THE UNIVERSITY OF CHICAGO

HOST HETEROGENEITY AND THE DYNAMICS OF PATHOGEN DIVERSITY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM: ECOLOGY AND EVOLUTION

BY
SYLVIA RANJEVA

CHICAGO, ILLINOIS
JUNE 2018

I would like to dedicate this thesis to my mother, Dr. Anna Mattila, who taught me the empowerment of education. It is an honor to follow in her example.

"Man's first glance at the universe discovers only variety, diversity, multiplicity of phenomena. Let that glance be illuminated by science..."

— Louis Pasteur

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am incredibly grateful for the support of outstanding scientific colleagues, collaborators, and mentors. Each of these individuals inspired me with their generosity, creativity, and kindness, and the largest privilege of my Ph.D. career was to learn from their example. Particularly, I would like to thank my thesis committee, Mercedes Pascual and Stefano Allesina, and my advisors, Greg Dwyer and Sarah Cobey. Mercedes and Stefano provided invaluable insight and feedback about the chapters of this dissertation, and they encouraged a spirit of lively debate at committee meetings that fueled my passion for the work.

I had the honor of mentorship from two leaders in the field, Greg and Sarah. Greg is a pioneer in the discipline of fitting nonlinear dynamical models to data, and from his example I have learned to embrace the "messy" complexity of real-world systems. Sarah is one of the formative minds of modern infectious disease epidemiology. Her work weaves ecological theory, immunology, and state-of-the-art mathematical techniques seamlessly into an emerging picture of host-pathogen interactions across scales of life. Sarah's creativity, rigor, and wide-reaching expertise in multiple scientific areas have set the model for my career in research.

Greg and Sarah worked tirelessly to encourage me and to shape my capacity to think critically about science. In 2012, I came to Greg's office with a background in engineering and with a half-baked scheme to make a career that combined two of my passions, math and the study of human disease. Science can be an intimidating world, and my lack of background in both statistics and ecology primed me with a sense of inadequacy. Greg took the time to listen to me, to think about my interests, and to introduce fundamental concepts in disease ecology in a way that was relatable and motivating. I attribute much of the passion that I have developed for disease ecology to that early encounter. As Greg's student, I realized that our first meeting was representative of the way that he approaches mentorship. He makes students feel capable, important, and inspired, all while

challenging them to continually improve. Sarah radiates enthusiasm, scientific curiosity, professionalism, and humility. When I joined Sarah's lab, math was intellectually "safe" for me, and she challenged me to also think critically about biology. Sarah taught me to expect more from myself than I thought I could. She was relentlessly patient, kind, and constructively critical. She held my work to a standard that reflected her strong belief in me, a belief that I grew to share as I developed under her mentorship. Sarah puts 150% of her own effort into anything that she asks of a student, and she has truly shaped my abilities and confidence approaching scientific questions.

I would like to thank members of the Dwyer and Cobey labs for the pleasure of their collaboration and camaraderie. Molly Gallagher, Colin Kyle, and Joe Mihaljevic were encouraging throughout my time in Greg's lab. Frank Wen, Marcos Vieira, Rahul Subramanian, Phil Arevalo, and Kangchon Kim made the Cobey lab a home for me, and a place that I am proud to represent. My Cobey lab colleagues balance immense talent and intellect with kindness, generosity, and humility. I would especially like to thank Ed Baskerville, who taught me how to program and who devoted substantial effort to my progress during my first year in the lab.

I am grateful for collaborators that shared their data and epidemiological expertise. Anna Giuliano, Ed Baskerville, and Ken Alexander were instrumental to the HPV project. Ben Cowling, Vicky Fang, and Rahul Subramanian made the influenza project possible. I am also grateful for substantial administrative support from both the MSTP program and from the Department of Ecology and Evolution.

I am immeasurably grateful for the support of friends and family. I would like to thank my MSTP classmates, who have been a family to me since we began this journey together. It has been a joy to watch our class grow together as scientists and as people. I would also like to thank my close friends outside of academia, who have been some of my biggest cheerleaders and greatest sources of comfort. I would like to particularly thank Ian, whose love, selflessness, and kindness never cease to amaze me. Ian inspires

me as a scientific colleague, as a friend, and as a partner, and I am excited to see where our life together takes us. Finally, I am grateful for my family members, who have made considerable sacrifices throughout my life to help me achieve my goals. Their love and encouragement motivates me every day.

ABSTRACT

A central goal of ecology is to explain the diversity of coexisting species. Analogously, fundamental questions in epidemiology center around coexistence in pathogen communities. Classical models of strain competition predict variable coexistence dynamics depending upon the strength of cross-immunity. However, heterogeneity among human hosts, through variation in population structure and immune responses, can also profoundly affect coexistence. This dissertation investigates the intersection of immune-mediated competition and host heterogeneity to explain the dynamics of pathogen diversity in two viral communities: human papillomavirus (HPV) and influenza A viruses. We test hypotheses about viral dynamics by fitting mechanistic models to longitudinal data. We show that the prevalence and coexistence of over 200 genetically distinct HPV types are maintained by recurring infection within individuals of type-specific high-risk subpopulations. We then show that the dynamics of immune protection after influenza infection differ between children and adults, signaling substantial variability in the population-level selective pressures that shape the diversity of influenza strains.

CHAPTER 1

INTRODUCTION

A fundamental goal of ecology is to explain the observed diversity of species in ecological communities. Analogously, pathogen diversity frames central questions in infectious disease epidemiology: What are the mechanisms that support pathogen diversity? How do interactions with the host immune system drive competition, selection, and variation in pathogen traits? What do observed patterns of diversity imply about the dynamics of host immunity across a population?

The population dynamics of coexisting species depend upon competition for shared resources. One of the earliest ecological theories about the "struggle for existence" is that competition between two species for limited resources in a perfectly overlapping environment will lead to competitive exclusion by the species with higher fitness [55]. Early coexistence theory suggests that species competing for shared, limited resources may coexist by occupying unique ecological niches [78]. Contemporary theory suggests that long-term coexistence is a balance between equalizing mechanisms that reduce differences in relative fitness [162, 27] and stabilizing mechanisms that reduce niche overlap [73, 27]. Niche differences stabilize competitive dynamics by promoting intra-species competition over inter-species competition, such that rarer species have higher per-capita population growth rates, a process known as negative frequency-dependent selection (NFDS) [27].

Similarly, the abundance of many pathogens depends on the dynamics of competition for susceptible hosts [102, 160, 28]. Pathogen community structure may arise by neutral processes (such as stochastic infection events or neutral mutations) or by niche differentiation that constrains the pathogens present in each host. We can consider individual hosts as partially isolated patches or habitats that comprise a population-level immune environment. Pathogen niches can therefore be described as the spectrum of processes that alter host-pathogen interactions. Types within a pathogen species may differ by neutral markers, by drug resistance determinants, by antigens that trigger immune responses,

by infectiousness or persistence, and often by a combination of traits. One classic view of pathogen diversity holds that strong diversifying selection from the host immune system drives niche partitioning by antigenic variation between closely-related pathogens. Many human pathogens can be categorized into distinct "strains", each defined by a set of polymorphic, non-overlapping antigens [68, 70, 7, 67]. Strains compete by inducing adaptive immune responses that are specific to shared antigens, limiting the growth rates of antigenically similar competitors [28]. The accumulation of specific immunity in the host population promotes the spread of rarer antigenic variants by NFDS. Such immune-mediated competition has been invoked to explain the antigenic and genetic diversity of influenza, pneumococcus, rotavirus, norovirus, *Neisseria meningitidis*, malaria, hepatitis C, HIV, trypanosomes, and other common pathogens [68, 107, 28, 102, 101, 160, 142].

Theoretical predictions about strain competition draw from ecological principles to quantify the effects of competition and host environmental variability on coexistence. Classical models of strain competition, which assume both NFDS and niche partitioning by antigenic variation, show that complex coexistence dynamics can arise depending on the degree of cross-immunity [69]. Under intense competition, long-term diversity is greatly restricted. Because immunity to a specific antigen phenotype dramatically reduces the future transmission success of strains possessing that variant, a single subset of strains with non-overlapping phenotypes dominates in prevalence over time. At very low levels of cross-immunity, strains can coexist at stable prevalence without discrete structure. Intermediate levels of competition can yield complex patterns of diversity, where the relative proportions of different variant groups exhibit cyclical or chaotic fluctuations over time.

While strain competition can lead to diversity under certain scenarios of cross-immunity, pathogen diversity can also arise if the fitness of related types varies across individual hosts [144, 89, 131, 47]. Two forms of host heterogeneity have particularly important implications for pathogen dynamics: variation in population structure and variation

in immune responses. Multiple forms of population structure affect disease dynamics and pathogen diversity. For example, community structure leads to spatial separation of pathogen strains. Well-defined neighboring populations with low connectivity can stably maintain diverse subsets of pathogen strains [20]. Age-assortative mixing leads to different transmission rates for the same circulating pathogen among individuals of different age groups [3]. Additionally, individuals' positions within social contact networks can separately affect their risk. Local clustering of hosts in highly connected high-risk subpopulations can impact epidemic dynamics and change the selection pressures acting on traits such as virulence and duration of the infectious period [143, 44, 90]. Heterogeneity in the host immune response to circulating pathogens also influences pathogen dynamics. Individuals may vary extensively in their immune response to circulating pathogens Extensions of strain theory demonstrate that diversity is maximized at narrow, strain-specific host responses [30]. Increasing either the number of immunodominant epitopes perceived by the immune system or the heterogeneity in the relative immunodominance of epitopes among hosts lowers the threshold of cross-immunity between strains at which chaotic strain dynamics, and hence competitive exclusion, can occur.

This dissertation investigates the intersection of host heterogeneity and immune-mediated competition in the dynamics of pathogen diversity. I examine the mechanisms that underlie coexistence in two viral communities with contrasting patterns of standing genetic diversity: human papillomaviruses (HPV), a group of over 200 double-stranded DNA virus types that exist at low, stable prevalence in humans but have questionable antigenic distinctions [16, 106], and influenza A viruses, a group of single-stranded RNA viruses that exhibit rapid replacement of antigenic variants within two major subtypes. Below, I provide a broad overview of the general arms of the human immune system that mediate selective pressure in viral communities. I then provide an overview of the major hypotheses and results for each chapter.

1.1 Overview of human immunity

The human immune system consists of innate and adaptive branches that act in coordination to control infection. Because interactions with host immunity define pathogen fitness, different forms of immunity generate different selective pressures on pathogen populations based on their function. The innate immune system involves pre-existing host defenses such as barriers formed by skin and mucosa, complement, and phagocytes. Innate immunity confers generally short-lived, non-specific protection. Adaptive immunity is especially important for the control of viral infections. The adaptive arm of the immune system consists of humoral, or B cell-mediated immunity, and cellular, or T cell-mediated immunity. B cells and T cells interact with pathogens at antigenic sites called epitopes. Epitopes that elicit the strongest immune responses after antigen priming are called "dominant", and epitopes that elicit weaker responses are "subdominant" [1]. B cells produce antibodies that bind to epitopes on the pathogen surface. Antibodies can both neutralize extracellular pathogens and recruit other immune defenses. T cells recognize antigens as peptides bound to self surface proteins of the major histocompatibility complex (MHC), which are encoded by human leukocyte antigen (HLA) genes. Therefore, T cell activation requires binding of antigenic peptides to MHC molecules. CD4+ T cells, or "helper T cells," act on other cells of the immune system to promote various aspects of the immune response, including B cell maturation and antibody production, macrophage activation, and enhanced activity of natural killer cells and CD8+ T cells [82]. The main function of virus-specific CD8+ T cells is that of cytotoxic T lymphocytes [82]. Upon infection, these cells are activated in the lymphoid tissues and recruited to the infection site to eliminate infected cells and prevent viral reproduction.

1.2 Overview of dissertation

1.2.1 Chapter 2. The dynamics of human papillomavirus diversity

The high prevalence of human papillomavirus (HPV), the most common sexually transmitted infection, arises from the coexistence of over 200 genetically distinct types. However, little is known about fundamental aspects of HPV ecology, such as mechanisms of type-specific host immunity and the strength of inter-type interactions in natural infection. Conceivably, type-specific or "homologous" immunity could drive type diversity by NFDS. Additionally, heterogeneity in host risk among sub-populations defined by demographic and behavioral traits could also support the transmission of multiple types. To test for homologous immunity and to identify mechanisms determining HPV transmission, we fitted nonlinear mechanistic models to longitudinal data on genital infections in unvaccinated men. Our results provide no evidence for homologous immunity, instead showing that infection with one HPV type strongly increases the risk of infection with that type for years afterwards. This increased risk occurs in both sexually active and celibate men, suggesting that it arises from auto-inoculation, episodic reactivation of latent virus, or both. Overall our results show that high HPV prevalence and diversity can be explained by a combination of a lack of homologous immunity, frequent reinfections, weak competition between types, and variation in type fitness between host subpopulations [142][1].

1.2.2 Chapter 3. Age-related variation in host immunity may restrict the standing diversity of circulating influenza A viruses.

Two subtypes of influenza A viruses, "pandemic" H1N1 and H3N2, currently circulate in humans. Despite high mutation rates and strong selection by host immunity on sur-

1. The results from Chapter 2 were published previously [142], and much of the chapter text appears as written in the published manuscript.

face viral proteins, strains within each subtype exhibit low standing diversity. Because of influenza's rapid evolution, individuals may be infected with the same subtype multiple times. Epidemiological and immunological evidence suggests that repeated exposures over a lifetime alter the specificity of influenza antibody responses, such that individuals may respond differently to the same circulating strain. Individual heterogeneity in the response to circulating strains that arises as a result of age-related changes in immune specificity could influence strain diversity. In this chapter, we quantified age-related differences in the dynamics of influenza protection by fitting mechanistic models to longitudinal serology in children and adults. Our results suggest that early immune responses target highly mutable sites on the head of the hemagglutinin surface protein, but that responses are focused over time towards other epitopes, which are generally more conserved. Our results imply that the immune responses in children and adults generate different selective pressures for influenza viruses.

CHAPTER 2

THE DYNAMICS OF HUMAN PAPILLOMAVIRUS DIVERSITY

2.1 Introduction

Human papillomavirus (HPV), a major cause of genital warts and anogenital and oropharyngeal cancers [59], is the most common sexually transmitted infection [166]. While the population prevalence of genital HPV is approximately 40% among women and 45% among men in the USA, over 200 HPV types have been identified, and the prevalence of individual types never exceeds 10% [71]. Understanding the factors that underlie the high diversity of low-prevalence HPV types requires defining fundamental aspects of HPV ecology, including mechanisms of host immunity, interactions between HPV types, and the characteristics of high-risk subpopulations, which are poorly understood.

It is unclear how HPV interacts with the immune system during infection, but in principle, distinctions between HPV types may arise due to acquired immunity from B cells and T cells. HPV types are defined by a 10% threshold of dissimilarity in the L1 gene, which codes for the major capsid protein [37]. The outer capsid modulates viral entry into host cells at the epithelial basement membrane [19], and the humoral response to infection is mainly type-specific anti-L1 antibodies [135]. Studies of HPV in T-cell-deficient people show that cellular immunity is important to control infection [152], but it is unclear how cellular immunity achieves type-specific recognition, if at all. In individuals with HPV16-related cancerous lesions, cytotoxic T lymphocytes specific to the E6 and E7 oncoproteins are correlated with reduced disease [121, 138]. The specificity of the T cell repertoire to other genes or to the majority of HPV types, however, is not well established.

Efforts to understand the effects of immunity on HPV dynamics must begin with homologous immunity, or protection against repeat infection with the same HPV type. Homologous immunity would limit the prevalence of any type through negative frequency dependence. The traditional assumption is that most HPV-infected individuals perma-

nently clear infection after 1-2 years [156, 147, 117], suggesting protective homologous immunity. In this scenario, the elevated cancer risk associated with particular HPV types results from a small fraction of persistent infections [117]. Longitudinal studies, however, have shown that hosts can be infected repeatedly [153]. Although type-specific antibodies may provide modest protection against future infection in women [8], serum antibody is not a marker of immune protection in men [132, 104]. Thus, the strength of homologous immunity remains unclear.

Evidence for competition between HPV types is also conflicting. Immune-mediated competition has been invoked to explain small increases in the prevalence of non-vaccine HPV types following HPV vaccination [85], and to explain weak cross-protection from the vaccine against related non-vaccine types [167]. Mathematical models show that type competition is consistent with observed patterns of HPV prevalence [43, 119]. Nevertheless, there is little empirical evidence for inter-type competition [146], as shown in part by elevated rates of multiple-type compared to single-type infections and frequent concurrent acquisition of HPV types [25].

Meanwhile, the risk of HPV infection depends on differences in demographic and behavioral risk factors between host subpopulations. For example, a host's number of sexual partners strongly affects infection risk [56, 59, 117], and some evidence suggests that oncogenic and non-oncogenic types are differently sensitive to the numbers of partners among hosts [159]. However, detailed comparisons of risk factors for infection with each HPV type have not been performed. Different risk factors would suggest differences between types among host subpopulations that could explain type prevalence.

To investigate the factors determining HPV prevalence and diversity, we fitted mechanistic models of HPV dynamics to an extensive longitudinal dataset. Mechanistic models have long been used in infectious disease ecology to quantify the biological processes that underly pathogen dynamics, but HPV has received comparatively little attention [147]. HPV models have generally focused on qualitative dynamics [113, 118, 130, 43] and on

predictions for health policy [42, 17], relying on informal methods to estimate parameters. Detailed longitudinal studies present an opportunity to use mechanistic models to rigorously test hypotheses with robust statistical methods. Here we use such methods to show that the prevalence and diversity of HPV types are best explained by the combined effects of a lack of competition within and between types, high rates of reinfection or persistence within individuals, and modest differences in high-risk subpopulations between types.

2.2 Results

2.2.1 Low-prevalence HPV types coexist over time

We fitted models to data from the HPV in Men (HIM) study, which tracks genital HPV infection and demographic and behavioral traits in unvaccinated men sampled at six month intervals over five years [57]. In these data, the mean prevalence of all HPV is 65% but no type has a mean prevalence greater than 10% (Fig. 2.1A), and the prevalence of individual types is roughly constant over time (Fig. 2.1B), demonstrating coexistence. The rank prevalence of HPV types is also constant among geographical locations, although absolute prevalence is higher in Brazil than in the USA or Mexico (Fig. 2.2).

We analyzed the five HPV types with the highest mean prevalences: HPV62, HPV84, HPV89, HPV16, and HPV51. We also analyzed HPV6, the ninth most prevalent type, because it is included in the quadrivalent vaccine and is highly associated with genital warts [59]. We accounted for previously identified risk factors for HPV infection in the HIM dataset [57, 56]. The risk factors for any HPV type include markers of increased exposure to infected sexual partners as well as non-sexual behaviors, such as tobacco use, and demographic traits. Several factors, notably circumcision and sexual orientation, differ in their effects among HPV types [2, 125]. Type-specific differences may reflect meaningful ecological distinctions. We therefore modeled the effects of a diverse set of

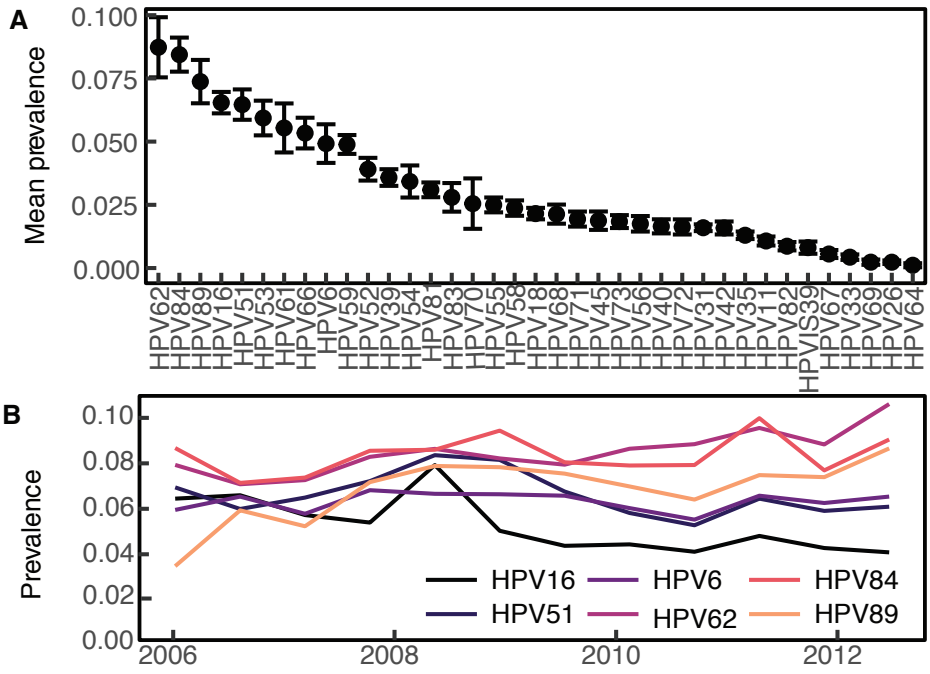

Figure 2.1: **A** Average prevalence of HPV types in the study population, with standard errors calculated across visits. **B** Prevalence over time of the six types analyzed.

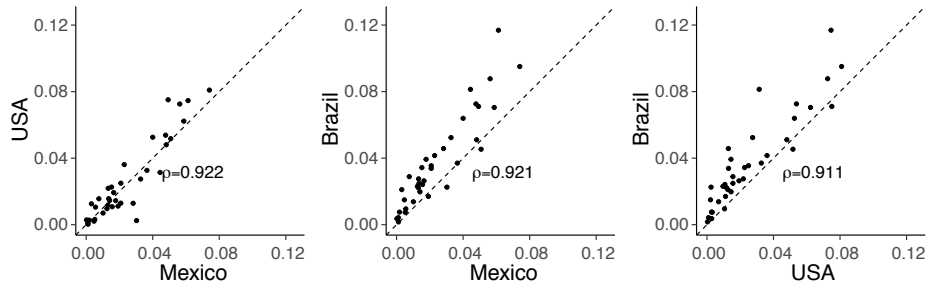

Figure 2.2: Comparison of the time-averaged prevalence of HPV types in the HIM Dataset. The black dotted lines indicates 1:1 correlation, and ρ denotes Pearson's correlation coefficient ($p < 10^{-15}$ with 35 degrees of freedom for each comparison).

risk factors.

2.2.2 *Past infection with HPV confers minimal protection against infection with the same type*

Our models test three hypotheses about HPV infection. Under the simplest, or "memoryless" model, infection risk depends only on the effects of host- and HPV-type-specific risk

factors (Eq. 2.1), with no consideration of immunity. Two more complex models account for the effects of previous infection with the same type (see Methods for model details).

