gene system". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1458 (2000), pp. 2183–2188.

[125] Sharon Y. Strauss. "Ecological and evolutionary responses in complex communities: implications for invasions and eco-evolutionary feedbacks". en. In: *Oikos* 123 (2014), pp. 257–266.

[126] Sheena G. Sullivan, Eric J Tchetgen Tchetgen, and Benjamin J. Cowling. "Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness". In: *American Journal of Epidemiology* 184.5 (2016), pp. 345–353.

[127] Dereck R Tait et al. "Final Analysis of a Trial of M72/AS01E Vaccine to Prevent Tuberculosis." In: *The New England journal of medicine* (2019), pp. 1–12.

[128] Douglas R Taylor et al. "The acquisition of hypovirulence in host-pathogen systems with three trophic levels". In: *The American Naturalist* 151.4 (1998), pp. 343–355.

[129] Terry M Therneau. "Package 'survival'". In: (2019).

[130] Stephen J Thomas et al. "Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine through 6 months". In: *New England Journal of Medicine* 385.19 (2021), pp. 1761–1773.

[131] John N. Thompson. "The Evolution of Species Interactions". en. In: *Science* 284 (1999), pp. 2116–2118.

[132] Charlotte Tollenaere et al. "A hyperparasite affects the population dynamics of a wild plant pathogen". In: *Molecular Ecology* 23.23 (2014), pp. 5877–5887.

[133] Arne Traulsen, Simon A Levin, and Chadi M Saad-Roy. "Individual costs and societal benefits of interventions during the COVID-19 pandemic". In: *Proceedings of the National Academy of Sciences* 120.24 (2023), e2303546120.

[134] Pauline Van den Driessche and James Watmough. "Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission". In: *Mathematical biosciences* 180.1-2 (2002), pp. 29–48.

[135] John. Vandermeer, Ivette. Perfecto, and Heidi Liere. "Evidence for hyperparasitism of coffee rust (Hemileia vastatrix) by the entomogenous fungus, Lecanicillium lecanii, through a complex ecological web". zu. In: *Plant Pathology* 58 (2009), pp. 636–641.

[136] James W Vaupel, Kenneth G Manton, and Eric Stallard. "The impact of heterogeneity in individual frailty on the dynamics of mortality". In: *Demography* 16.3 (1979), pp. 439–454.

[137] F. Raúl Velázquez et al. "Rotavirus Infection in Infants as Protection against Subsequent Infections". In: *New England Journal of Medicine* 335.14 (Oct. 1996), pp. 1022–1028.

[138] Suzanne Verver et al. "Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis". In: *American Journal of Respiratory and Critical Care Medicine* (2005).

[139] Merryn Voysey et al. "Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK". In: *The Lancet* 397.10269 (2021), pp. 99–111.

[140] Caroline E Wagner et al. "Economic and behavioral influencers of vaccination and antimicrobial use". In: *Frontiers in Public Health* 8 (2020), p. 614113.

[141] Caroline E Wagner et al. "Vaccine nationalism and the dynamics and control of SARS-CoV-2". In: *Science* 373.6562 (2021), eabj7364.

[142] Ning Wang et al. "Identification of a Hyperparasitic Simplicillium obclavatum Strain Affecting the Infection Dynamics of Puccinia striiformis f. sp. tritici on Wheat". In: *Frontiers in Microbiology* 11 (2020), p. 1277.

[143] Joshua S Weitz et al. "Modeling shield immunity to reduce COVID-19 epidemic spread". In: *Nature medicine* 26.6 (2020), pp. 849–854.

[144] SR Weldon, MR Strand, and KM Oliver. "Phage loss and the breakdown of a defensive symbiosis in aphids". In: *Proceedings of the Royal Society B: Biological Sciences* 280.1751 (2013), p. 20122103.

[145] Heather J. Whitaker et al. "Tutorial in biostatistics: The self-controlled case series method". In: *Statistics in Medicine* (2006).

[146] Beyene Zewdie et al. "Temporal dynamics and biocontrol potential of a hyperparasite on coffee leaf rust across a landscape in Arabica coffee's native range". In: *Agriculture, Ecosystems and Environment* 311 (2021), p. 107297.