Our three models differ only in their assumptions about the instantaneous per-capita infection risk, or "force of infection" [86], for host i with HPV type j at time t, given by λ_{ijt}:

(i) In the memoryless model, the force of infection depends only on host-specific risk factors and the type-specific baseline force of infection, λ_{0_j}:

$$\lambda_{ijt} = \lambda_{0_j} f(\overrightarrow{\alpha_j} \boldsymbol{C_{it}}), \tag{2.1}$$

The vector $\overrightarrow{\alpha_j}$ scales the effect of each of the M covariates on the force of infection, where $\boldsymbol{C_{it}}$ is the covariate matrix:

$$f(\overrightarrow{\alpha_j} \boldsymbol{C_{it}}) = e^{\alpha_{j,1} C_{1it} + \alpha_{j,2} C_{2it} + \alpha_{j,3} C_{3it} + ... + \alpha_{j,M} C_{Mit}}. \tag{2.2}$$

Each model included five continuous or ordinal covariates and six binary covariates (M = 11 individual-level risk factors).

(ii) In the homologous immunity model, protective immunity reduces the probability of reinfection ($p_{\text{subsequent}}$) relative to the probability of an initial infection (p_{first}) with type j:

$$\begin{aligned} p_{\text{first}} &= (1_{t+\Delta t} | 0_t) = (1 - e^{-\lambda_{0_j} f(\overrightarrow{\alpha_j} \boldsymbol{C_{it}}) \Delta t}), \\ p_{\text{subsequent}} &= p_{\text{first}} (1 - (1 - d) e^{-w(t - t_{\text{clr}})}), \end{aligned} \tag{2.3}$$

The strength of immunity d is constrained to fall between 0 and 1, so that d reduces the probability of reinfection over time interval Δt. Note that for d = 1, the homologous immunity model is identical to the memoryless model. After the previous infection clears at time t_{clr}, immunity wanes at rate w.

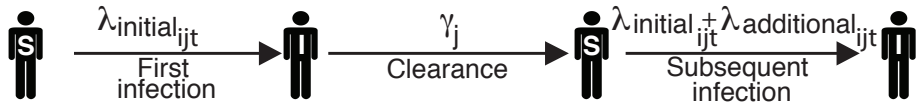

Figure 2.3: Dynamics for one individual host i and one HPV type j under the additional risk model (Eq. 2.4). "S" and "I" denote susceptible and infected, respectively. The duration of each infection is drawn from a Gamma distribution with mean $1/\gamma_j$.

(iii) In the additional risk model, the risk of an initial infection is determined as in the memoryless model (Eq. 2.1), but the risk of reinfection is allowed to be higher. The force of infection thus includes an additional risk factor, which is distinct from the host-specific covariate effects (such as the number of recent sexual partners) and describes only the effect of previous infection (Eq. 2.4, Fig. 2.3):

$$\lambda_{ijt} = \lambda_{0_j} f(\overrightarrow{\alpha_j} \boldsymbol{C_{it}}) + \mathrm{I}_{\mathrm{prev}} d_{jc_i} e^{-w_j(t-t_{\mathrm{clr}})}. \tag{2.4}$$

The additional risk d_{jc_i} depends on the infecting type j and on the sexual subclass c_i of host i. The variable $\mathrm{I}_{\mathrm{prev}}$ indicates whether there was a previous infection with type j.
We included an observation model that determines the probability of the observed data using the sensitivity and specificity of the HPV genotyping test.

We fitted our models to the data using a likelihood-based nonlinear fitting routine [80]. For model selection, we used the corrected Akaike Information Criterion (AICc), which balances the better fit of more complex models against the parsimony of simpler models, while including a correction for finite sample size [76]. The best model has the smallest AICc value.

We first tested whether infection with an HPV type reduces the risk of subsequent infection with the same type, allowing for variation in the degree and duration of protective homologous immunity (Eq. 2.3). Notably, the homologous immunity model reduced to the memoryless model for each HPV type ($d = 1$, Eq. 2.3) at the point estimate of the maximum likelihood. The likelihood of the homologous immunity model for each type was thus effectively identical to that of memoryless model (SI Appendix, Table S1). Small

discrepancies in the likelihoods in the two models arise from the Monte Carlo error associated with the use of simulation-based estimation of the likelihood (SI Appendix 6). The additional parameter in the homologous immunity model therefore increased model complexity without meaningfully improving the likelihood, yielding higher AICc scores. There is thus no evidence for homologous immunity against any of the HPV types that we studied, suggesting that intra-type competition is weak or absent.

Because separate types are less closely related, competition between types should be weaker still. Previous work has nevertheless speculated that there may be cross-immunity between HPV types. In particular, virus-like particles of HPV16 can induce a low level of neutralizing antibodies against HPV31, and clinical trial data suggest that current vaccines, which immunize against HPV16, provide partial protection against HPV31 [39, 87]. We therefore tested for competition between HPV16 and HPV31 by fitting a model in which previous infection with HPV31 affects the risk of infection with HPV16 (see Additional models). Our estimate of the additional risk of HPV16 infection given previous HPV31 infection is centered around 1 (1.3, 95% C.I.: [0.5 , 2.0]), suggesting that the two types do not compete. This lack of an effect may be partly due to the low prevalence of HPV31 (Fig. 2.1) and correspondingly low statistical power, but the lack of even a trend toward competition nevertheless suggests that there is no interaction. We therefore conclude that neither intra-type nor inter-type competition has strong effects on HPV dynamics.

2.2.3 Past infection with HPV strongly increases the risk of future infection with the same type

The additional risk model fits the data vastly better than the homologous immunity and memoryless models for all 6 HPV types (Table 2.1).

To determine whether the additional risk was due to repeated exposure to the same HPV type by infected sexual partners, we fitted the magnitude of the additional risk d sep-

Type	Model (n parameters)	ΔAICc
HPV62	Memoryless (13)	389.6
	Homologous immunity (15)	396.3
	Additional risk (17)	0
HPV16	Memoryless (13)	117.9
	Homologous immunity (15)	123.1
	Additional risk (17)	0
HPV89	Memoryless (13)	137.4
	Homologous immunity (15)	141.2
	Additional risk (17)	0
HPV51	Memoryless (13)	63.9
	Homologous immunity (15)	67.3
	Additional risk (17)	0
HPV84	Memoryless (13)	270.3
	Homologous immunity (15)	276.3
	Additional risk (17)	0
HPV6	Memoryless (13)	121.7
	Homologous immunity (15)	126.8
	Additional risk (17)	0

Table 2.1: Comparison of candidate Models. The ΔAICc gives the AICc score of each candidate model relative to the best-fit model, such that the best-fit model has ΔAICc = 0.

arately for three sexual subclasses: celibate individuals, individuals reporting one recent sexual partner, and individuals reporting multiple recent sexual partners. We included only people who remained in the same sexual subclass for at least three consecutive years. Celibate individuals were defined as reporting no recent receptive or insertional anal, vaginal, or oral sex with male or female partners. Participants were asked separately at each visit about insertional and receptive sexual practices and were asked separately about all sexual practices with male or female partners. Vaginal sex was defined explicitly as ”your penis in your partner’s vagina,’’ anal sex was defined explicitly as ”your penis in your partner’s anus or your partner’s penis in your anus, and oral sex was defined explicitly as ”your penis in your partner’s mouth or your partner’s penis in your mouth.’’

For all six types, previous infection significantly raises the risk of subsequent homologous infection in each sexual subclass (Fig. 2.4A). The high additional risk experienced by

celibate individuals (d_{celibate}) strongly suggests that serial infections are driven by factors other than sexual transmission. Crucially, our estimate of the additional risk is uncorrelated with our estimate of baseline infection risk, λ_{0j} (Fig. 2.5), showing that the high estimated additional risk does not reflect statistical non-identifiability between additional risk and baseline risk. Estimates of the additional risk between different sexual subclasses have the same magnitude (Fig. 2.4A), demonstrating that variation in the additional risk across sexual subclasses is minimal.

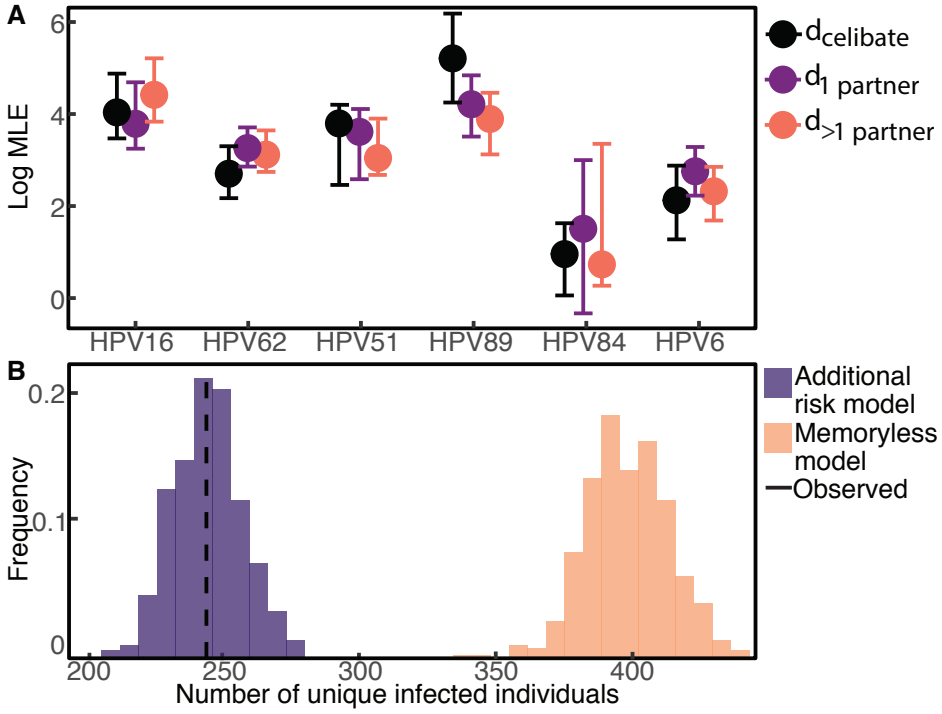

Figure 2.4: **A** Log estimate of the additional risk d by sexual subclass for each HPV type (MLE and 95% confidence interval). **B** Predicted number of unique individuals infected with HPV16 during follow-up across 1,000 simulations under the additional risk model, as compared to the memoryless model and the data.

To quantify the impact of previous infection, we calculated the effect of previous infection on the total risk of a subsequent homologous infection at $t = 0$, 1, and 3 years after clearing the previous infection. Even several years after the initial infection is cleared, the additional risk due to previous infection accounts for more than 90% of the force of infection. Moreover, immediately following infection, the one-year probability of reinfec-

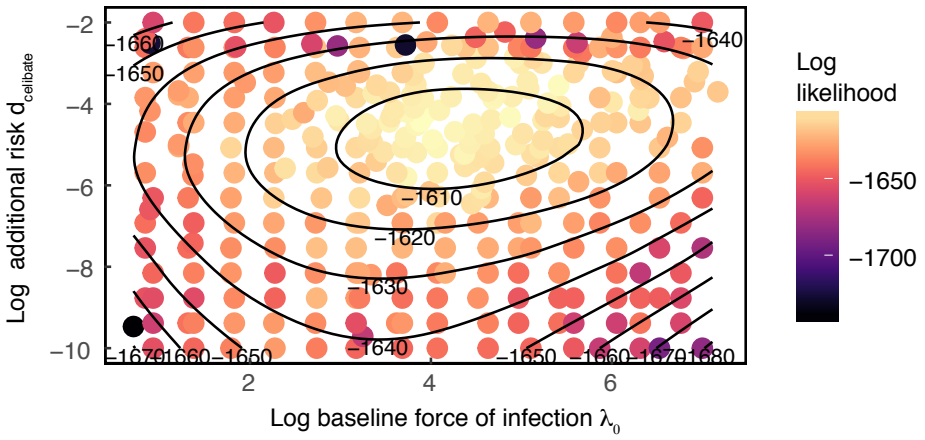

Figure 2.5: Bivariate likelihood profile for the additional risk d_{celibate} and the baseline infection risk λ_0 for HPV16.

tion with HPV16 (Eq. 2.7) is on average 20.4-fold higher than the probability of infection in a naive individual. The average increase is 19.1–20.5-fold among HPV types. Even three years after clearing an infection, the probability of reinfection with HPV16 in the following year remains 13.5-fold higher (7.4–20.5-fold among types).

The strong additional risk independent of sexual subclass is consistent with two major biological explanations: repeat infections, presumably due to autoinoculation between anatomical sites, or episodic reactivation of latent infection. Because our models account for 11 different risk factors besides previous infection, it is unlikely that the increased risk is due to confounding by unmeasured individual-level risk factors. Completely ruling out unmeasured covariates is impossible without controlled experiments. As an initial test, however, we repeated our estimation of the additional risk d using a model that includes additional measured covariates (17 in total). In this more complex model, the additional risk d still accounts for more than 90% of the force of infection following the initial infection, and the model with more covariates did not provide a better fit to the data than the best-fit model with 11 covariates (see Additional models). This result suggests that the increased risk is not due simply to unmeasured individual-level risk factors, while emphasizing the inability of traditional risk factors to explain the vast majority of HPV infections.

2.2.4 *Modest differences in host-specific risk factors suggest ecological differences between HPV types and highlight high-risk subpopulations*

Although the additional risk conferred by past infection is substantial, a model without host-specific risk factors fits the data far worse than the full additional-risk model (Eq. 2.4) for all types (see Additional models). Moreover, the effects of the host-specific risk factors vary among HPV types. To understand this variation, we inserted our estimates of the baseline force of infection and the covariate effects for each HPV type into the best-fit model to calculate the distribution of infection risk in individuals who have never been infected. The expected time to infection ($1/\lambda_{ijt}$), a measure of risk, is generally low but varies by orders of magnitude among individuals in the naive population (Fig. 2.6).

Thus, the risk of initial infection with any type is concentrated in a few high-risk individuals. This feature captures a major pattern in the data. Under the additional risk model, the population prevalence of an HPV type arises from repeated infections in relatively few individuals. In the memoryless model, by contrast, the initial infection risk is higher, previous infection has no effect on subsequent risk, and prevalence arises from fewer infections in more individuals.

Simulations from the maximum likelihood estimates of the two models demonstrate that both reproduce the population-level prevalence in the data across visits (Fig. 2.7). However, only the additional risk model accurately predicts the fraction of individuals that ever experience infection, which the memoryless model overestimates (Fig. 2.4B).

To assess the ability of the best-fit model to capture the observed dynamics, we simulated infection data for the n = 1,099 individuals that we included when making inferences, preserving the visit dates and covariate data for each individual. We compared the results of 1,000 simulations from both the memoryless model and the additional risk model (the best-fit model) to the data. Both the memoryless model and the additional risk model captured the observed prevalence over patient followup (Fig. 2.7), but the additional risk model accurately predicted the total number of unique individuals that were

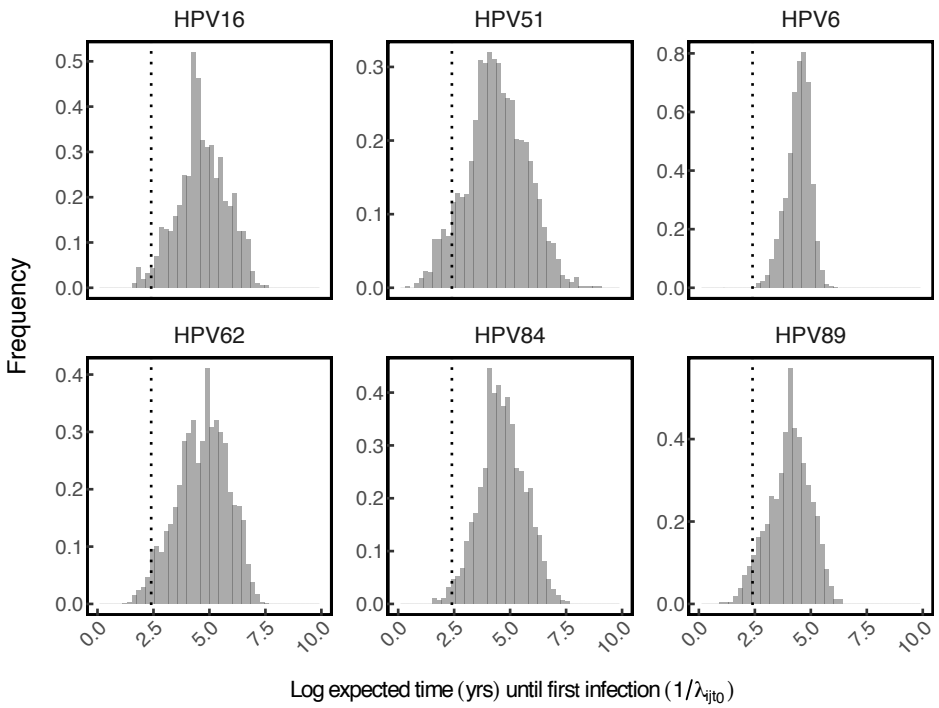

Figure 2.6: Distribution of the (log) baseline expected time to infection ($t_{\text{expected}} = 1/\lambda_{ij}$) in uninfected individuals assuming no prior infection, such that $\lambda_{ij} = \lambda_{0_j} f(\overrightarrow{\alpha_j} \boldsymbol{C_{i0}})$. For each individual, λ_{ij} was calculated using the maximum likelihood estimate for each element in $\overrightarrow{\alpha}$ and the individual-specific covariates $\boldsymbol{C_i 0}$, which were reported at the baseline visit ($t = 0$). The y axis reports frequency, while the vertical dashed line in each panel marks an expected time to infection of 10 years. Thus, only the portion of the distribution to the left of the dashed line in each panel represents individuals for whom the the expected infection time falls within the next ten years.

infected at any point during the study. The memoryless model, in contrast, overestimated the number of unique infected individuals (main text Fig. 2.4B). The best-fit model thus captures important qualitative aspects of the dynamics, supporting the results of model selection. The success of the additional-risk model provides strong support for our use of a Bernoulli likelihood function (Eq. 2.7). An implicit assumption of the Bernoulli likelihood function is that the error in the model's predictions is due only to measurement error rather than to systematic bias in the model's predictions. The ability of the best model to reproduce both the prevalence data, and the number of unique individuals infected, strongly suggests that any such systematic bias is reasonably weak.

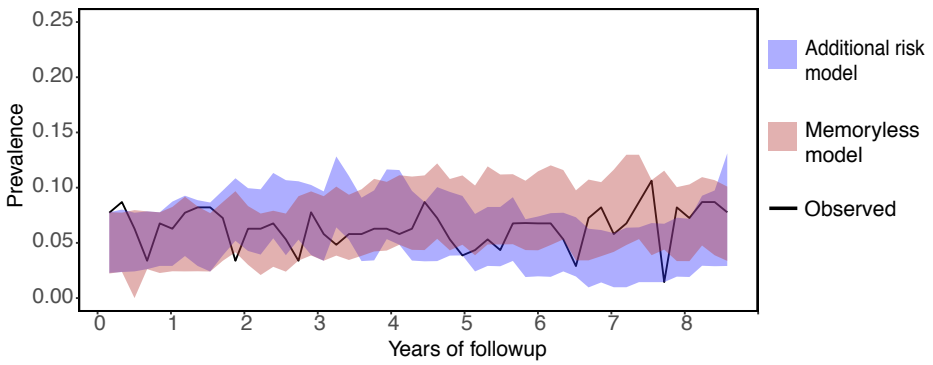

Figure 2.7: Comparison of the dynamics under candidate models with the observed HPV16 infection data. Fraction of infected individuals at various time-points during followup. The shaded areas are bounded by the 2.5% and 97.5% quantiles from 1,000 simulations of each model. The solid line gives the observed fraction infected among the individuals used to simulate the data.

The success of the additional-risk model provides strong support for our use of a Bernoulli likelihood function (Eq. 2.7). An implicit assumption of the Bernoulli likelihood function is that the error in the model's predictions is due only to measurement error rather than to systematic bias in the model's predictions. The ability of the best model to reproduce both the prevalence data, and the number of unique individuals infected, strongly suggests that any such systematic bias is reasonably weak.

Unmeasured assortative mixing [99] and simultaneous exposure to multiple HPV types through sexual partnerships [109] can confound associations between HPV incidence and past exposures. Residual confounding may be a problem because the initial infection risk is low and sensitive to host-specific risk factors. However, such confounding would be minimized in celibate individuals. We therefore re-estimated the magnitude of the additional risk d_{celibate} by fitting the additional risk model (Eq 2.4) to data from celibate individuals only. In the celibates-only analysis, the risk conferred by previous infection was still high (see Additional models) and consistent with our previous estimates. The high additional risk inferred for the celibate individuals and the consistency of the additional risk for all three sexual subclasses suggest that the additional risk reflects autoinoculation or reactivation rather than confounding from unmeasured sexual

activity.

The effects of each covariate also show interesting similarities and differences between types (Fig. 2.8).

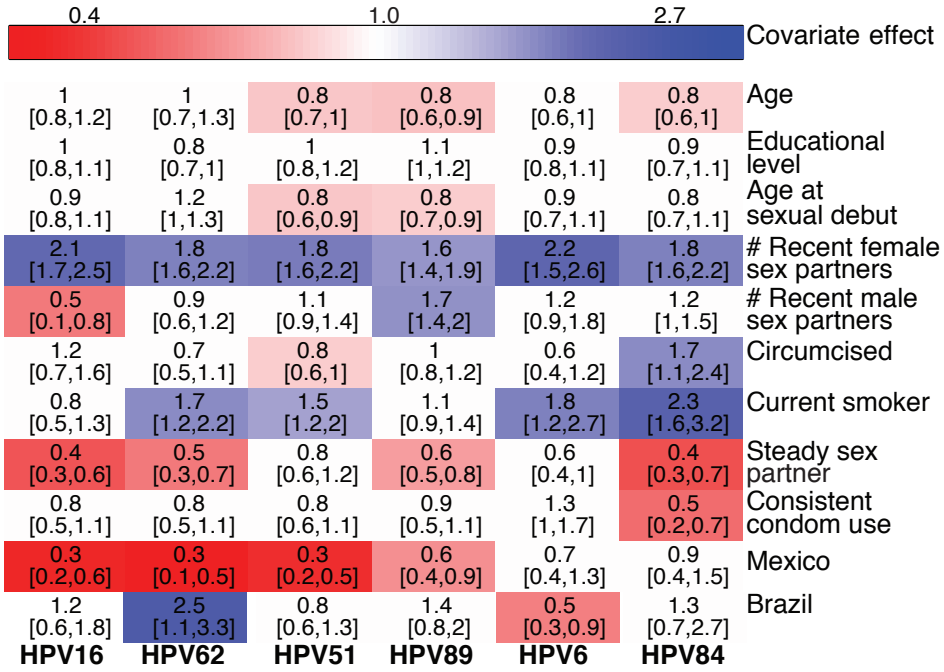

Figure 2.8: Inferred effect of host-specific covariates (maximum likelihood parameter estimate and the 95% confidence interval) Colored cells denote statistically significant positive(blue) and negative (red) effects.

Having more recent female partners strongly increases risk for all HPV types, emphasizing the importance of heterosexual transmission. The addition of a single female sexual partner raises instantaneous infection risk by 80% to 120% among types compared to the risk in individuals with no recent female partners. The number of male partners, however, has divergent effects. The addition of a single male sexual partner *reduces* instantaneous risk with HPV16 by 50%, but *increases* the risk of infection with types HPV84 and HPV89 by 20% and 70%, respectively. Most other covariates were significant for only some types, indicating subtle differences between types.

Our estimates of the effects of type-specific risk factors are generally consistent with previous studies, but there are some exceptions. First, a previous analysis of the HIM

data showed no difference in HPV16 incidence between men who have sex with men and men who have sex with women [125], but our results instead show that risk of HPV16 infection declines with an increase in the number of male sex partners, suggesting that HPV16 infection depends largely on heterosexual transmission. Meanwhile, our results show that the risk of infection with HPV51, HPV89, HPV6, and HPV84 increases with the number of male or female partners, consistent with previous work, although the effect was statistically significant only for HPV89. Finally, the previous analysis concluded that circumcision increases the risk of HPV51 [2], but our results suggest that circumcision instead reduces the risk of HPV51. Our modeling approach, however, may provide greater statistical power than previous analyses. First, our models account for variation in risk per unit time, and they describe the dynamics in a way that is unconstrained and unbiased by the frequency of observations. Second, we infer the contribution of host-specific risk factors to infection risk separately from the dynamics of previous infection. We therefore have reason to believe that our results are robust and that differences with previous analyses arise from differences in model structure.

2.2.5 Estimated parameters for the best-fit model

The values of the maximum likelihood estimate for the best-fit model parameters are given in Table 2.2.

2.2.6 Additional models

1. *Testing for homologous immunity by excluding individuals with prevalent infection*
   Including individuals infected with HPV at baseline could have enriched the data for individuals with less immunity to HPV, causing the additional risk model to be favored over the homologous immunity model. Therefore, to help validate the results of the model selection, we fitted each of the three candidate models (the

Parameter	MLE	Confidence interval	Type
λ_{0j}	-4.6	[-5.2,-4]	HPV62
	-4.6	[-5.2,-4.3]	HPV84
	-4.3	[-4.7,-3.6]	HPV89
	-4.2	[-5.1,-3.5]	HPV16
	-4.3	[-5,-3.8]	HPV51
	-5.1	[-5.7,-4.5]	HPV6
p_{initial} (logit scale)	-2.6	[-2.8,-2.5]	HPV62
	-2.6	[-2.7,-2.5]	HPV84
	-2.8	[-3.1,-2.6]	HPV89
	-2.9	[-3.2,-2.8]	HPV16
	-2.9	[-3.1,-2.8]	HPV51
	-3.2	[-3.4,-2.8]	HPV6
d_{celibate}	2.7	[2.2,3.3]	HPV62
	1.0	[0.1,1.6]	HPV84
	5.2	[4.3,6.2]	HPV89
	4.0	[3.5,4.9]	HPV16
	3.9	[2.4,4.2]	HPV51
	2.1	[1.3,2.9]	HPV6
$d_{1\text{ partner}}$	3.3	[2.9,3.7]	HPV62
	1.5	[-0.3,3]	HPV84
	4.2	[3.5,4.8]	HPV89
	3.8	[3.2,4.7]	HPV16
	3.6	[2.5,4.1]	HPV51
	2.8	[2.2,3.3]	HPV6
$d_{>1\text{ partner}}$	3.1	[2.7,3.6]	HPV62
	0.7	[0.3,3.4]	HPV84
	3.9	[3.1,4.5]	HPV89
	4.4	[3.8,5.2]	HPV16
	3	[2.7,3.9]	HPV51
	2.3	[1.7,2.9]	HPV6
w	-0.3	[-0.5,-0.1]	HPV62
	-0.5	[-0.8,-0.3]	HPV84
	0	[-0.3,0.2]	HPV89
	0.3	[0.1,0.5]	HPV16
	-1.3	[-1.6,-0.8]	HPV51
	0	[-0.2,0.3]	HPV6

Table 2.2: Values of the estimated parameters. Estimates are on a log scale unless otherwise indicated.

memoryless model, main text Eq. 1, the homologous immunity model, main text Eq. 3, and the additional risk model, main text Eq. 4) to the subset of individuals

that were uninfected for at least two visits at baseline. We found strong support for the additional risk model in this subpopulation (Table 2.3, consistent with our major findings. For each type, the homologous immunity model again reduced to the memoryless model (the maximum likelihood was achieved at $d = 1$ for all types), affirming the lack of homologous immunity in any type.

Type	Model (n parameters)	ΔAICc
HPV62	Memoryless (13)	52.5
	Homologous immunity (15)	55.3
	Additional risk (17)	0
HPV16	Memoryless (13)	17.8
	Homologous immunity (15)	21.5
	Additional risk (17)	0
HPV89	Memoryless (13)	22.0
	Homologous immunity (15)	25.3
	Additional risk (17)	0
HPV51	Memoryless (13)	33.1
	Homologous immunity (15)	38.4
	Additional risk (17)	0
HPV84	Memoryless (13)	47.9
	Homologous immunity (15)	51.1
	Additional risk (17)	0
HPV6	Memoryless (13)	30.6
	Homologous immunity (15)	34.1
	Additional risk (17)	0

Table 2.3: Comparison of the candidate models fit to the subset of the study of population without prevalent infection at baseline. The ΔAICc gives the AICc score of each candidate model relative to the best-fit model, such that the best-fit model has ΔAICc = 0.

2. *Additional risk model fit to celibate individuals only*

   We fitted the additional risk model to data from the celibate individuals ($n = 237$). Celibate individuals were those who reported no receptive or insertional vaginal, anal, or oral sex between visits. For all analyses, we included only individuals that remained celibate for a minimum of three years, and we estimated the magnitude of the additional risk dcelibate only during the period of celibacy. Consistent with our main findings, when fitting the model to only celibate individuals, we recover

the strong effect of previous infection with each HPV type on the risk of subsequent infection with the same type (Table 2.4). The effect of previous infection on the total risk of subsequent infection in the celibate population accounts on average for over 96% of the infection risk across types, even at three years after infection clearance.

Type	d_{celibate} (MLE)	95% Confidence Interval
HPV62	1.6	[0.1, 3.4]
HPV84	1.5	[0.4,9.8]
HPV89	5.8	[1.1,10.1]
HPV16	3.4	[1.3,5.4]
HPV51	3.0	[0.9,6.8]
HPV6	2.5	[0.8, 5.2]

Table 2.4: Maximum likelihood estimate of the additional risk d_{celibate} inferred using data from only celibate individuals. All values are reported on a log scale.

3. *Additional risk only model*

   In this model, the force of infection λ_{ijt} for individual i with type j at time t was determined only by the baseline infection risk for type j and by the effect of previous infection, so that the behavioral and demographic risk factors had no effect. The force of infection is then:

$$\lambda_{ijt} = \lambda_{0_j} + \mathrm{I}_{\text{prev}} d_{jc_i} e^{-w_j(t-t_{\text{clr}})} \tag{2.5}$$

   For each HPV type, this model out-performed the memoryless model but it performed much worse than the model that also took into account the covariates (Table 2.5).

4. *HPV16/HPV31 interaction model*

   To test whether infection with HPV31 affects the risk of infection with HPV16, we fitted a model in which the force of infection of HPV16 depends on whether an individual has ever been infected with HPV31. In this model, we included a covariate variable $\mathrm{I}_{\text{HPV31}}$, updated at each visit, that indicated whether individual i was

Type	ΔAICc
HPV16	50.6
HPV51	11.7
HPV6	8.3
HPV62	66.8
HPV84	37.3
HPV89	22.5

Table 2.5: Performance of the additional risk only model for each HPV type relative to the full additional risk model. The ΔAICc gives the AICc score of each candidate model relative to the best-fit model, such that the best-fit model has ΔAICc = 0.

currently or previously infected with HPV31. The force of infection λ_{ijt} was thus:

$$\lambda_{ijt} = \lambda_{0_j} f(\overrightarrow{\alpha_j} \boldsymbol{C_{it}}) \alpha_{\text{HPV31}} \text{I}_{\text{HPV31}} + \text{I}_{\text{prev}} d_{jc_i} e^{-w_j(t-t_{\text{clr}})}, \tag{2.6}$$

where α_{HPV31} gives the effect of previous or current infection with HPV31 on the risk of infection with HPV16. Our goal was to estimate the direction and strength of interaction between HPV16 and HPV31. Our estimate of the interaction parameter was centered around 1 (MLE 1.3, CI[0.5,2.0]), indicating no significant interaction.

5. *A model with additional covariates*

   To estimate the contribution of the additional risk to the force of infection when allowing for a large number of host risk factors, we added additional covariates to the additional risk model for HPV16, using the full set of 17 original candidate covariates (Fig. 2.10). We used the maximum likelihood estimates of the model parameters to estimate the fraction of the force of infection made up by the additional risk d among individuals. This model involved inference for n=1019 individuals that had complete covariate information at baseline for the full set of covariates. To compare this model to the best-fit additional risk model (with 11 covariates), we fitted the original additional risk model for the same n=1019 individuals and again calculated the maximum likelihood. The additional risk d still accounts on average

for over 90% of the risk for several years after infection clearance (Fig. 2.9).

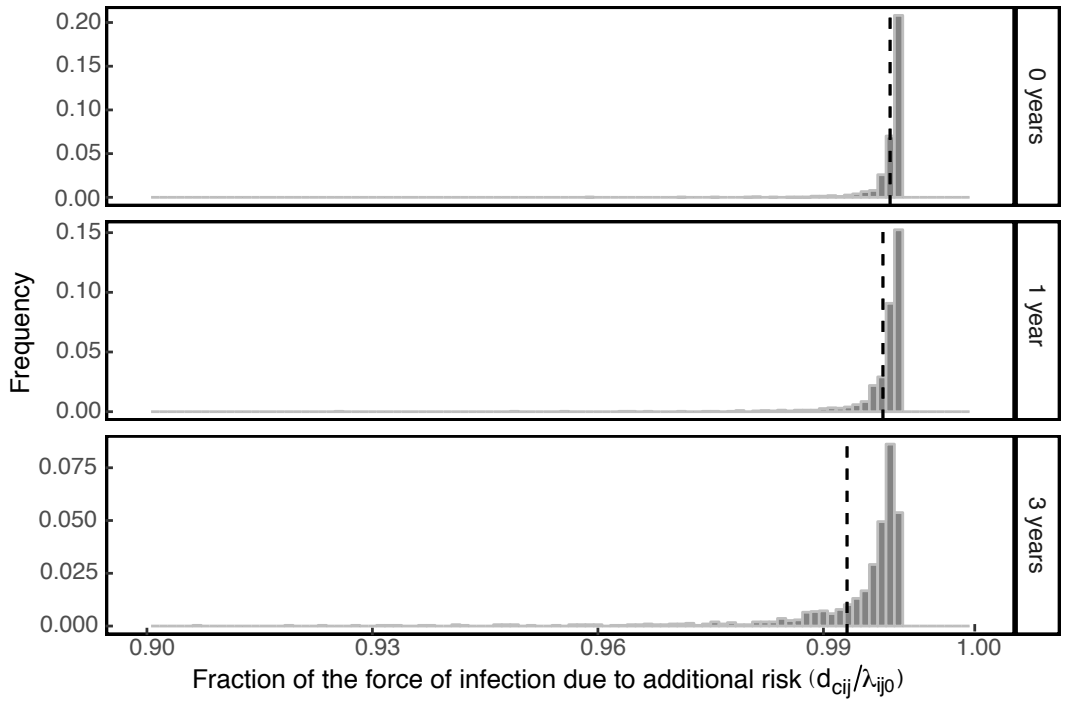

Figure 2.9: Contribution of the effect of previous infection to the overall infection risk among individuals for HPV16, using a model with 17 covariates. Histograms show the distribution of the fraction of the overall force of infection λ_{ijt} made up by the additional risk d in the population at various times post-clearance of the precedent infection. The mean value for each distribution is given by the vertical dotted line.

The maximum likelihood and AICc for the model with additional covariates were -1468.4 (SE .63) and 2983.9, respectively, whereas the maximum log likelihood and AICc for the model without additional covariates, fitted to the data from the same individuals, were -1472.8 (SE .86) and 2979.6, respectively. Thus, the ΔAICc for the additional covariates model is 4.3, showing that the improvement in the likelihood from the additional complexity of the additional covariates does not provide a better explanation of the data than the additional risk model.

2.3 Discussion

Understanding the dynamics of HPV transmission is crucial to explain the coexistence of over 200 low-prevalence HPV types. Our results show little evidence of competition within and between HPV types, and instead they show high rates of reinfection or persistence by the same type. The inferred lack of both homologous and heterologous immunity implies that HPV type diversity cannot be explained by negative frequency-dependent selection, which promotes the coexistence of other immunologically distinct pathogen strains. Our results instead suggest that high HPV prevalence results from continuous or repeated infection with multiple, apparently independent virus types. Types differ slightly in their risk factors, suggesting that ecological niche partitioning could further promote coexistence.

Our conclusion that there is no homologous immunity in men supports the hypothesis that genital infection differs between men and women. Although the durations and distributions of types in genital HPV infection are similar in men and women [71, 59], the prevalence of genital HPV is higher in men [71]. Acquired immunity has been proposed to explain declining cervical HPV prevalence by age in women in some countries [15, 149], whereas HPV prevalence does not change with age in men [57, 71]. Furthermore, the seroprevalence of some types is higher in women than in men from the same source population [111], although homologous protection in women is still likely to be weak [8].

The high risk of recurring infection is consistent with either autoinoculation across anatomic sites or with episodic reactivation of latent infection. Studies showing type-concordant HPV infection across anatomic sites support the importance of autoinoculation. First, HPV DNA has been detected on the fingers of patients with genital warts [154], suggesting transmission between the hands and the genitals. Second, an analysis of a subcohort of the HIM study demonstrated a 1.5 to 15-fold increase in the risk of anal HPV infection after genital infection [125], and a similar analysis of women found a

20-fold increase in the risk of recurrent anal HPV infection given infection at the cervix [60]. Moreover, 63% of the cases of recurrent anal HPV occurred without a self-reported history of anal sex [60]. Significantly, the magnitude of the effect in these studies is consistent with our estimate that previous infection yields a 20-fold increase in the risk of subsequent infection.

Apparent reinfection could instead be due to the reactivation of latent virus. Whether HPV persists latently remains unknown [147, 64], but an animal model of cottontail oral papillomavirus provides evidence of a latent reservoir [108]. Furthermore, sexually inactive HIV-positive women have a higher risk of recurrent HPV infection compared to HIV-negative controls [161], and among HIV-positive females, higher CD4+ T cell counts are negatively correlated with recurrent infection risk [157]. The latter two studies have been interpreted as evidence of reactivation of latent infection [64], but either effect may instead result from increases in susceptibility to autoinoculation. Furthermore, the studies in HIV-positive women suggest that reactivation requires immune suppression, and may therefore be rare in healthy individuals.

Vaccine efficacy trials suggest that reinfection is more common than reactivation. Women with previous HPV-related disease who received the quadrivalent vaccine were protected against HPV-related lesions [84]. Similarly, one study showed 100% efficacy of the HPV vaccine against HPV-related disease in individuals with serological evidence of past HPV 6, 11, 16, or 18 infection [129]. HPV vaccines prevent infection by inducing antibodies that block viral entry, and such antibodies would likely not prevent reactivation. If indeed vaccines do not affect the disease course of reactivated infections, then these trials suggest that the vaccine diminishes the incidence of HPV lesions by blocking true reinfection. Vaccinating children before sexual debut clearly reduces infection rates [110], but if vaccination prevents autoinoculation, then vaccinating previously infected individuals may also reduce HPV prevalence.

The type-specific effects of demographic and behavioral risk factors that we inferred

suggest that modest differences exist in the host subpopulations supporting each HPV type. These differences highlight functional ecological distinctions between types that may further contribute to type coexistence For all types, the major determinant of infection risk is the number of recent female sex partners, suggesting a central role for heterosexual transmission. Additionally, current smoking increases risk in most types, consistent with other studies [56, 58, 71, 134]. Smoking can suppress mucosal and cellular immunity [98], but its effects may be confounded with other potentially unobserved, high-risk behaviors. Although shared risk factors account for most of the initial infection risk for each type, there are important distinctions too. For instance, the effect of circumcision on the risk of HPV infection is type-specific. Several randomized, controlled trials of male circumcision have demonstrated that circumcision protects against any HPV infection [6, 65], whereas observational studies have instead documented either no effect or increased risk [71, 105]. These conflicting results may have arisen from confounding by differences between types. Foreskin consists of unkeratinized mucosal epithelium, whereas the exposed tissue in circumcised men is made up of stratified squamous epithelium [65]. Our finding that circumcision increases the risk of infection with some HPV types (e.g. HPV84) and decreases the risk with others (e.g. HPV51) could reflect the adaptation of HPV types to different epithelia in the male genital tract.

Our modeling approach has several limitations, notably that we cannot distinguish between reactivation and autoinoculation, and that unmeasured covariates could have affected our results. While we identified celibate individuals based on detailed self-reporting on anal, vaginal, and oral sex, celibacy did not account for all forms of non-penetrative sexual contact, such as kissing and masturbation. Furthermore, the construction of covariates from risk-factor data may have led to bias. While we based our selection and specification of the covariates on previous analyses, we cannot ignore the possibility of residual bias. For example, for recent male sex partners, if any sexual activity with male partners significantly affected risk, but the distinction between two and three recent

partners was negligible, then we could have underestimated the effect of male sexual partnership activity by inferring a linear coefficient. Furthermore, we inferred a linear coefficient to estimate the general effect of increasing educational level on HPV risk, but changes between discrete educational levels may correspond imperfectly to proportional changes in risk, which may have limited our power to detect a true effect. Dichotomizing condom use into more or less than 50%, a necessary measure given the small fraction of individuals that reported 100% condom use, could have introduced bias. Particularly, we may underestimate the protective effect of condoms if they are only strongly effective at 100% usage. However, at least one previous study identified a significant reduction in the risk of HPV infection with condom use when condom use was measured as use more than 50% of the time [172]. We cannot completely rule out the possibility that unobserved properties of the sexual contact network affected our results. In addition, our model of homologous immunity assumes that protection arises immediately after infection, but protection may be lagged [118].

2.4 Methods

2.4.1 Data

The HPV in Men (HIM) study is a multinational cohort study of HPV infection in unvaccinated men aged 18-70 years. The study enrolled 4,123 participants between 2005 and 2009 and tracked men longitudinally over five years. Men were recruited from three cities: Tampa, Florida, USA; Cuernavaca, Mexico; and Sao Paulo, Brazil. Detailed study methods are described elsewhere [57]. The human subjects committees of the University of South Florida (Tampa, FL, USA), Centro de Referencia e Treinamento em Doenabas Sexualmente Transmisseveis e AIDS (Sao Paulo, Brazil), and Instituto Nacional de Salud Publica de Mexico (Cuernavaca, Mexico) approved all HIM study procedures. All participants provided written informed consent. The data for each individual consist of binary

time series of infection status with each HPV type for up to 10 clinic visits (median = 10 visits).

We excluded individuals that failed to meet the full eligibility criteria described by the HIM study [57], which included no prior diagnosis of genital cancer, warts, HIV, or other STIs. We identified 3,656 eligible participants from the 4,123 men enrolled in the HIM study as of October 2014. Of the eligible participants, we excluded $n = 575$ participants that had missing data at the time of enrollment for the 11 covariates that we included in our three candidate models. We then divided the remaining 3,081 participants into three sexual subclasses based on the number of recent sex partners. For all covariates, "recent" activity indicates activity in the past six months, or since the last clinic visit if the last clinic visit was more than six months prior, as reported at each clinic visit. The three sexual subclasses were celibate individuals, individuals with one recent sex partner, and individuals with two or more recent sex partners. To account for the effects of sexual subclass on the infection risk, we restricted our analysis to include only the $n = 1,099$ individuals who remained in the same subclass for at least three years over the course of their participation in the study.

2.4.2 Covariate variables

Covariates, including those used to determine sexual subclass were derived from a risk factor questionnaire that was administered at each visit, which was described and validated previously [57, 126]. The survey was conducted via computer-assisted self- interviewing (CASI) to preserve patient anonymity during the response process. Questions covered sociodemographic characteristics, sexual behaviors, sex partnerships, and condom use. Participants were asked to recall their recent behavior, such as recent number of male or female sexual partners, where "recent" referred to behavior over the past six months. Participants had the option of refusing to answer any question, and refusals were treated as missing values as in Giuliano et al. [57], such that a missing covari-

ate was assigned its value at the closest visit. Covariates were selected for inclusion in the model based on known risk factors identified in the literature for HPV in men [57, 56, 59, 125, 71, 2]. We also included country of residence as a covariate. The full set of candidate covariates (Fig. 2.10) included race (black/African American or other), ethnicity (Hispanic or other), age, age at sexual debut, lifetime numbers of male and female sexual partners, numbers of recent male and female sexual partners, numbers of new male and female sexual partners, presence of a steady sexual partner, marital status, level of education, circumcision status (confirmed by a clinician at the baseline visit), whether or not the participant was a current smoker, whether or not the participant used condoms for the majority of recent sexual encounters, Brazilian nationality, and Mexican nationality, with US nationality used as the baseline.

To decrease statistical non-identifiability and to increase computational feasibility, we reduced the full candidate set of covariates to a subset based on observed correlations (Fig. 2.10). Among highly correlated pairs of similar covariates, we heuristically selected one representative. For example, we chose to include recent numbers of sexual partners instead of lifetime numbers of sexual partners and the presence of a steady sexual partner instead of marital status. The variables describing race and ethnicity were strongly correlated with country of origin, so we excluded them. Table 2.6 gives the final subset of covariates included in the model.

Through this selection process, we identified five continuous covariates that we included as linear predictors (Table 2.6). The continuous and ordinal covariates that we modeled as linear predictors were age, educational level, age at sexual debut, number of recent female sex partners, and number of recent male sex partners.

Following previous analyses from the HIM study group [125, 56], we specified these covariates according to a previous study [126] that evaluated the test-retest reliability of the risk factor questionnaire across the three study sites and languages. Nyitray et al. identified the covariate specifications that maximized the calculated reliability coefficient

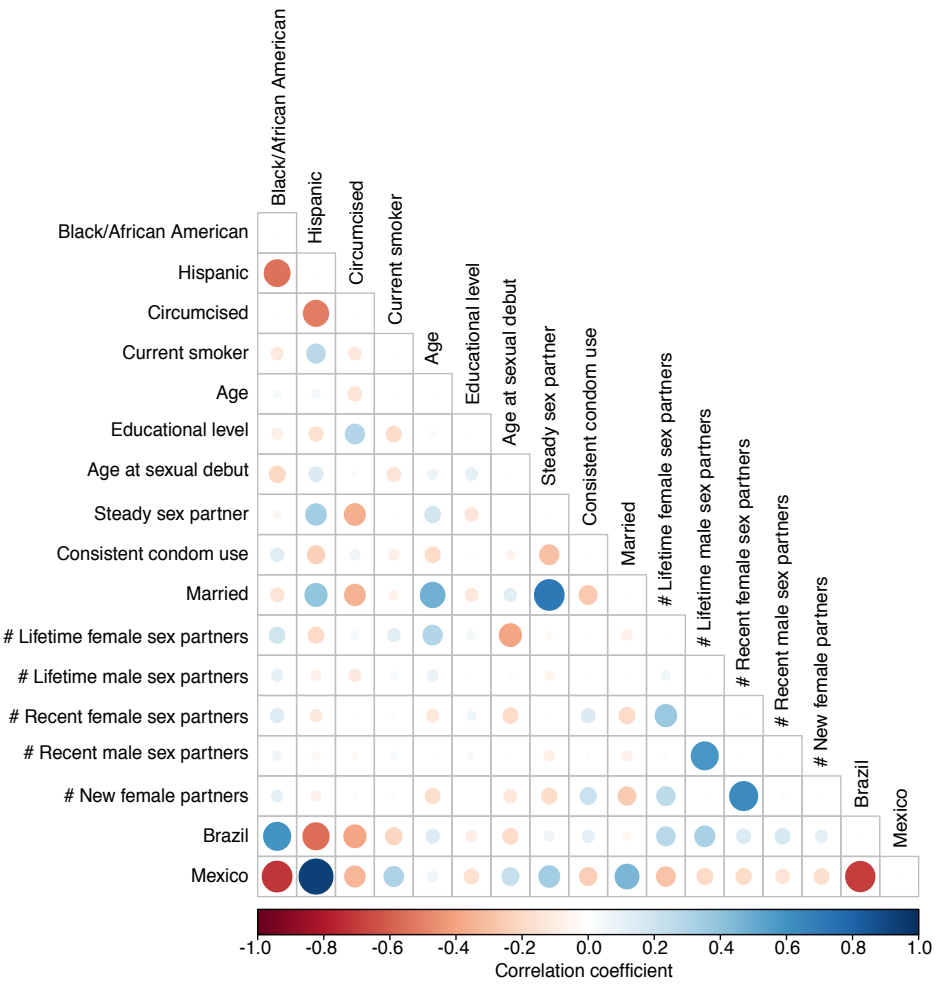

Figure 2.10: Correlations between the full candidate set of self-reported covariates at the baseline visit. Pearson product-moment correlations were calculated between continuous variables, polyserial correlations (inferred latent correlations between continuous and categorical variables) were calculated between numeric and binary variables, and polychoric correlations (inferred latent correlations between categorical variables) were calculated between binary variables. All correlations shown were significant at the $\alpha = .05$ significance level based on tests for bivariate normality.

of the response between geographical sites. For recent sexual partnerships, for example, they found that across-site and across-time reliability was maximized when recent sexual partnerships were grouped as 0, 1, 2, and ≥ 3 sexual partners of either gender over the previous six months. We preserved this variable grouping and found that less than 4% of the individuals in our data reported >3 recent sexual partnerships at any visit. Binning of ≥ 3 sexual partnerships thus likely did not affect the estimated covariate effect. For

educational level, the HIM study group identified a binned ordinal set of educational levels (0-4) that were similar across cultures. Nyitray et al. found that covariate responses had substantial reliability regardless of participant age, and the previous analyses of the HIM dataset found no association of age with HPV incidence across different binnings of age. Furthermore, previous analyses identified a simple positive relationship between HPV incidence and younger age at sexual debut [56]. We therefore used age and age at sexual debut as continuous predictors without binning.

To model the effect of the covariates on the instantaneous risk of infection λ_{ijt}, we used a log-linear model (Eq. 2.4), such that each covariate effect α is directly interpretable as a multiplicative effect on λ_{ijt}. Time-varying covariates (age, recent numbers of male and female sex partners, the presence of a steady sex partner, smoking status, educational level, and condom use) were updated at each visit.

	Name	Structure	Notes
1	Age	Continuous	
2	Educational Level	Ordinal	
3	Age at sexual debut	Continuous	
4	Recent female sexual partners	Continuous	0,1,2, ≥ 3
5	Recent male sexual partners	Continuous	0,1,2, ≥ 3
6	Circumcision status	Binary	
7	Smoking status	Binary	Current smoker
8	Steady sex partner	Binary	Presence of a steady sexual partner
9	Consistent condom use	Binary	$\geq$50% condom use
10	Brazil	Binary	Country of residence
11	Mexico	Binary	Country of residence

Table 2.6: Covariates included in the analyses.

2.4.3 Model of HPV dynamics

Our models are two-state discrete time partially observed Markov processes (POMPs), in which an individual is either infected (1) or uninfected (0) at any time with an HPV type. Infection of individual i with HPV type j occurs at rate λ_{ijt}, such that the probability of

infection over Δt is:

$$p(1_{t+\Delta t}|0_t) = 1 - e^{-\lambda_{ijt}\Delta t}. \tag{2.7}$$

Because we cannot directly observe infection events, the measurement model relates each observation Y_{ijt} to the latent state X_{ijt}:

$$Y_{ijt} \sim \text{Bernoulli}(X_{ijt}(1 - P_{\text{FP}}) + (1 - X_{ijt})P_{\text{FN}}), \tag{2.8}$$

where P_{FP} and P_{FN} are the rates of false positives and false negatives, respectively. The duration of infection was drawn from a gamma distribution, where the shape (k_j) and scale (θ_j) were fixed to match the empirical distribution of infection durations in the data for type j (Table 2.8 and Fig. 2.11).

2.4.4 Model parameters

1. *Estimated parameters*

   A description of the estimated parameters is given in Table 2.7.

2. *Fixed parameters*

   The false positive and false negative rates of HPV detection were set equal to the high sensitivity (96%) and specificity (99%) of the Roche Linear Array HPV genotyping test reported by the manufacturer (Roche Diagnostics), which has been confirmed by other analyses [63]. The duration of each simulated infection with each HPV type j was drawn from a $\Gamma(k_j, \theta_j)$ distribution, where k_j and θ_j were fixed according to the empirical distribution of infection durations in the data for type j (Fig. 2.11, Table 2.8).

   Following Giuliano et al. [57], we required two consecutive negative visits following a positive visit for any HPV type to constitute an observed clearance. Of the

Parameter group	Name	Type
Baseline infection risk	λ_{0j}	Estimated
Infection duration	Gamma shape (k_j)	Fixed
Infection duration	Gamma scale (θ_j)	Fixed
Covariate effects (α)	Age	Estimated
Covariate effects (α)	Educational status	Estimated
Covariate effects (α)	Age at sexual debut	Estimated
Covariate effects (α)	Circumcision status	Estimated
Covariate effects (α)	No. recent female sex partners	Estimated
Covariate effects (α)	No. recent male sexual partners	Estimated
Covariate effects (α)	Steady sexual partner	Estimated
Covariate effects (α)	Consistent condom use	Estimated
Covariate effects (α)	Current smoking status	Estimated
Covariate effects (α)	Mexican nationality (α_{MX})	Estimated
Covariate effects (α)	Brazilian nationality (α_{BZ})	Estimated
Initial conditions	Probability of initial infection (P_{initial_j})	Estimated
Initial conditions	Remaining ($F_{\text{remaining}}$)	Nuisance
Initial conditions	Previous infection probability (p_{past})	Nuisance
Measurement model	False positive rate (P_{FP})	Fixed
Measurement model	False negative rate (P_{FN})	Fixed
Homologous immunity	Magnitude of scaled risk (d)	Estimated
Homologous immunity	Rate of waning of immunity (w)	Estimated
Additional risk (d)	d_{celibate}	Estimated
Additional risk (d)	$d_{1\text{partner}}$	Estimated
Additional risk (d)	d_{multiple}	Estimated

Table 2.7: Description of model parameters

Type	k_j	θ_j
HPV62	1.20	0.81
HPV84	1.67	0.59
HPV89	1.37	0.70
HPV16	1.71	0.52
HPV51	1.33	0.75
HPV6	1.71	0.51

Table 2.8: Values of the parameters describing infection durations, as estimated from observed infection and clearance events in the data.

consecutive negative tests, the first was assumed to be the first true negative observation. Thus, we treated 1-0-1 transitions as false negatives, thereby adjusting data

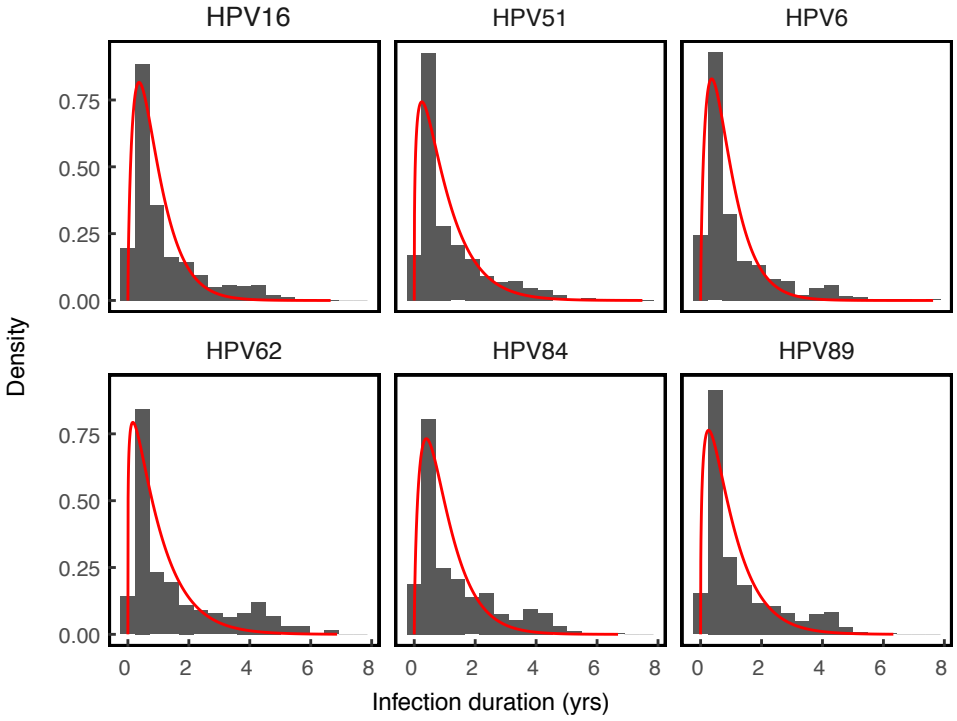

Figure 2.11: Empirical distribution of infection durations for each HPV type (grey histogram) with an overlying gamma distribution having the same mean and coefficient of variation as in the data (red).

a priori based on the assumption that infection in such cases was actually continuous. Our justification was two-fold: first, this approach allowed us to account for false negatives in HPV sampling beyond the laboratory specifications of the HPV test that were included in the observation model. Second, the approach ensured that the high rates of reinfection estimated by the best-fit model were not the result of failing to account for false clearances, ensuring in turn that our estimates of reinfection rates would be conservative.

3. *Nuisance parameters*

   Two nuisance parameters were used to generate the initial conditions for model simulations but were not estimated from the data. For individuals that were initially infected during any simulation, the fraction of an infection duration that they had experienced prior to time $t = 0$ was given by the nuisance parameter $F_{\text{remaining}}$,

which was drawn from a uniform (0,1) distribution for each model realization. For individuals that were initially uninfected during any simulation, the probability that the individual had previously been infected at some point in time was given by p_{past}, which was drawn from a uniform (0,1) distribution for each model realization.

2.4.5 Likelihood-based inference

Inference by maximum likelihood was carried out using multiple iterated filtering (MIF) [80]. Briefly, iterated filtering is an algorithm that uses sequential Monte Carlo (SMC) to approximate maximum likelihood estimates of parameters from POMP models. SMC uses a population of particles drawn from the parameters of a given model to generate Monte Carlo samples of the latent dynamic variables and evaluate the likelihood of observed time series [38]. Iterated filtering successively filters the particle population, perturbing the parameters between iterations. The perturbations decrease in amplitude over time, allowing convergence at the maximum likelihood estimate.

The extension of SMC and iterated filtering to longitudinal panel data has been previously described [145], and we extended longitudinal POMP methods to binary data. The data for each HPV type is a set of binary time series, or panel units, describing the observed infection trajectory for each individual. The panel POMP contains a POMP model for each individual, and individuals share parameters. To evaluate the likelihood of a shared parameter set, SMC is carried out over the time series for each individual to generate a log-likelihood for the corresponding panel unit. The log likelihood of the panel POMP object is the sum of the individuals' log likelihoods. All optimization routines were carried out using 20,000 particles to overcome high Monte Carlo error.

For each model, we initialized the iterated filtering with 100 random parameter combinations. Optimization involved series of successive MIF searches, with the output of each search serving as the initial conditions for the subsequent search. The likelihood of the output for each search was calculated by averaging the likelihood from ten passes

through the particle filter, each using 50,000 particles. The optimization was repeated until additional operations did not arrive at a higher maximum likelihood.

For model selection, we used the corrected Akaike Information Criterion (AICc) [76]. We obtained maximum likelihood estimates for each parameter and associated 95% confidence intervals by constructing likelihood profiles. We used Monte Carlo Adjusted Profile methods [79] to obtain a smoothed estimate of the profile that accounts for the increased Monte Carlo error associated with longitudinal data. The lower and upper limits of the 95% confidence interval were the points that lay 1.92 log-likelihood units below the maximum likelihood estimate on the smoothed profile curve. These points correspond to one-half the 95% critical value for a χ^2 distribution with one degree of freedom.

2.4.6 Monte Carlo error in inference from binary panel data

Advances in simulation-based Monte Carlo methods have made it possible to fit complex models to large datasets. We took advantage of extensions of multiple iterated filtering (MIF) [80] to the case of panel data. Iterated filtering uses sequential Monte Carlo, also known as particle filtering, to estimate the likelihood of partially observed Markov process (POMP) models. Sequential Monte Carlo uses stochastic simulations of dynamical models to produce successive populations of weighted particles. Each particle represents a Monte Carlo sample from the probability density of the latent process, conditional on the parameters and the previous observations. As the particle population is propagated along the time series, the particles are weighted and resampled at each data point, and the likelihood of each observation is estimated as the weighted average of the particles.

Large data sets and complex models can result in non-negligible Monte Carlo error in estimated likelihoods. The structure of panel data, a collection of time series that are dynamically independent apart from shared model parameters, yields high Monte Carlo error that often makes it infeasible to calculate the likelihood with an error of less than one log likelihood unit [79]. This is important because a standard approach to calculating

95% confidence intervals relies on the observation that parameter values with log likelihood scores that are within 1.92 units of the maximum log likelihood fall within the 95% confidence interval [13]. Because high levels of Monte Carlo error can make it difficult to accurately estimate likelihoods, high Monte Carlo error rates also make it difficult to estimate 95% confidence intervals. Ionides et al. [79] show that one solution to this problem is to approximate the likelihood in the region of the maximum likelihood by fitting a quadratic to to likelihood scores from a large sample of parameter values, which can in turn be used to directly estimate the confidence bounds. Ionides et al. validated this approach with panel data [79], and here we apply it to the case of binary panel data.

Before carrying out the model fitting, we tested our approach by quantifying the effect of particle size on the likelihood for a given set of parameters. As is often the case in simulation-based approaches, the Monte Carlo error in our simulations is high enough that most particles have very low likelihoods. As a result, our likelihood estimates at first improve rapidly with increases in the number of particles (Fig. 2.12).

For particle numbers above about 5000, however, further increases in particle numbers have at most weak effects. We therefore used 20,000 particles for the iterated filtering and 50,000 particles for evaluations of the likelihood by particle filtering. We also accounted for Monte Carlo error in maximum likelihood estimation by initiating a large number of independent MIF searches (n = 100) at random parameter values for any given model. To identify the MLE for a given model, we required that three searches independently arrive within two log likelihood units of the maximum likelihood value.

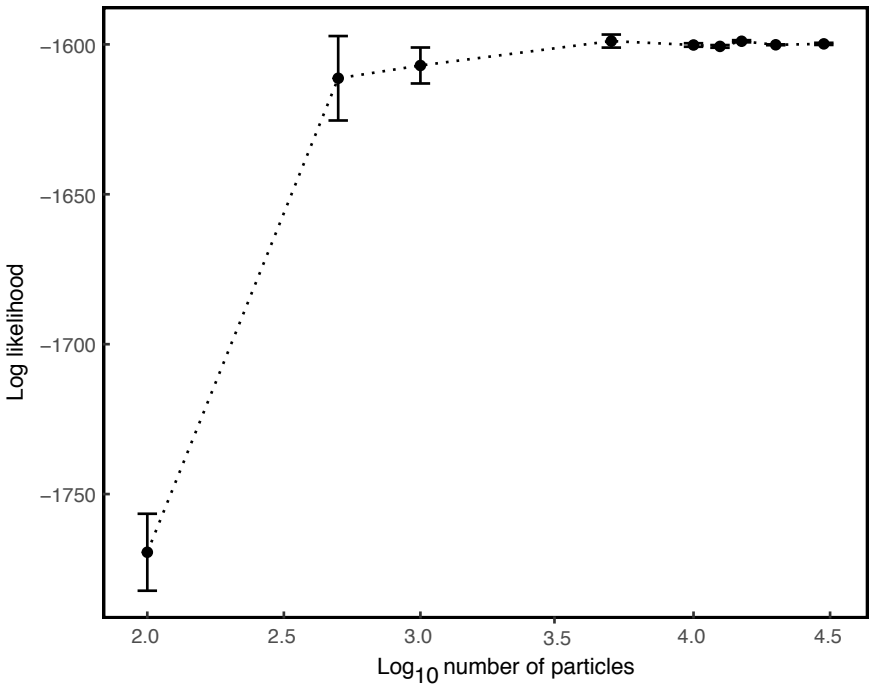

Figure 2.12: Likelihood of the maximum likelihood parameter set for HPV16 calculated at increasing particle sizes. Point estimates and error bars represent the mean and standard error, respectively, of 10 particle filter replicates.

CHAPTER 3

AGE-RELATED DIFFERENCES IN INFLUENZA IMMUNE DYNAMICS

3.1 Introduction

Circulating influenza viruses evolve rapidly in response to strong selective pressure by antibodies to the viral surface proteins hemagglutinin (HA) and neuraminidase (NA) [170, 23]. Two subtypes of influenza A viruses currently circulate in human populations, defined by antigenic differences in HA and NA. One consequence of influenza's rapid evolution is that individuals are infected multiple times over a lifetime, such that the cumulative factors that shape individuals' immune responses to circulating strains are seldom observed directly. The dynamics of protection after influenza infection are incompletely understood. The extent to which infection with one subtype reduces the risk of infection with the same or a different subtype is still unclear. It is also unclear how changes in individuals' immune responses with exposure history affect the dynamics of protection after infection. Understanding the dynamics of protection, both among individuals and within individuals over time, is important for understanding influenza evolution and population dynamics, and for vaccination strategy.

Individuals can mount different humoral and cellular immune responses to the same circulating strains. The specificity of the immune response evolves over time with repeated exposures, leading to age-related differences in protection. Infection can trigger pre-existing immune responses [51, 83] by the reactivation of B cells targeting previously encountered sites [49, 4, 88]. Furthermore, concentrations of stalk-directed antibodies increase with age [114, 120], suggesting that the immune response evolves to target more conserved regions.

The specificity of antibody repertoires may also be determined by factors independent of infection history, such as the immunogenicity of viral epitopes and even chance

[31]. Heterogeneity in T cell-mediated protection can arise from genetic factors. Cell-mediated immune responses directed against conserved proteins of influenza A viruses are believed to play an important role in both homosubtypic and heterosubtypic protection [155, 171, 14, 124, 140, 12], mainly by decreasing viral shedding rather than by preventing infection [66]. Polymorphisms human leukocyte antigen (HLA) genes correlate with differences in vaccine-induced humoral immunity [97] and in clinical outcomes after natural infection [72].

The notion that individuals vary in their immune response to circulating strains is further supported by evidence that an individual's earliest influenza infection significantly affects susceptibility to later infections. The "HA imprinting" hypothesis [61] suggests that primary infection with virus of one HA group reduces the rate of severe infection with virus of the same HA group later in life. Imprinting results in epidemiological signatures of subtype-level susceptibility by birth year, a proxy for early exposures.

Understanding the dynamics of influenza immunity requires observations of single individuals over time. While population-level influenza-like-illness (ILI) and subtype frequency data reveal important patterns of flu activity, they obscure the individual-level processes that give rise to such population trends. Infections often leave signature rises in subtype-specific antibodies that are reflected in sera. Therefore, longitudinal serology provides a basis to infer the history of individuals' infections, some of which are sub-acute [22]. The haemagglutination inhibition (HI) assay, which measures antibody against HA head, is the traditional method to measure titers against influenza, and HI titers correlate with protection [123, 74]. Serological surveillance relies on threshold values of cross-sectional titers and fold-increases in titers to indicate susceptibility and infection [141]. However, multiple studies document individual- and subtype-level variability in the titer response to influenza infections [141, 24, 49, 177], suggesting that non-specific thresholds are inappropriate in some individuals. To accurately infer the dynamics of protection from longitudinal data, it is important to understand variation in immune cor-

relates among individuals.

Mechanistic models can be used to estimate the strength and duration of homosubtypic and heterosubtypic protection. Previous models have not explicitly estimated the duration of protection after influenza infection at an individual level. Population-level transmission models coupling influenza-like-illness (ILI) time series with antigenic changes in seasonal viruses [10, 40] demonstrate the importance of waning immunity due to antigenic evolution for inter-annual epidemic dynamics. One model estimated that individuals lose immunity to circulating viruses over 2-10 years predominantly by antigenic drift [10]. Another [40] suggested that protection against a single H3N2 strain could last as long as 30 years, but that evolutionary changes every 1-2 years, predicted either by virus genotype or by antigenic cluster transitions, drive the loss of protection.

Several models of HI titer dynamics have quantified the the antibody response to influenza infection over multiple timescales, though they did not estimate the strength and duration of antibody-mediated protection. A mechanistic model fitted to cross-sectional serology suggested that the antibody response to H3N2 infection involves quickly decaying cross-reactivity against related strains, such that strains circulating 2.4 years apart were only 50% cross-reactive [95]. By reconstructing probable patterns of infection for individuals, the same model predicted that H3N2 infections occur at a median interval of 5 years. This analysis could not examine immune mechanisms acting at different (short-term and long-term) timescales within individuals. However, a complementary model incorporating longitudinal data [93] suggests that a strong, short-lived cross-reactive antibody response decays over time to a long-lived response against a narrower set of strains.

Previous models fitted to HI titers suggest that protection may differ between children and adults. Statistical analyses of paired pre- and post- epidemic sera suggest that children have a larger antibody boost than adults immediately after infection [141, 175], and that a given antibody titer is more protective in children than in adults [141]. Furthermore, there is evidence that individuals mount higher antibody titers against strains

encountered early in life than they do to subsequent infections, suggesting that adults and children may respond differently to the same circulating strain [100].

Existing models of HI titer dynamics make limiting assumptions about the immune dynamics of influenza infection. No previous models simultaneously account for the effects of pre-existing immunity, for individual heterogeneity in the infection response, and for forms of protection that are uncorrelated with HI titer. Statistical models fitted to paired pre- and post-epidemic titers [177, 141, 52, 175] or to titers over several months during multiple epidemic waves [176] focus on the period from acute infection to convalescent titer, ignoring the dynamics over longer time periods or multiple infections. These models account for heterogeneity in the infection response by age [175] or by individual-level effects [176, 52, 141, 177]. However, most ignore the effects of potential pre-existing immunity on the titer dynamics or include vague metrics of pre-existing immunity (such as propensity for high titers [141] or a binary variable for high pre-existing titer [52]) as covariates. All previous statistical and mechanistic models consider HI titer as the only determinant of protection. Furthermore, all previous models are limited by the timescale of the serological observations, ignoring the possibility of unobserved infections between data points.

To quantify the dynamics of protection after influenza infection, we fit mechanistic models of influenza dynamics to longitudinal serology. We test hypotheses about the dynamics of influenza immunity in children and adults by comparing the contributions of HI-mediated and non-HI-mediated immunity to homosubtypic protection. We introduce a new approach that tracks the dynamics of individuals over time, taking advantage of simulation-based inference methods for individual based models. Our models account for pre-existing immunity and for individual-level heterogeneity in the infection response. We find that different factors contribute to protection in children and adults, suggesting that immune repertoires change with age and repeated exposures to target more conserved viral epitopes. Homosubtypic protection wanes with a half-life of 2-4 years

across age-groups and subtypes, consistent with the findings of multiple population-level studies [10, 95]. We also quantify the short-term antibody response after PCR-confirmed infection, finding variability in the post-infection titer boost among subtypes, age-groups, and individuals. Simulations of the individual-level immune processes yield signatures of homosubtypic protection in the emergent population-level dynamics, and the simulated attack rates are consistent with population-level estimates [163].

3.2 Data

We fit models to longitudinal serology from the Kiddivax study, a household cohort study conducted in Hong Kong between 2008 and 2014. The details of the study have been described in detail elsewhere [34, 33]. Briefly, one child in each of 796 households was randomized to receive the 2009-2010 seasonal trivalent inactivated influenza vaccine (TIV) or placebo between August 2009 and February 2010. The study followed households over five years post-enrollment, collecting serology approximately every six months and respiratory specimens during reported illnesses. We use two overlapping subsets of data from the Hong Kong Kiddivax study for two separate analyses. The first analysis quantifies the short-term titer response after PCR-confirmed infection. The second, our main analysis, quantifies the dynamics of homosubtypic or heterosubtypic protection, and for a separate analysis.

For our main analysis, we fit models to HI titers from the $n = 706$ individuals (including 114 children $\leq$15 y old) from the study that were not vaccinated as part of the study and reported no vaccination at any season during followup. We classify adults as individuals $>$15 yo. Individuals had a median of 12 clinic visits spaced at a median of 6.4 months over a median 4.1 years of followup.

We fit a separate sub-model of the short-term post-infection HI titer dynamics to data from individuals with PCR-confirmed infection. We estimate the magnitude and variability of the titer boost following infection and the dependence of the titer boost on the

pre-infection titer. We fit sub-models to data from $n = 50$ individuals (including $n = 29$ children $\leq$15 y old) with PCR-confirmed H3N2 infection and $n = 78$ individuals (including $n = 42$ children $\leq$15 y old) with PCR-confirmed pH1N1 infection. For H3N2, the median time between the pre-infection titer measurement and the PCR-positive swab was 5.3 months, and the median time between the PCR-positive swab and the post-infection titer measurement was 2.6 months. For pH1N1, the median time between the pre-infection titer measurement and the PCR-positive swab was 2.4 months, and the median time between the PCR-positive swab and the post-infection titer measurement was 6.6 months. Using the raw data, we performed a linear regression of the post-infection titer on the pre-infection titer, and we excluded one individual with $\Delta t > 1$y between the pre- and post-infection titer measurements.

3.3 Modeling approach

Our models of influenza immune dynamics track individuals' HI titers and susceptibility to infection over time (Fig. 3.1, Model Overview). We assume that, after infection at time $t^{\mathrm{X}}_{\mathrm{i},s}$, the titer $h_{i,s}(t)$ undergoes an acute boost, which wanes over one year, potentially leaving a long-term boost that does not wane. We assume that protection after infection may be HI-mediated (such that susceptibility is a function of the HI titer), non-HI mediated (such that protection wanes at a constant rate after infection), or a combination of the two.

We first infer the short term-titer dynamics after infection (Fig. 3.1, Step 1a) using a sub-model fitted to the pre- and post- infection sera from individuals with PCR-confirmed infection. We then fit models to the full longitudinal data to test hypotheses about the dynamics of protection (Fig. 3.1, Step 1b), fixing the parameters that govern the short-term titer dynamics. The "Complete model description" details the full model.

From the maximum likelihood parameter estimates for the best-fit models, we simulate individuals' latent infections, titers, and susceptibility (Fig. 3.1, Step 2). We estimate

population-level patterns of infection and protection that arise from the individual dynamics over the study period (3.1, Step 2a). We also evaluate the ability of our models to reproduce empirical patterns in the longitudinal titers (Fig. 3.1, Step 2b).

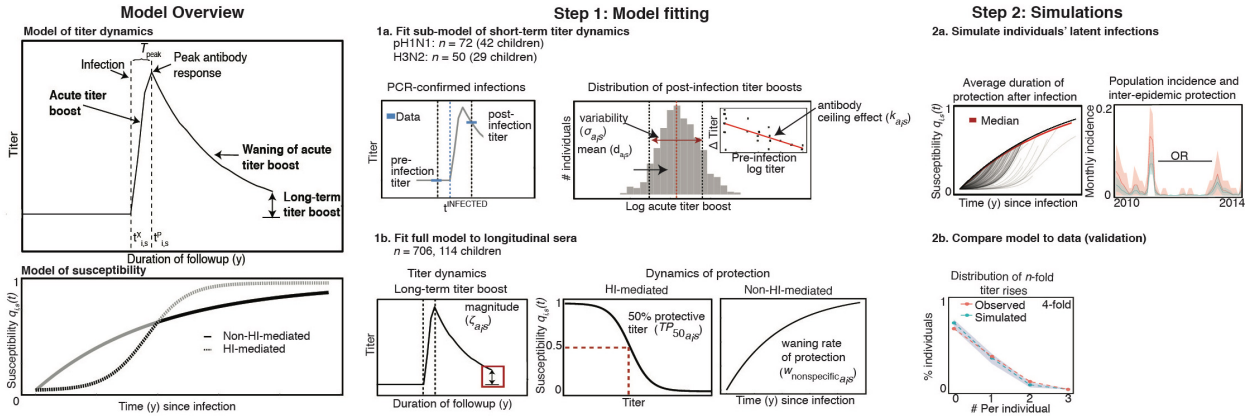

Figure 3.1: Schematic of modeling approach. **Model Overview**: Schematic of HI titer for individual i against subtype s after infection at time $t^{X}_{i,s}$ (top) and schematic of susceptibility $q_{i,s}(t)$ after infection under non-HI-mediated and HI-mediated protection (bottom). The solid and dashed lines in the bottom figure show the dynamics of susceptibility under each form of protection in isolation. The black trajectory shows the dynamics under both forms in combination, where the susceptibility at any time is the minimum value predicted by either form. **Step 1**: (1a) Inference approach for the sub-model of the short-term post-infection titer dynamics. The data for each individual are the time of PCR-confirmed infection and the closest pre- and post-infection titer measurements. (1b) Inference approach for the full model of the immune dynamics, fitted to the full longitudinal serology. **Step 2**: (2a) Simulated population-level dynamics. From the latent infections and susceptibility for each individual, we track the loss of protection after infection in the population. We also estimate the incidence and the odds ratios (OR) of protection between epidemics. (2b) Model validation. We compare the simulated and observed distributions of n-fold titer rises among individuals.

3.4 Results

3.4.1 Homosubtypic protection arises from a combination of HI-mediated immunity and non-HI-mediated immunity in adults, while protection is dominated by HI-mediated immunity in children.

Homosubtypic protection may arise due to mechanisms undetectable by HI, such as cellular immunity, anti-HA antibodies directed at sites away from the receptor binding domain, or anti-NA antibodies [114, 120, 26], and the importance of these unobserved mechanisms may change with repeated exposures or age. We therefore evaluate three hypotheses about the factors that underlie homosubtypic protection in children and adults, reflected in individuals' latent susceptibiliy $q_{i,s}(t)$:

1. Protection consists of a combination of HI-mediated and non-HI-mediated factors

$$q_{i,s}(t) = \min\left\{q_{1_{i,s}}(t), q_{2_{i,s}}(t)\right\}, \tag{3.1}$$

   with $q_{1_{i,s}}(t)$ and $q_{2_{i,s}}(t)$ given by Eqs. 3.7 and 3.8, respectively. Here, the susceptibility at any time is the minimum predicted by either form of protection (Fig. 3.1, Model Overview).

2. Protection arises only from non-HI-mediated factors

$$q_{i,s}(t) = q_{2_{i,s}}(t) \tag{3.2}$$

   such that the degree of protection depends on the rate $w_{\text{nonspecific}_{a_i,s}}$ of immune waning (Fig. 3.1 Step 1b).

3. Protection arises only from HI-mediated factors

$$q_{i,s}(t) = q_{1_{i,s}}(t) \tag{3.3}$$

such that the degree of protection depends on the 50% protective titer $TP_{50_{a_i,s}}$ (Fig. 3.1, Step 1b).

For both subtypes, the best-fit model in children includes only HI-mediated protection (Table 3.1). In adults, there is strong support for both models that include a contribution of non-HI-mediated protection. This result suggests that early in life, protection against influenza infection is dominated by immune responses that correlate well with HI titer, namely anti-HA-head-directed antibodies. However, over time, other processes become important, such that HI titer alone is not the best predictor. This conceptual model is consistent with the observation that children in this study have higher baseline titers than adults for both subtypes (Fig. 3.8), where baseline titers are approximated by the lowest observed titer over follow-up. Children have both a larger fraction of detectable baseline titers and a higher mean baseline titer for both subtypes, signaling greater pre-existing HI-mediated immunity.

Subtype		**Model (*n* parameters)**	**Log likelihood (se)**	**ΔAICc**
H3N2	Adults	Combined (4)	-5217.9 (0.7)	0
		Non-HI-mediated (3)	-5218.3 (0.8)	1.2
		HI-mediated (3)	-5242.2 (0.4)	46.6
	Children	Combined (4)	-1511.8 (0.6)	2.5
		Non-HI-mediated (3)	-1524.0 (0.5)	24.8
		HI-mediated (3)	-1511.6 (0.9)	0
pH1N1	Adults	Combined (4)	-4492.1 (0.7)	0
		Non-HI-mediated (3)	-4493.4 (0.9)	0.6
		HI-mediated (3)	-4520.7 (0.2)	55.1
	Children	Combined (4)	-1370.0 (0.5)	2.0
		Non-HI-mediated (3)	-1381.8 (0.8)	23.8
		HI-mediated (3)	-1370.1 (0.4)	0

Table 3.1: Model comparisons for the complete model. Under the "combined" model, protection arises from a combination of HI-mediated and non-HI-mediated mechanisms.

3.4.2 *Natural infection generates age-specific homosubtypic protection that declines within several years.*

Using the best-fit models for each subtype in adults (Table 3.2), we quantified the dynamics of homosubtypic protection. Infection in adults results in non-HI-mediated immunity

Subtype	Parameter		MLE [95% CI]
H3N2	Long-term boost		
		$\zeta_{\text{adults,s}}$	0 [0, 0.001]
		$\zeta_{\text{kids,s}}$	0.02 [0, 0.04]
	50% protective titer		
		$\text{TP}_{50_{\text{kids},s}}$	1:60 [1:42, 1:122]
		$\text{TP}_{50_{\text{adults},s}}$	1:330 [1:110, 1:5120]
	Half-life		
	non-HI-mediated	from $w_{\text{nonspecific}_{\text{adults},s}}$	2.1 y [1.3, 3.3]
	immunity	$w_{\text{nonspecific}_{\text{kids},s}}$	N/A
pH1N1	Long-term boost		
		$\zeta_{\text{adults,s}}$	0.01 [0, 0.03]
		$\zeta_{\text{kids,s}}$	0.2 [0.1, 0.4]
	50% protective titer		
		$\text{TP}_{50_{\text{adults},s}}$	1:540 [1:90, 1:5120]
		$\text{TP}_{50_{\text{kids},s}}$	1:30 [1:10, 1:45]
	Half-life		
	non-HI-mediated	from $w_{\text{nonspecific}_{\text{adults},s}}$	3.4 y [2.6, 4.7]
	immunity	$w_{\text{nonspecific}_{\text{kids},s}}$	N/A

Table 3.2: Maximum likelihood estimates and associated uncertainty.

that wanes with a half-life of 3.4 y [2.6, 4.7] for pH1N1 and 2.1 y [1.3, 3.3] for H3N2. HI-mediated immunity makes only a weak contribution to protection against both pH1N1 and H3N2, indicated by the high $\text{TP}_{50_{\text{adults}}}$ for both subtypes. Notably, the $\text{TP}_{50_{\text{adults}}}$ for the best-fit models in adults is not identifiable from the rate of non-specific immune waning (Fig. 3.5). When we force the models to exclude non-HI-mediated immunity, we estimate a lower $\text{TP}_{50_{\text{adults}}}$ for both pH1N1 ($\text{TP}_{50_{\text{adults}}}$ = 1:15 [1:9, 1:25], Fig. 3.7) and H3N2 ($\text{TP}_{50_{\text{adults}}}$ = 1:45 [1:37,1:120], Fig. 3.7). Infection in adults does not result in a long-term titer boost in either pH1N1 or H3N2 (the 95% confidence interval for $\zeta_{\text{adults},s}$ includes 0

for both subtypes).

Given our estimates of the parameters that govern the individual-level dynamics, we simulated the models to estimate the duration of protection among adults. Using 1,000 simulations from the MLEs of the best-fit models for H3N2 and pH1N1, we track the latent susceptibility in adults following an initial infection over followup (Fig. 3.2). For both subtypes, the waning rate $w_{\text{nonspecific}_{\text{adults},s}}$ of non-HI-mediated immunity (Eq. 3.8) determines the loss of protection such that the population regains 50% susceptibility in roughly 3.5 y and 2 y for pH1N1 and H3N2, respectively. Therefore, the individual trajectories in Fig. 3.2 are constrained from above by the curve describing non-HI-mediated protection (Fig. 3.1, Step 1b).

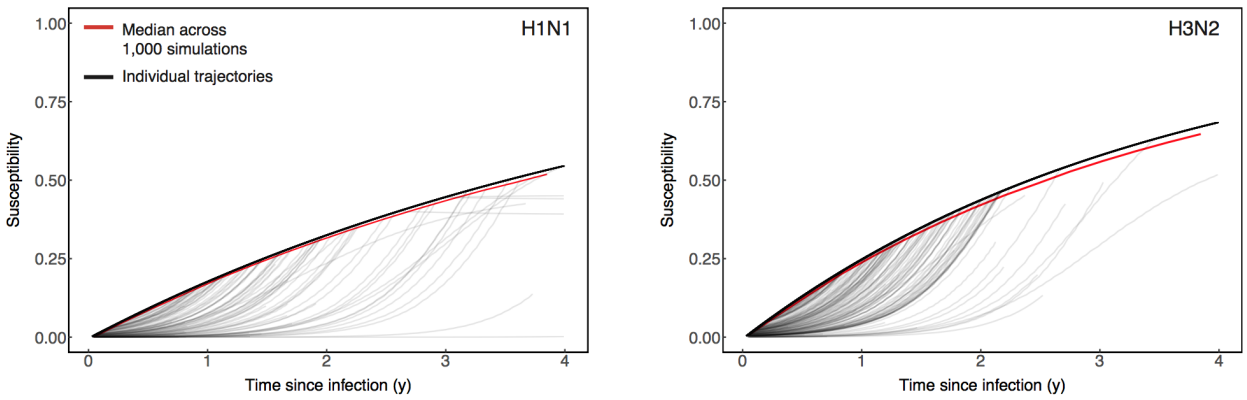

Figure 3.2: Susceptibility in adults following simulated infection (at time $t = 0$) with pH1N1 and H3N2. The black lines represent individual trajectories from one simulation, and the red line represents the median among individuals over 1,000 simulations.

Compared to adults, children have a more variable duration of protection. Because susceptibility in children depends only on HI titer, the dynamics of protection are sensitive to variability in the post-infection titer boost. For both pH1N1 and H3N2, we estimate substantial variation in the short-term titer dynamics after PCR-confirmed infection (section 3.8.1.1). The variability arises both from stochastic variation in the magnitude of short-term titer boost and from the antibody ceiling effect (Table 3.8), in which higher pre-infection titers diminish the realized boost. Therefore, while the median susceptibil-

ity among children reaches 50% approximately 4 years after infection with pH1N1 and over roughly 2 years for H3N2 (Fig. 3.3), the individual trajectories differ according to variation in the pre-infection titers and in the extent of short-term boosting. The shape of the individual trajectories reflects the logistic relationship between titer and susceptibility under HI-mediated protection (Eq. 3.7 and Fig. 3.1, Step 1b). Infection with pH1N1 generates a long-term boost in titer that is 20% the value of the acute boost ($\zeta_{\text{kids,s}}$ = 0.2 [0.1, 0.4], Table 3.2), such that individuals gain long-term protection as their baseline titer is boosted above the TP_{50} through repeated exposures. In H3N2, by contrast, we estimate only a small long-term boost ($\zeta_{\text{kids,s}}$ = 0.01 [0, 0.03], Table 3.2).

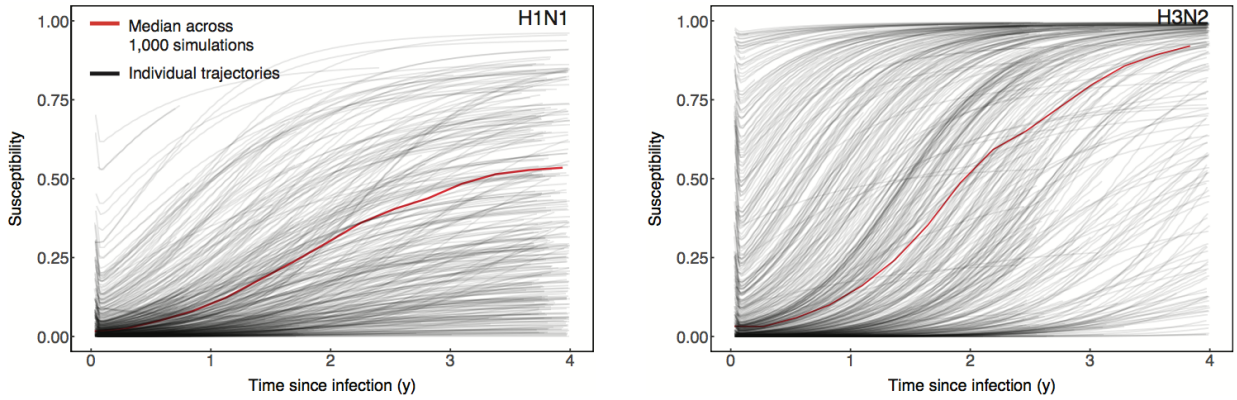

Figure 3.3: Susceptibility in children following simulated infection (at time $t = 0$) with pH1N1 and H3N2. The black lines represent individual trajectories from one simulation, and the red line represents the median among individuals over 1,000 simulations.

3.4.3 The models reproduce population-level patterns of infection and homosubtypic protection

Our models recover population-level patterns of infection during the study period for both subtypes. From the simulated latent infections for H3N2 and pH1N1, we infer the monthly incidence in children and adults (Fig. 3.4) and the cumulative incidence over the epidemics that occurred during the study period (Table 3.5). Our models predict slightly higher incidence in children than adults for both subtypes, and incidence closely tracks

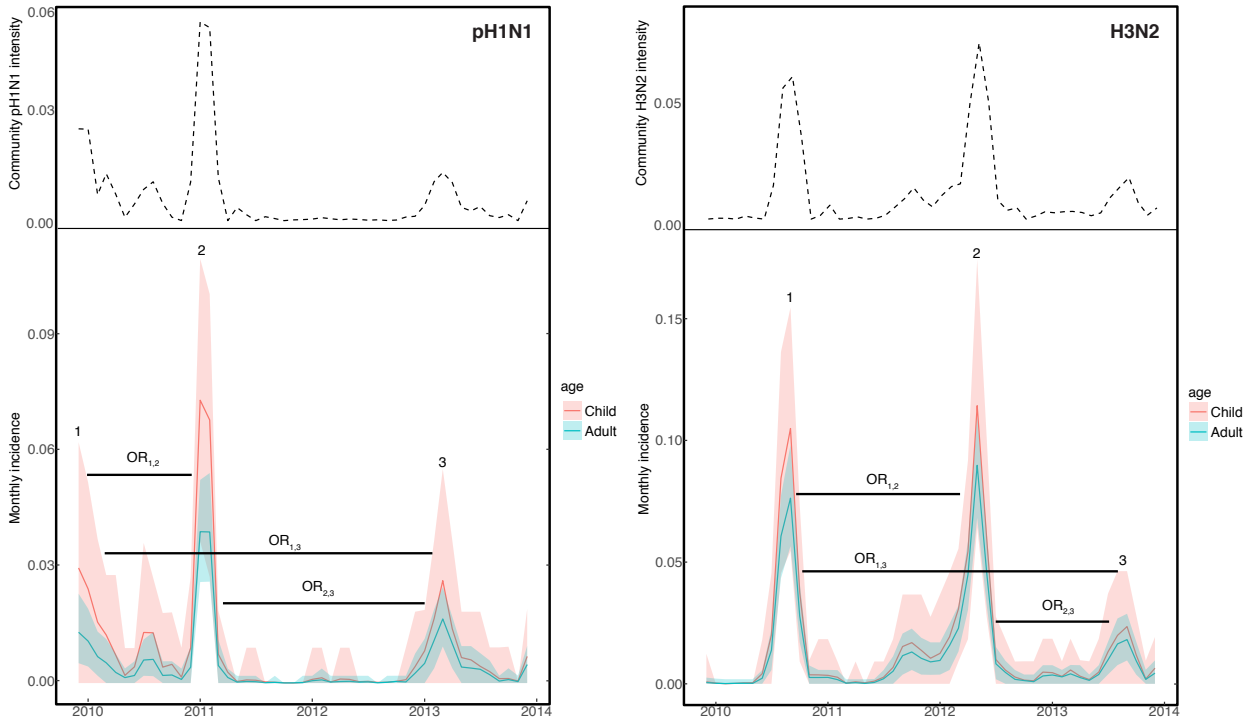

Figure 3.4: Simulated monthly incidence for pH1N1 (left) and H3N2 (right) in children and adults, averaged over 1,000 simulations. The shaded areas are bounded by the 2.5% and 97.5% quantiles from the simulations. The black dashed line represents the monthly community intensity of H3N2. Horizontal black bars denote inter-epidemic periods for odds ratios (OR).

the monthly subtype-specific intensity of influenza illness estimated by a Hong Kong sentinel surveillance network. The range of inferred cumulative incidences over the observed pH1N1 epidemics is 4-8% in adults and 6-17% in children. For H3N2, the epidemic incidences range from 5-17% in adults and from 5-24% in children. Our estimates are also within range of multiple estimates of seasonal influenza incidence in the United States [158, 163]. Note that in our models, the community influenza intensity drives infection risk (Eq. 3.6) and thus affects the timing but not the absolute magnitude of the latent epidemics. The incidence itself is an emergent property of the dynamics, dependent on the inferred subtype-specific scaled transmission rate $\beta_s(t)$ and on the inferred dynamics of protection.

The simulated infections reproduce signatures of homosubtypic protection. We es-

timate the odds ratios of protection between epidemics (Table 3.3). We find evidence of inter-epidemic homosubtypic protection for children with pH1N1 between 2009 and 2011, consistent with a previous analysis of this trial [33]. There is also homosubtypic pH1N1 protection in adults for two inter-epidemic periods, between 2009 and 2011 and between 2011 and 2013. We find homosubtypic H3N2 protection for both children and adults during two inter-epidemic periods, between 2010 and 2012 and between 2012 and 2013.

Subtype			OR [95% CI]	Estimate from [33]
pH1N1	Adults	$OR_{1,2}$	0.4 [0.3, 0.5]	
		$OR_{2,3}$	0.6 [0.5, 0.7]	
		$OR_{1,3}$	0.8 [0.7, 1.0]	
	Children	$OR_{1,2}$	0.6 [0.5, 0.8]	0.27 [0.10, 0.76]
		$OR_{2,3}$	0.8 [0.7, 1.1]	
		$OR_{1,3}$	0.9 [0.8, 1.2]	
H3N2	Adults	$OR_{1,2}$	0.5 [0.4, 0.6]	
		$OR_{2,3}$	0.6 [0.4, 0.6]	
		$OR_{1,3}$	0.7 [0.7, 1.0]	
	Children	$OR_{1,2}$	0.8 [0.7, 0.8]	0.39 [0.18, 0.83]
		$OR_{2,3}$	0.7 [0.7, 0.9]	
		$OR_{1,3}$	0.9 [0.9,1.1]	

Table 3.3: Inter-epidemic odds ratios of infection, predicted from 1,000 simulations of the models for pH1N1 and H3N2 at the MLEs.

3.4.4 Neither group-level HA imprinting nor heterosubtypic immunity reduce subtype-level susceptibility.

The dynamics of protection against circulating influenza viruses may vary among individuals based on differences in primary influenza exposures and on previous infection with heterologous subtypes. We therefore estimate the effects of group-level HA imprinting and heterosubtypic infection on susceptibility.

We estimate the effect α_{imp_s} of primary infection with a subtype of one HA group (or

group-level HA imprinting) on the risk of infection with subtype s by

$$\lambda_{\text{imp}_{i,s}}(t) = \lambda_{i,s}(t)(\alpha_{\text{imp}_s})(p_{\text{imp}_{i,s}}) + \lambda_{i,s}(t)(1 - p_{\text{imp}_{i,s}}). \tag{3.4}$$

We calculate the probability $p_{\text{imp}_{i,s}}$ that an individual's first influenza exposure was to subtype s based on the individual's birthdate, the current date, and historical incidence data (Fig. 3.10A). Birth-year effects are confounded by age-specific differences in risk. However, data from individuals of similar age but different imprinting subtypes contain information about imprinting effects. We therefore fit the imprinting models for pH1N1 and H3N2 to data from middle-aged adults (35-50 yo), where we observe a transition between Group 1 (H2N2) and Group 2 (H3N2) imprinting viruses (Fig. 3.10B). A previous analysis [54] We estimate the effect of homosubtypic imprinting for H3N2 and we estimate the group-level effect for pH1N1 using H2N2 as the imprinting subtype (Table 3.7). The likelihood profiles for the imprinting effect (Fig. 3.10C) are centered around one (no effect) for both subtypes.

To estimate the duration of heterosubtypic protection, we fit a two-strain model of the pH1N1 and H3N2 dynamics. We fix the model parameters that govern homosubtypic immunity for each subtype according to the MLEs of the best-fit single strain models (Table 3.2). Let $q_{\text{homosubtypic}_{i,s}}$ denote the susceptibility to subtype s determined only by homosubtypic protection. Heterosubtypic protection following infection with subtype m contributes to the susceptibility against subtype s such that

$$q_{i,s}(t) = \min\left\{q_{\text{heterosubtypic}_{i,s}}(t), q_{\text{homosubtypic}_{i,s}}(t)\right\}. \tag{3.5}$$

We estimate the rate of waning of heterosubtypic protection, $w_{\text{nonspecific}_{a_i,m}}$ (Eq. 3.15), assuming that the rate is symmetric between subtypes. The likelihood profile for $w_{\text{nonspecific}_{a_i,m}}$ supports an extremely high rate of heterosubtypic immune waning (Fig. 3.11), such that our estimate of the half life of heterosubtypic protection includes zero (MLE 0.006y, 95%

CI [0.0,0.01]).

3.5 Model validation and sensitivity analysis

Simulations from our models at the maximum likelihood parameter estimates reproduce the observed distribution of fold titer rises in the data (section 3.8.2.1, Figs. 3.16, 3.15). Our inference results are robust to assumptions about the initial conditions and to alternate scaling of the community flu intensity during the 2009 pH1N1 pandemic (section 3.8.2.3). Furthermore, our assumptions about the measurement error are consistent with the error from estimated from replicate titer measurements in the data (section 3.8.2.4).

3.6 Discussion

Host immunity partially regulates the diversity of influenza viruses [46, 91]. However, the dynamics of protection and the contribution of individual variability in the immune response are poorly understood. Here we quantify the age-specific dynamics of homosubtypic protection. We find that HI-mediated immunity alone correlates with protection in children, while non-HI-mediated immunity correlates with protection in adults. Our results therefore suggest that circulating influenza strains face different selective pressures in children and adults. Broadly, our results underscore the need for a more precise understanding of the factors that determine the variable response to infection among individuals, and for better correlates of immune protection in adults.

We find that different immune factors correlate with homosubtypic protection in children and adults. HI-mediated immunity alone best explains the dynamics in children, suggesting that protective responses in childhood are directed against epitopes on the HA head. Non-HI-mediated immunity correlates with the dynamics of protection in adults, suggesting that the antibody repertoire evolves with age and/or repeated exposures to focus on more conserved epitopes. This model is consistent with epidemiological evidence

that levels of stalk-directed antibodies correlate positively with age [133]. Protection in both adults and children may also arise from responses not specific to HA, such as antibody and T cell responses to NA, and T cells to conserved internal proteins [45, 92]. However, our model suggests that mechanisms that are uncorrelated with responses directed against HA play a minor role in protection for children. Our work supports age- and subtype- level differences in the protective HI titer (TP_{50}), consistent with previous literature [175, 9]. In children, our estimates of the TP_{50} for both subtypes are in line with the 1:20 to 1:80 range identified previously (Table 3.2) [175, 123, 74]. In adults, by contrast, we cannot estimate the $TP_{50_{adults}}$ independently of the rate of non-specific immune waning $w_{\text{nonspecific}_{\text{adults}}}$ (Table 3.2, Fig. 3.5). We estimate a lower TP_{50} in adults for both subtypes when we force the model to exclude non-HI-mediated protection (Fig. 3.7). However, model selection indicates that non-HI-mediated immunity is important for the dynamics of protection (Table 3.1). The finding that different factors determine homosubtypic protection in children and adults results implies that the antigenic phenotype of any given strain varies among individuals. Our results therefore imply that infection in children and adults generates different population-level selective pressures for influenza strains.

We estimate that homosubtypic protection in both children and adults for both subtypes wanes with an average half life of 2-4 years. Notably, this timescale is consistent with the 2.4 y half-life of cross-reactive immunity that was previously estimated for H3N2 [95], and with the estimated decay of immunity over 2-10 years due to antigenic evolution [10]. While protection wanes consistently in adults, the duration of protection varies substantially among children (Figs. 3.3, 3.2). If individuals respond differently to the same virus, then correlates of protection might also vary among individuals, making correlates challenging to identify. Variable mechanisms of host protection against a circulating pathogen imply that some induced immune responses have no perceivable effect on the pathogen growth rate within certain individuals [30]. This could explain the fact that,

while anti-HA head antibody responses best explain protection in children, including long-lived immunity in children to pH1N1, the same responses cannot explain protection in adults.

We can connect the individual dynamics to population trends. Consistent with evidence that individuals' antibody repertoires are shaped by early exposures to particular HA variants [48, 100], our results suggest that protection against circulating strains in younger individuals correlates with antibodies against the HA head. However, antigenic evolution may diminish the protection afforded by responses that target earlier strains over time. Population-level transmission models suggest that this process partly drives epidemic dynamics [40]. Given the loss of protection from HA head-directed antibody responses targeting earlier strains, protective immune responses in older individuals may target epitopes away from the HA head.

We estimate substantial individual variation in the short-term titer boost after infection for both pH1N1 and H3N2, explained partially by differing pre-infection levels of HA head antibodies. The antibody ceiling effect that we detect in children and adults for both subtypes (Tables 3.6, 3.8) has been previously described both for antibodies against the HA head following vaccination [137] and, recently, for antibodies against the HA stalk following viral challenge [133]. Both the ceiling effect and the nonspecific variability in the magnitude of the short-term boost (Table 3.6) suggest that threshold correlates of infection (indicated by fold-increases in titer) may not be predictive for all individuals.

We find no evidence that HA imprinting or heterosubtypic immunity mediate susceptibility to circulating subtypes. Two epidemiological studies support birth-year effects consistent with HA imprinting [61, 54]. Previous analyses of HA group-level imprinting have suggested that imprinting reduces the rate of severe infection, though it may not prevent infection altogether [61]. As serological testing in this study did not require symptomatic infection, our data may not be able to detect such an effect. As with imprinting, heterosubtypic immunity may act to reduce the severity of illness but not to prevent

infection, consistent with hypotheses that cross-reactive T cells decrease viral shedding [66, 92]. Coupling data on disease severity, viral shedding, and serology could elucidate the existence of both imprinting and heterosubtypic immunity. Additionally, the discordance of pH1N1 and H3N2 epidemic peaks (Fig. 3.4, 3.19) over the study period may prevent us from estimating within-season effects of heterosubtypic protection.

This work has multiple limitations. First, though our models support substantial variability in the short-term titer boost following infection, our data lack multiple PCR-confirmed infections in single individuals. Therefore, we cannot distinguish the non-specific variability at each infection ($\sigma_{a_i,s}$) from consistent differences between individuals. Second, our models assume a well-mixed population with age-specific contact rates, ignoring age-assortative mixing that might alter the exposure risk among individuals. Third, though our results provide insight into differences between children and adults, we cannot model the evolving response in individuals over a lifetime, including in infants and in the elderly. Therefore, we cannot estimate how age-related phenomena such as maternal antibody protection and immunosenescence affect the response to infection.

3.7 Methods

3.7.1 Complete model description

1. *Exposure to infection*

   Individuals' risk of exposure is based on current subtype-specific influenza activity and age-specific contact rates [50, 175]. For individual i, the risk of infection with subtype s, $\lambda_{i,s}(t)$, depends on the community-level risk and the individual's susceptibility,

$$\lambda_{i,s}(t) = q_{i,s}(t)\beta_{c_{a_i}}\beta_{\text{scaled}_s}L_s(t) \tag{3.6}$$

where $q_{i,s}(t)$ is the subtype- and individual-specific susceptibility (or per-infectious-contact probability of infection), $\beta_{c_{a_i}}$ is the contact rate for age class a_i, and $L_s(t)$ is a proxy for the flu activity for subtype s. The parameter β_{scaled_s} scales the flu intensity to determine the per-infectious-contact transmission rate at time t. We estimate β_{scaled_s} for each subtype by fitting the model to the combined data for children and adults, and we then fix β_{scaled_s} to fit models of the protective dynamics separately in children adults. We calculate $L_s(t)$ from weekly GP data as (ILI/total GP consultations)(% positive). We impose a minimum threshold $L_{s_{\min}}(t) = 10^{-5}$.

2. *Susceptibility to infection based on HI titer to the infecting strain, non-HI-mediated protection, or both*

   We model susceptibility to infection with subtype s as two separate processes, one that depends on the HI titer against the infected strain (the HI-mediated component), and another that accounts for protection by other unobserved processes (the non-HI-mediated-component).

   The HI-mediated component of susceptibility $q_{1_{i,s}}(t)$ is a logistic function of the HI titer [32, 41], demonstrated in Fig. 3.1, Step 1b. Because previous studies suggest that the dependence of susceptibility on titer is age specific [9], we estimate the relationship separately for children and adults. The susceptibility of individual i to subtype s at time t , $q_{1_{i,s}}(t)$, is given by the logistic function

$$q_{1_{i,s}}(t) = 1 - \frac{1}{1 + e^{\phi_{a_i}(\log(h_{i,s}(t)) - \log(\text{TP}_{50_{a_i,s}}))}}, \tag{3.7}$$

   where $h_{i,s}(t)$ is the latent titer and ϕ_{a_i} and $\text{TP}_{50_{a_i,s}}$ are age-dependent scaling constants.

   The non-HI-mediated component of susceptibility $q_{2_{i,s}}(t)$ assumes complete protec-

tion that wanes at a constant rate after infection

$$q_{2_{i,s}}(t) = 1 - e^{-w_{\text{nonspecific}_{a_i,s}}(t - t^{\text{X}}_{\text{i,s}})} \tag{3.8}$$

where $w_{\text{nonspecific}_{a_i,s}}$ is the rate of waning protection, fitted separately for children and adults. Figure 3.1, Step 1b shows the

The susceptibility $q_{i,s}(t)$ can then be modeled by either component or by a combination of the two.

3. *Boost and waning following infection*

   When individual i is infected with subtype s, antibody titers begin increasing from the time of infection $t^{\text{X}}_{i,s}$ and eventually peak. The acute boost occurs occurs according to f_{rise}

$$f_{\text{rise}}(h_{i,s}(t), t^{\text{X}}_{i,s}, t) = h_{i,s}(t^{\text{X}}_{\text{i,s}})^{(1-k_{a_i,s})} d_{i,s}(t^{\text{X}}_{\text{i,s}}) \left(1 - e^{-r(t - t^{\text{X}}_{\text{i,s}})}\right), \tag{3.9}$$

   where $t^{\text{X}}_{i,s}$ and $h_{i,s}(t^{\text{X}}_{i,s})$ give the time and titer, respectively, at the most recent infection, r gives the rate of titer rise following infection, $d_{i,s}(t^{\text{X}}_{i,s})$ is the magnitude of the short-term boost. The age- and subtype-specific parameter $k_{a_{i,s}}$ governs the dependence of the titer boost on the pre-infection titer, allowing for an antibody ceiling effect [81] whereby individuals with high pre-infection titers experience diminished short-term boosting.

   We estimate the parameters $k_{a_{i,s}}$, $d_{a_{i,s}}$, and $\sigma_{a_i,s}$ that describe the short-term titer dynamics following infection separately from a sub-model fit to data from individuals with PCR-confirmed infection (Section 3.8.1.2). We then fix the values of these parameters in the main model. Multiple studies demonstrate heterogeneity in the short-term titer rise following infection [52, 49]. Therefore, we allow for variability

in the magnitude of the short-term boost for each infection such that

$$log(d_{i,s}(t^{\mathrm{X}}_{\mathrm{i,s}})) \sim \mathcal{N}(d_{a_i,s}, \sigma_{a_i,s}) \tag{3.10}$$

where $d_{a_{i,s}}$ and $\sigma_{a_i,s}$ give the age- and subtype-specific log mean and standard deviation, respectively of the boost magnitude.

After antibodies peak at time $t^{\mathrm{P}}_{\mathrm{i,s}}$, the titer wanes exponentially at rate w to an individual's subtype-specific baseline titer $h_{\mathrm{baseline}_{i,s}}(t)$. Therefore, the titer after the peak short-term response is given by

$$f_{\mathrm{wane}}(h_{i,s}(t^{\mathrm{P}}_{\mathrm{i,s}}), t^{\mathrm{P}}_{\mathrm{i,s}}, t) = \left(h_{i,s}(t^{\mathrm{P}}_{\mathrm{i,s}}) - h_{\mathrm{baseline}_{i,s}}(t)\right) e^{-w(t-t^{\mathrm{P}}_{\mathrm{i,s}})}. \tag{3.11}$$

Infection may cause a long-term boost $d_{\mathrm{longterm}_{i,s}}(t^{\mathrm{X}}_{\mathrm{i,s}})$ that does not wane, where $d_{\mathrm{longterm}_{i,s}}(t^{\mathrm{X}}_{\mathrm{i,s}})$ is defined as a fraction $\zeta_{a_i,s}$ of the acute boost

$$d_{\mathrm{longterm}_{i,s}}(t^{\mathrm{X}}_{\mathrm{i,s}}) = \zeta_{a_i,s} d_{i,s}(t^{\mathrm{X}}_{\mathrm{i,s}})). \tag{3.12}$$

The long-term boost updates the baseline titer following each infection at time $t^{\mathrm{X}}_{\mathrm{i,s}}$ such that

$$h_{\mathrm{baseline}_{i,s}}(t) = h_{\mathrm{baseline}_{i,s}}(t^{\mathrm{X}}_{\mathrm{i,s}}) + d_{\mathrm{longterm}_{i,s}}(t^{\mathrm{X}}_{\mathrm{i,s}}). \tag{3.13}$$

Let T_{peak} denote the (fixed) length of time between infection and peak titer level. The complete expression for $h_{i,s}(t)$ is then

$$\begin{cases} h_{i,s}(t) = h_{i,s}(t^{\mathrm{X}}_{\mathrm{i,s}}) + f_{\mathrm{rise}}(h_{i,s}(t^{\mathrm{X}}_{\mathrm{i,s}}), t^{\mathrm{X}}_{\mathrm{i,s}}, t), & \text{for } t - t^{\mathrm{X}}_{\mathrm{i,s}} < T_{\mathrm{peak}}, \\ h_{i,s}(t) = h_{i,s}(t^{\mathrm{X}}_{\mathrm{i,s}}) + f_{\mathrm{wane}}(h_{i,s}(t^{\mathrm{P}}_{\mathrm{i,s}}), t^{\mathrm{P}}_{\mathrm{i,s}}, t), & \text{for } t - t^{\mathrm{X}}_{\mathrm{i,s}} \geq T_{\mathrm{peak}}. \end{cases} \tag{3.14}$$

4. *Heterosubtypic immunity*

Heterosubtypic immunity acts as a non-specific form of protection against subtype s following infection with subtype m at time $t^{\mathrm{X}}_{\mathrm{i},m}$ and wanes at rate $w_{\mathrm{nonspecific}_{a_i,m}}$

$$q_{\mathrm{hetero}_{i,s}}(t) = 1 - e^{-w_{\mathrm{nonspecific}_m}(t - t^{\mathrm{X}}_{i,m})}. \tag{3.15}$$

3.7.2 *Initial conditions*

We assign each individual's initial latent subtype-specific baseline titer, $h_{\mathrm{baseline}_{i,s}}(0)$ based on the lowest observed titer, $h_{\mathrm{obs}_{\mathrm{min}_{i,s}}}$. Because an observed HI titer represents the lower bound of a two-fold dilution, we draw $h_{\mathrm{baseline}_{i,s}}(0)$ for each realization of the model according to

$$h_{\mathrm{baseline}_{i,s}}(0) \sim U(h_{\mathrm{obs}_{\mathrm{min}_{i,s}}}, 2h_{\mathrm{obs}_{\mathrm{min}_{i,s}}}). \tag{3.16}$$

The values of the initial latent titer $h_{i,s}(0)$ and the initial susceptibility $q_{i,s}(0)$ depend on the time of most recent infection, which may have occurred before entry in the study. To initialize the latent states for each individual, we draw the time of the most recent infection from the density of subtype-specific flu intensity $L_s(t)$ in the seven years before the first observation. In this way, we account for known epidemic activity in Hong Kong before the beginning of the study (Fig. 3.6). For children less than 7 y old, the distribution is truncated at birth, and the density includes the probability that the child is naive to influenza infection.

Measurement model and likelihood function

The measurement model accounts for error in the titer measurements as well as the effect of discretization of titer data into fold-dilutions. The observed titer values are fold-dilutions in the range [<1:10, 1:10, 1:20 ... , 1:5120]. Consistent with other models

[176, 175, 95], we define a log titer ($logH$) for any observed titer (H),

$$logH = log_2(\frac{H}{10}) + 2, \tag{3.17}$$

such that the observed titers take on discrete values in the range [1,11]. We transform both the observed and latent titers as in Eq. 3.17. We assume that the observed log titer $logH_{\text{obs}i,s}(t)$ against subtype s is normally distributed around the latent log titer $logH_{i,s}(t)$:

$$logH_{\text{obs}i,s}(t) \sim \mathcal{N}(logH_{i,s}(t), \sigma) \tag{3.18}$$

where σ gives the standard deviation of the measurement error. Following other analyses that quantified the measurement error associated with different titer values [24, 49], we assign a lower value of the measurement error (σ = 0.74 log titer units, Table 3.4) for undetectable (<1:10) titers. The observed titer is censored at integer cutoffs, such that the likelihood of observing $logH_{\text{obs}i,s}(t) = k$ given latent titer $logH_{i,s}(t)$ is given by

$$\mathcal{L}(k \mid \theta, logH_{i,s}(t)) = \begin{cases} f(logH_{i,s}(t) \leq k), & k = 1 \\ f(k \leq logH_{i,s}(t) \leq k+1), & 2 \leq k \leq 10 \\ f(logH_{i,s}(t) \geq k), & k = 11 \end{cases} \tag{3.19}$$

where θ gives the parameter vector and f is specified as in Eq. 3.18.

Table 3.4 summarizes the estimated model parameters and latent states.

3.7.3 Likelihood-based inference

The titer dynamics for each individual are a partially observed Markov process (POMP). The model for each influenza subtype is a "panel POMP" object, or a collection of the individual POMPs with shared subtype-specific parameters. We use multiple iterated filtering, MIF [80] for simulation-based maximum likelihood inference. Iterated filtering

uses sequential Monte Carlo (SMC) to estimate the likelihood of observed time series. In SMC, a population of particles is drawn from the parameters of a given model to generate Monte Carlo samples of the latent dynamic variables. To evaluate the likelihood of a shared parameter set, SMC is carried out over the time series for each individual to generate a log-likelihood for the corresponding panel unit. The log likelihood of the panel POMP object is the sum of the individuals' log likelihoods. Iterated filtering successively filters the particle population, perturbing the parameters between iterations. The perturbations decrease in amplitude over time, such that the algorithm converges over time to the maximum likelihood estimate. For each model, we initialize the iterated filtering with 100 random parameter combinations. We perform series of successive MIF searches, with the output of each search serving as the initial conditions for the subsequent search. We use 10,000 particles for each optimization routine. The likelihood of the output for each search is calculated by averaging the likelihood from ten passes through the particle filter, each using 20,000 particles. We repeat the optimization until additional operations fail to arrive at a higher maximum likelihood. A typical MIF search from one starting parameter set requires about 36 cpu-hours of computing resources. For model selection, we used the corrected Akaike Information Criterion (AICc) [76]. We obtained maximum likelihood estimates for each parameter and associated 95% confidence intervals by constructing likelihood profiles.

3.7.4 Calculating imprinting probabilities

We calculate the probability that an individual's first influenza A virus was with a particular subtype (H1N1, H3N2, or H2N2) or that the individual was naive to infection at each year of observation. We assume that the first infection occurred between the ages of 6 months and 12 years old, as infants are protected by maternal antibodies for the first six months of life [11] and over 95% of children have been infected by influenza A before the age of 12 [151, 150]. Following the original imprinting model by Gostic and colleagues

[61], we estimate the probability that an individual with birth year i has his or her first IAV infection in calendar year j

$$\epsilon_{i,j} = \frac{(1-a)^{j-1}a}{\Sigma_{j=1}^{i+12}(1-a)^{j-1}a}. \tag{3.20}$$

Here, a is the constant annual attack rate in seronegative children as estimated by Gostic and colleagues (a = 0.28, [61]). Given observation year y, the probability that individual i was first infected in year j is:

$$\epsilon_{ij|y} = \begin{cases} \frac{a}{N_{i|y}} & y \geq i+12 \\ \\ \frac{a(\Pi_{k=1}^{j-1}(1-a))}{N_{i|y}} & y < i+12 \end{cases} \tag{3.21}$$

where $N_{i|y}$ is a normalizing factor that enforces the assumption that all individuals have their first infection by age 12 and ensures that all probabilities sum to one for individuals that are $>$ 12 years old at the observation date. The normalization factor does not apply to individuals that are $<$ 12 years old, who have some probability of being naive to infection. We combine the probabilities of the age of first infection with historical influenza A subtype frequency data from Hong Kong (or from the Southeast Asia region for years in which data from Hong Kong is unavailable) to determine the probability that an individual with birth year i had his or her first exposure to subtype S in year j:

$$p_{imp_{S,i|y}} = \Sigma_{j=i}^{y} f_{S|j}\epsilon_{i,j|y}. \tag{3.22}$$

Here, $f_{S|j}$ gives the fraction of positive specimens of subtype S out of all positive community surveillance samples that are positive for influenza A [168]. For individuals younger than 12 years old during the year of observation, the probability that an individual was

naive in observtaion year y is:

$$p_{naive_{i|y}} = 1 - \Sigma_{j=i}^{y} \epsilon_{i,j|y}. \tag{3.23}$$

3.8 Additional Information

	Notation	Type
Parameter		
Magnitude of short-term boost	$d_{\text{kids},s}$ $d_{\text{adults},s}$	Estimated (sub-model)
Variability of short-term boost	$\sigma_{\text{kids},s}$ $\sigma_{\text{adults},s}$	Estimated (sub-model)
Antibody ceiling effect	$k_{\text{kids},s}$ $k_{\text{adults},s}$	Estimated (sub-model)
Magnitude of long-term boost	$\zeta_{\text{kids},s}$ $\zeta_{\text{adults},s}$	Estimated
50% protective titer (TP_{50})	$\text{TP}_{50_{\text{kids},s}}$ $\text{TP}_{50_{\text{adults},s}}$	Estimated
Waning rate of non-HI-mediated immunity	$w_{\text{nonspecific}_{\text{kids},s}}$ $w_{\text{nonspecific}_{\text{adults},s}}$	Estimated
Waning rate heterosubtypic	$w_{\text{nonspecific}_m}$	Estimated
Scaled transmission rate	β_s	Estimated
Age-specific contact rate	$\beta_{c_{a_i}}$	Fixed
Infection duration	$\Gamma(\mu, \sigma_{\text{dur}})$	Fixed
Rate of short-term titer rise	r	Fixed
Duration: infection to peak titer	T_{peak}	Fixed
Rate of short-term titer waning	w	Fixed
Measurement error	σ $\sigma_{\text{undetectable}}$	Fixed
Latent state		
Titer	$h_{i,s}(t)$	Simulated
Baseline titer	$h_{\text{baseline}_{i,s}}(0)$	Fixed with error
Susceptibility	$q_{i,s}(t)$	Simulated
Time of infection	$t^{\text{X}}_{\text{i,s}}$	Simulated

Table 3.4: Model parameters and latent states.

Subtype		Epidemic	Simulated incidence	Observed incidence by titers
pH1N1	Adults	1	0.04 [0.02, 0.05]	0.05
		2	0.08 [0.07, 0.11]	0.12
		3	0.04 [0.03, 0.05]	0.06
	Children	1	0.08 [0.03, 0.13]	0.10
		2	0.17 [0.10, 0.21]	0.19
		3	0.06 [0.02, 0.10]	0.06
H3N2	Adults	1	0.17 [0.14, 0.19]	0.16
		2	0.19 [0.16, 0.22]	0.15
		3	0.05 [0.03, 0.06]	0.03
	Children	1	0.24 [0.17, 0.30]	0.21
		2	0.24 [0.17, 0.32]	0.24
		3	0.05 [0.02, 0.10]	0.03

Table 3.5: Cumulative incidence over observed epidemics. The simulated incidence was estimated from the latent simulated infections. The main and bracketed values give the median and 95% quantiles, respectively, from 1,000 simulations of the models at the maximum likelihood estimate parameter values. The observed incidence was estimated by 4-fold titer consecutive titer rises in the observed data.

Subtype		**Parameter**	MLE [95% CI]
H3N2	Adults	$d_{a_{i,s}}$	4.6 [3.1, 7.1]
		$\sigma_{a_{i,s}}$	1.4 [0.9, 2.3]
		$k_{a_{i,s}}$	1.0 [0.6 , 2.5]
	Children	$d_{a_{i,s}}$	5.1 [3.9,7.2]
		$\sigma_{a_{i,s}}$	1.5 [1.0, 2.2]
		$k_{a_{i,s}}$	0.5 [0.3, 1.0]
pH1N1	Adults	$d_{a_{i,s}}$	2.6 [2.4,5.2]
		$\sigma_{a_{i,s}}$	1.9 [1.4, 2.6]
		$k_{a_{i,s}}$	0.3 [0.1, 1.8]
	Children	$d_{a_{i,s}}$	4.5 [3.5, 5.5]
		$\sigma_{a_{i,s}}$	0.9 [0.7, 1.3]
		$k_{a_{i,s}}$	0.8 [0.6, 1.2]

Table 3.6: Maximum likelihood estimates of the parameters that govern the short-term titer dynamics, with associated uncertainty.

Infecting Subtype	**Imprinting Group**	**Population subset**	**MLE [95% CI]**
pH1N1	Group 1	Ages 35 - 50	1.0 [0.6, 1.8]
H3N2	Group 2	Ages 35 - 50	0.8 [0.4, 1.2]

Table 3.7: Maximum likelihood estimate of the group-level imprinting effects ($\alpha_{\text{imp}_{\text{Group1}}}$ and $\alpha_{\text{imp}_{\text{Group2}}}$) among subsets of the study population.

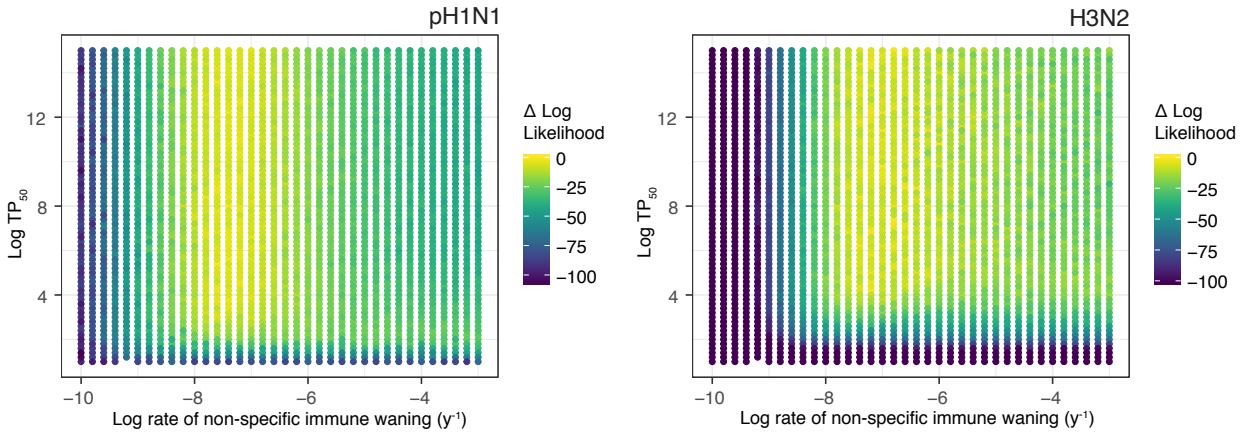

Figure 3.5: Bivariate likelihood profile of the log TP_{50} and the log rate of non-specific immune waning in adults for pH1N1.

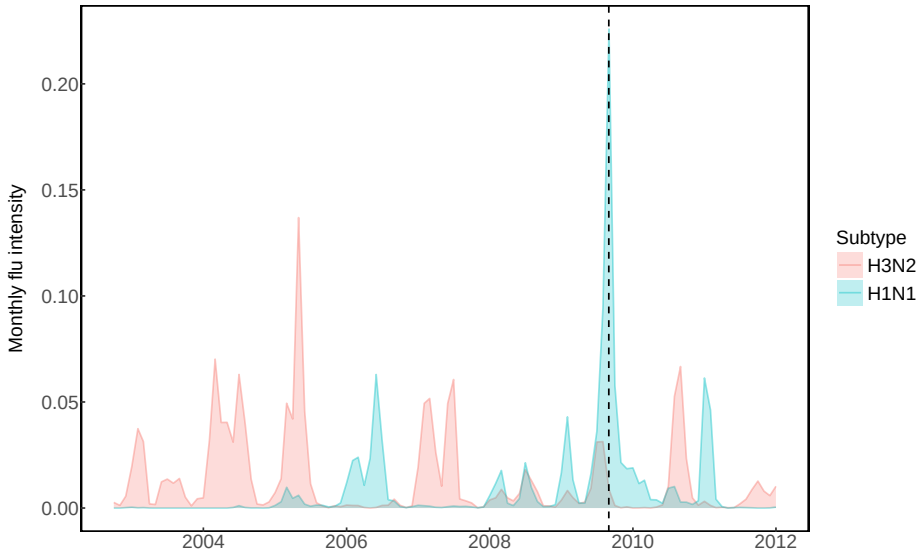

Figure 3.6: Subtype-specific flu intensity (ILI × % positive) in Hong Kong prior to the start of the study. The black vertical dashed line denotes the earliest observation date.

3.8.1 Short-term titer dynamics after PCR-confirmed infection

3.8.1.1 Infection generates a variable short-term homosubtypic antibody boost that declines with increasing pre-infection titer.

We find substantial variability in antibody titer responses after PCR-confirmed infection with H3N2 and pH1N1, respectively, for both children and adults (Table 3.6), which is

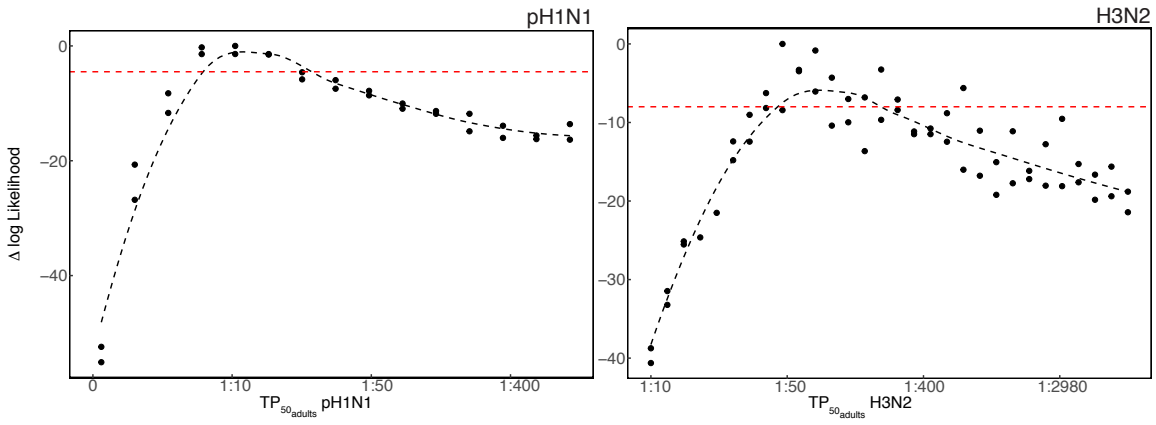

Figure 3.7: Likelihood profile of the TP_{50} for pH1N1 and H3N2. The red dotted line gives the threshold of 95% significance based on the spline (black dotted line).

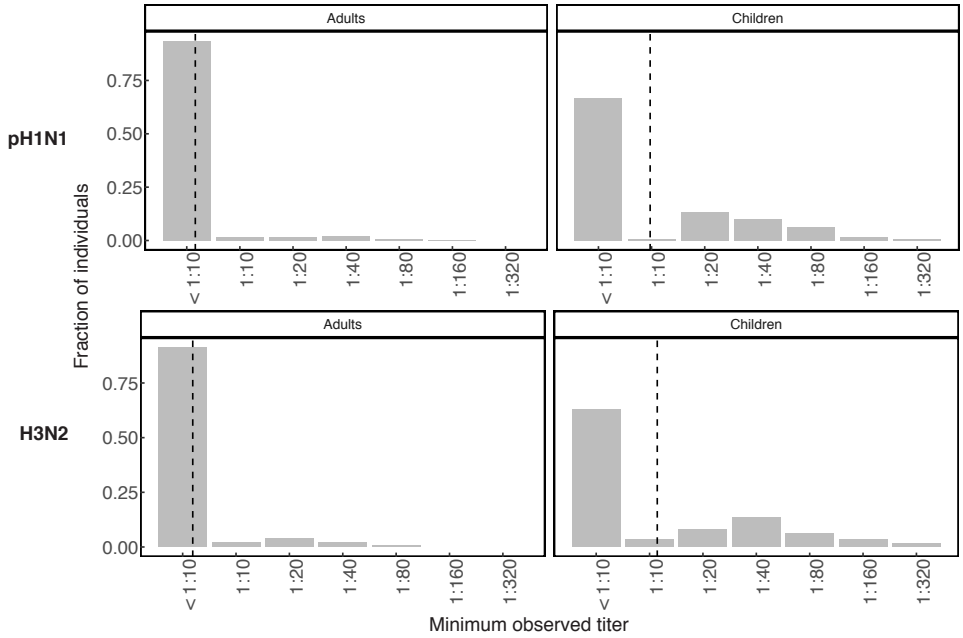

Figure 3.8: Distribution of the minimum observed (baseline) titers in children and adults for pH1N1 and H3N2. The vertical dashed line gives the geometric mean baseline titer.

consistent with other analyses [52, 49]. The inferred standard deviation of the lognormal titer boost distribution (Eq. 3.12) ranges from 0.9 to 1.9 log titer units among children and adults for pH1N1 and H3N2 (Table 3.6). The mean magnitude of the boost is higher for H3N2 than for H1N1 in both age groups. The variability in the acute infection response and the difference in the response between subtypes and age groups suggest that

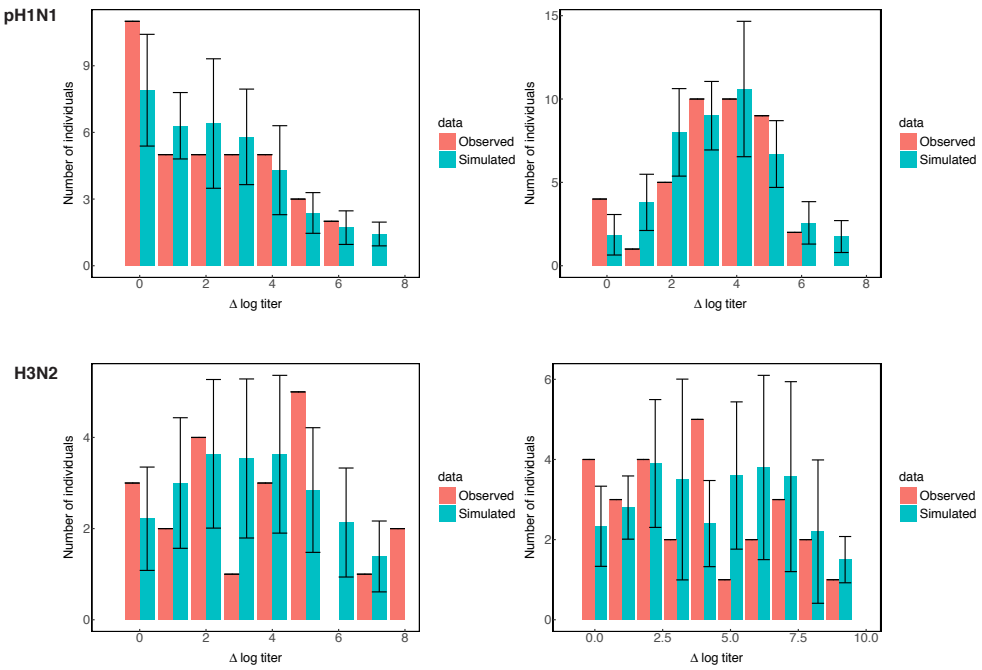

Figure 3.9: Observed vs. simulated distribution of titer boosts from the sub-model in individuals with PCR-confirmed infections, represented as the change in log titer between the pre- and post-infection observations. Error bars give the 95% CI among simulations.

threshold titer values used in sero-surveillance may not reliably predict infection in all individuals [74, 32].

Consistent with evidence from multiple studies of post-vaccination titer dynamics [81, 128], our models support an antibody ceiling effect for both subtypes in children and adults (Table 3.8), such that higher pre-infection titer diminishes the realized boost. The raw data for both subtypes exhibit a linear decline in the difference between pre- and post-infection log titers with increasing pre-infection titer (Fig. 3.14). In the models of the short-term titer dynamics, we tested for the presence of an antibody ceiling effect via the parameter $k_{a_i,s}$ (Eq. 3.9). For both subtypes, models that include the antibody ceiling effect ("with k") models that do not ("without k") in children and adults (ΔAICc > 2, Table 3.8). Therefore, part of the individual variation in the acute infection response can be explained by differences in pre-existing titers. Simulations from the MLEs of the best fit models of the short-term dynamics reproduce the shape of the observed distribution of

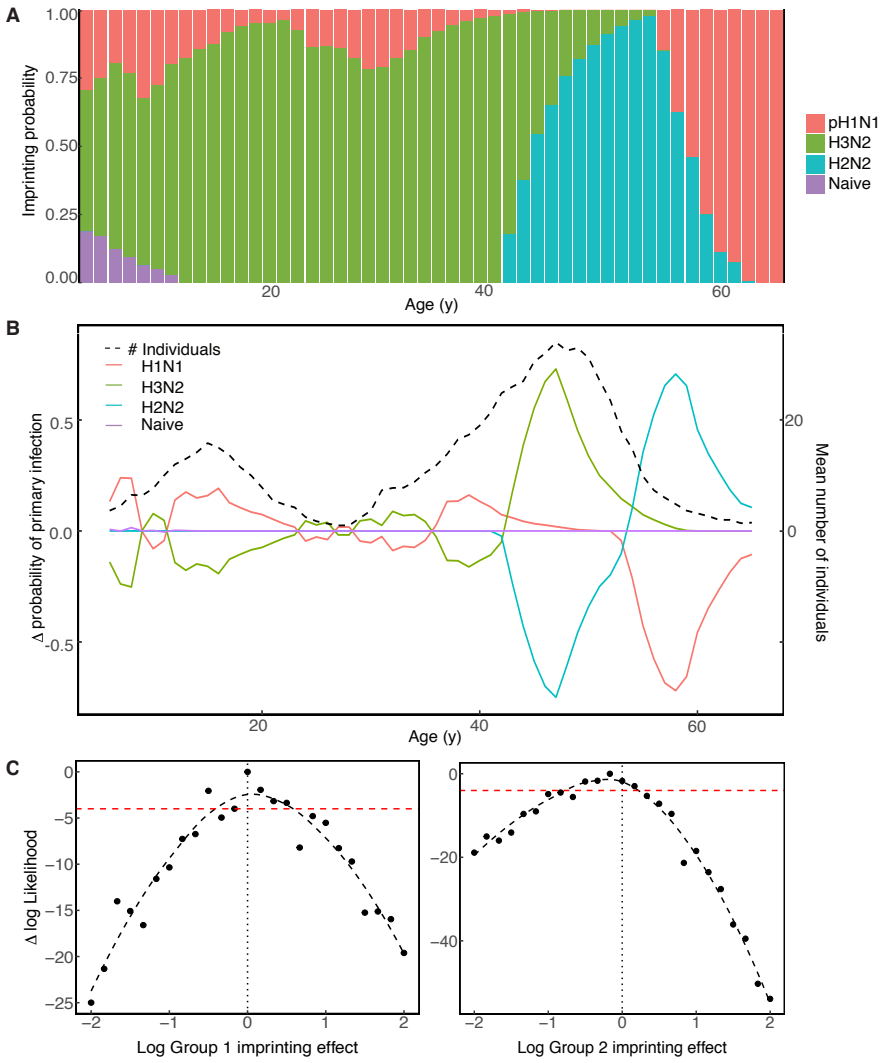

Figure 3.10: **A** Probability of imprinting by historically circulating influenza A subtypes by age in 2009. **B** Change in the mean probability of primary infection with historically-circulating subtypes by age between 2009 and 2014. The black dashed line gives the mean number of individuals by age that were observed in the data between 2009 and 2014. **C** Likelihood profiles for the effect of imprinting by H2N2 on the risk of infection with pH1N1 (left) and the effect of imprinting by H3N2 on the risk of H3N2 infection (right). Values of the log parameter less than 0 (vertical dotted line) indicate a protective imprinting effect. The red dashed horizontal line gives the threshold for statistical significance at a 95% level.

titer boosts in children and adults following PCR-confirmed infection for both subtypes (Fig. 3.9).

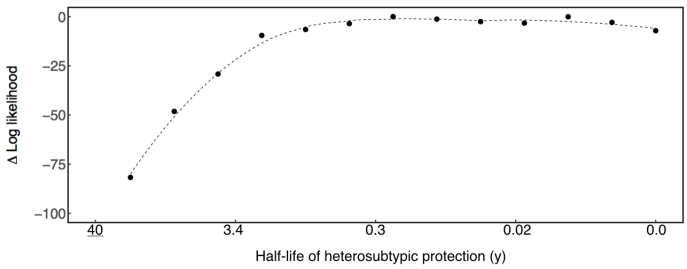

Figure 3.11: Likelihood profile for the rate of waning of heterosubtypic protection.

Subtype		**Model**	**Log likelihood (SE)**	**ΔAICc**
H3N2	Adults	with k	-45.5 (0.04)	0
		without k	-48.6 (0.03)	3.6
	Children	with k	-77.8 (0.07)	0
		without k	-88.0 (0.05)	13.0
pH1N1	Adults	with k	-79.5 (0.06)	0
		without k	-81.7 (0.03)	2.1
	Children	with k	-91.8 (0.01)	0
		without k	-95.9 (0.04)	5.6

Table 3.8: Model comparisons for sub-model of short-term boosting.

3.8.1.2 Model of short-term antibody boost following PCR-confirmed infection

To estimate the mean magnitude $d_{a_{i,s}}$ and variability $\sigma_{a_{i,s}}$ of the short-term titer boost and the dependence $k_{a_{i,s}}$ of the boost on the pre-infection titer, we fit models to the observed titers preceding and immediately following a subtype-specific positive swab in individuals with PCR-confirmed infection (Fig. 3.12).

We fix the pre-infection latent titer $h_{i,s}(0)$ to the observed pre-swab titer $h_{\text{OBS1}_{i,s}}$, allowing for two-fold uncertainty in the measured titer as in Eq. 3.16. We fix the latent time of infection $t_{\text{i,s}}^{\text{INF}}$ based on the date of the positive swab, assuming that the swab occurred during an infected period that we draw from a gamma distribution with fixed parameters (Table 3.4). We model the dynamics of the short-term titer rise as in Eq. 3.9, with the rate of rise r and time T_{peak} between infection and peak titer fixed as in Table 3.4. Following the peak titer response, we assume that the titer wanes at rate w (fixed as in Table 3.4) until the time of the second observed value $h_{\text{OBS2}_{i,s}}$.

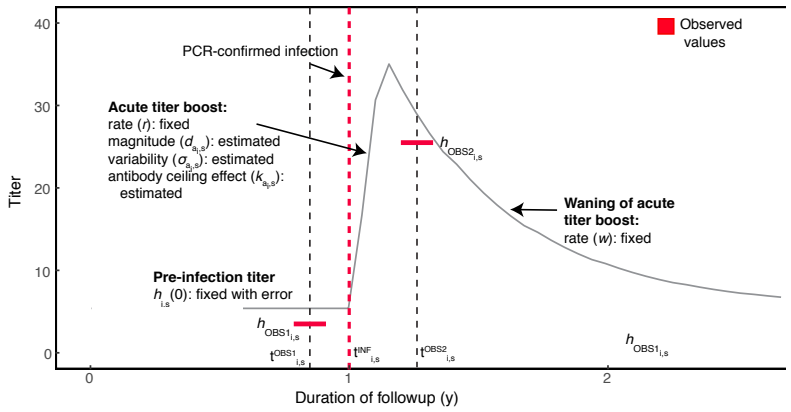

Figure 3.12: Schematic of the acute HI titer dynamics for individual i against subtype s, given PCR-confirmed infection at time $t^{\text{INF}}_{\text{i,s}}$.

3.8.1.3 Observed titer boosts secondary to symptomatic vs. non-symptomatic infections and primary vs. secondary infections

Our model of the short-term titer dynamics does not distinguish between symptomatic infections and non-symptomatic infections that may have been detected incidentally given illness in another household member. We define symptomatic infection as the presence of symptoms consistent with acute respiratory illness (ARI) during the two weeks prior to a PCR-confirmed infection. Based on data from household symtpom diaries, approximately 70% of infections in both children and adults for both subtypes were symptomatic (Table 3.9).

Subtype		% Symptomatic infections	% Primary infections
H3N2	Adults	76.2%	38.1%
	Children	69.0%	75.8%
pH1N1	Adults	69.4%	38.9%
	Children	64.3%	54.8%

Table 3.9: Fraction of children and adults with symptomatic infections (defined by an ARI in the two weeks prior to PCR-confirmed infection) and primary infections (defined by the absence of infection with or without ARI symptoms in other household members in the two weeks prior to PCR-confirmed infection) for pH1N1 and H3N2.

Children were more likely than adults for both subtypes to have a primary, or index

case infection, meaning that no other household members had a PCR-confirmed infection or symptoms of an ARI in the two weeks prior to confirmed infection. We compare the distributions of titer changes (the difference between the pre- and post-infection log titers normalized by the time between the pre- and post-infection sample dates) between symptomatic and non-symptomatic infections and between primary and secondary infections (Fig. 3.13).

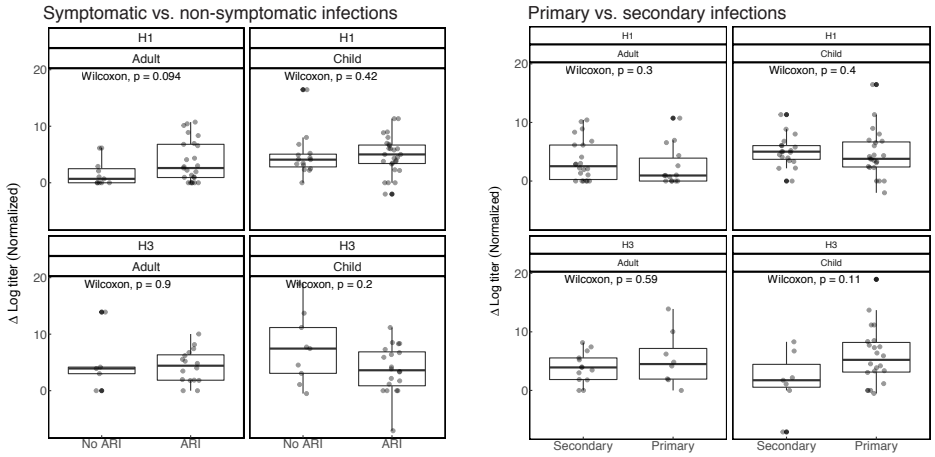

Figure 3.13: Distribution of normalized titer boosts after PCR-confirmed infections based on whether infections were symptomatic or non-symptomatic (Left) and on whether infections were primary or secondary (Right). Normalized titer boosts are calculated as the difference in the log titer between pre- and post- infection sera divided by the length of time in years between the pre- and post-infection samples. Box plots give the median and interquartile range of the normalized titer boosts, and the individual data points are overlain with horizontal jitter. Differences in the mean of the distributions are determined by non-parametric Wilcoxon tests at a 0.05 significance level.

We find no statistically significant difference in the mean normalized titer boost between symptomatic and non-symptomatic infections for either subtype in children or adults. Similarly, we find no statistically-significant differences when comparing primary and secondary infections. Therefore, the raw data suggest that such differences in the nature of the PCR-confirmed infections do not meaningfully affect our inference of the short-term dynamics.

3.8.1.4 *Evidence of the antibody ceiling effect in the raw data*

The raw data for both subtypes exhibit a linear decline in the difference between pre- and post-infection log titers with increasing pre-infection titer (Fig. 3.14). The linear decline

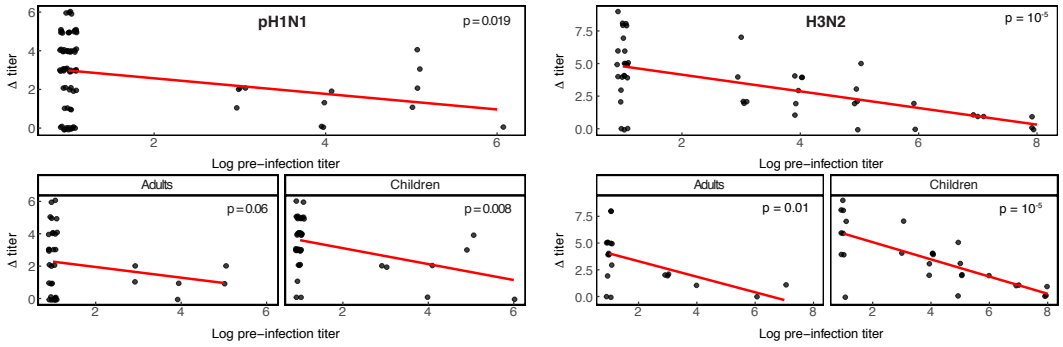

Figure 3.14: Difference in the log pre- and post-infection titer as a function of the log pre-infection titer for individuals with PCR-confirmed infection with pH1N1 (left) and H3N2 (right). The top panel for each subtype gives the relationship for the aggregate data in children and adults. Note that log titers are defined as in Eq. 3.17.

is statistically significant with $p < 0.02$ for both pH1N1 and H3N2. When we stratify the analysis in children and adults for each subtype, the relationship is statistically significant with $p \leq 0.01$ for children with both subtypes and for adults with H3N2 ($p = 0.06$ for adults with pH1N1).

3.8.2 *Model validation and sensitivity analysis*

3.8.2.1 *The model reproduces the observed distribution of titer rises among individuals.*

We compared the observed number of 2-,4-, and 8-fold increases in consecutive titer measurements for H3N2 and pH1N1 to the distribution obtained from 1,000 simulations of the model at the maximum likelihood parameter estimates (Figs. 3.15, 3.16). The model accurately reproduces the observed distributions in children and adults for both subtypes.

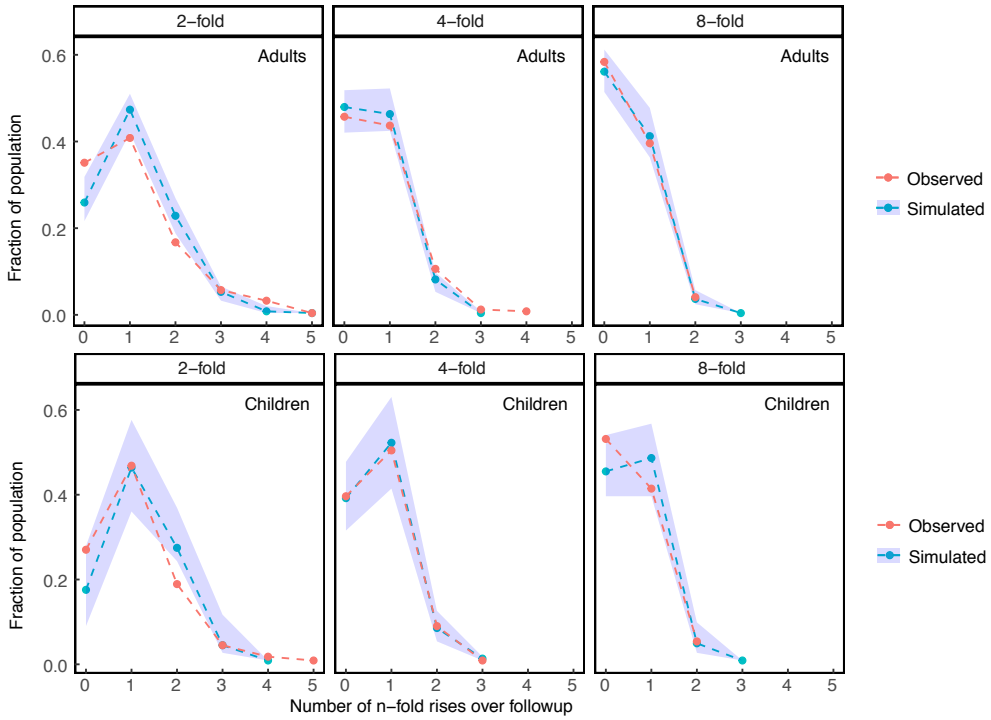

Figure 3.15: Observed vs. simulated distribution of consecutive 2-fold, 4-fold, and 8-fold titer rises per individual in the H3N2 data. The dashed blue lines give the median number of individuals with n = 0,1,... 2- or 4-fold titer rises from 1,000 simulations of the model, and the shaded blue areas are bounded by the 2.5% and 97.5% quantiles.

3.8.2.2 *The maximum likelihood parameter estimates are robust to assumptions about the initial conditions.*

For the main analysis, we draw each individual's time of most recent pre-observation infection from the density of the subtype-specific influenza intensity in the seven years prior to the first observation. For comparison, we fitted the best-fit model for each subtype in children and adults using two alternative assumptions about the initial conditions. First, we drew the time of most recent infection from the density of the subtype-specific influenza intensity over the five years prior to the first observation ("Five years", Figs. 3.17,3.18). Second, we drew the time of most recent infection uniformly over the seven years prior to the first observation rather than using $L_{s,t}$ ("Uniform draw", Figs. 3.17, 3.18). The maximum likelihood estimates for the alternative models fall within the 95% CI of the parameter estimates from the main analysis .

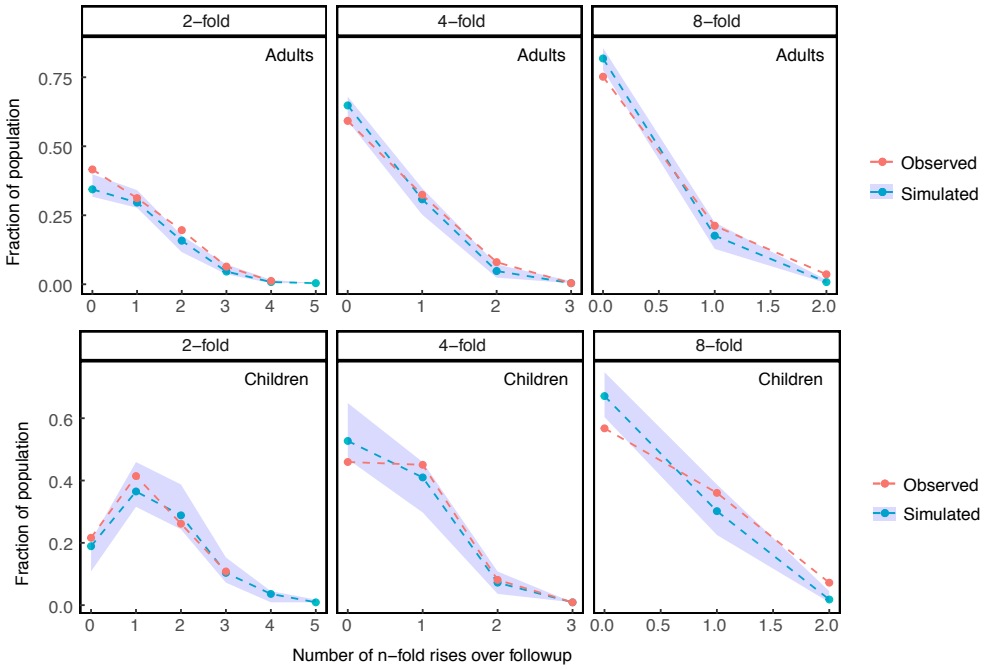

Figure 3.16: Observed vs. simulated distribution of consecutive 2-fold, 4-fold, and 8-fold titer rises per individual in the pH1N1 data. The dashed blue lines give the median number of individuals with $n = 0,1,...$ 2- or 4-fold titer rises from 1,000 simulations of the model, and the shaded blue areas are bounded by the 2.5% and 97.5% quantiles.

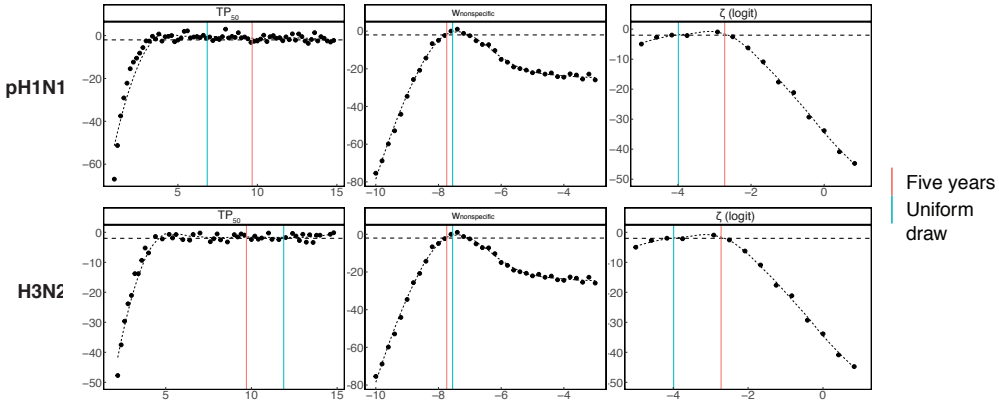

Figure 3.17: Likelihood profiles for the estimated parameters of the best-fit models in **adults** for pH1N1 and H3N2. The dashed horizontal line gives the threshold for statistical significance at a 95% level. The vertical lines denote the MLEs from models under alternative initial conditions.

3.8.2.3 The inference results are robust to rescaling of the community intensity of pH1N1 during the 2009 pandemic.

During the first wave of pandemic influenza H1N1 in 2009, the sentinel surveillance system was affected by increased reporting rates and by changes in health-care seeking

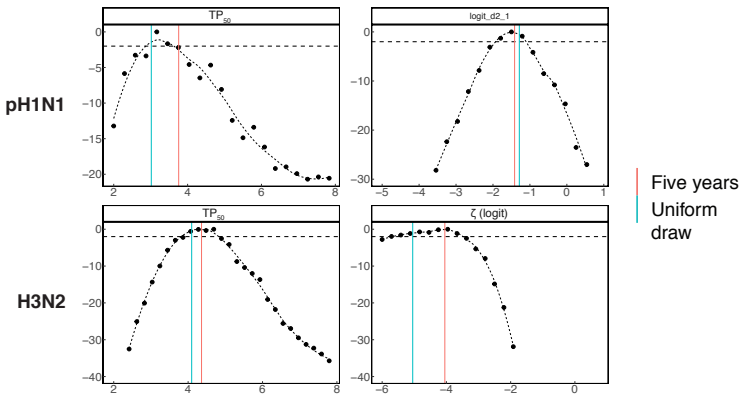

Figure 3.18: Likelihood profiles for the estimated parameters of the best-fit models in **children** for pH1N1 and H3N2. The dashed horizontal line gives the threshold for statistical significance at a 95% level. The vertical lines denote the MLEs from models under alternative initial conditions.

behaviors due to increased media and government attention to influenza [33, 112]. We re-fitted our models of pH1N1 after scaling the community flu intensity to reflect these differences. A previous study estimated separate scaling factors for the relationship between the pH1N1 intensity proxy and the risk of infection before and after a November 2009 change point during the 2009 pandemic in children and adults [164]. We rescaled our estimate $L_s(t)$ of the 2009 pandemic pH1N1 intensity by multiplying the intensity before the change point by the ratio ρ of the estimated post- and pre- change point scaling factors in children ($\rho = 0.25$) and adults ($\rho = 0.29$). Fig. 3.19 shows the rescaled intensity. Notably, our observations begin at the end of the 2009 pandemic. Fewer than 6% of observations in children and fewer than 5% of observations in adults occurred before the November 2009 change point. Fewer than 1% of observations in children and adults occurred before October 2009. The models recovered the same maximum likelihood parameter estimates given the rescaled pandemic intensity (Fig. 3.20).

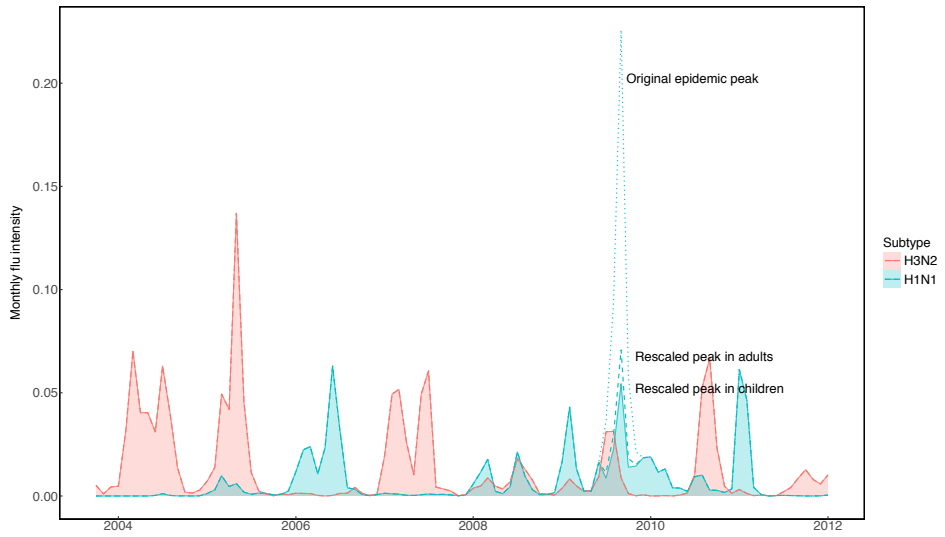

Figure 3.19: Rescaled community intensity of pH1N1 during the 2009 pandemic in adults (dashed blue line) and in children (solid blue line and shading) compared to the original intensity reported by community surveillance (blue dotted line).

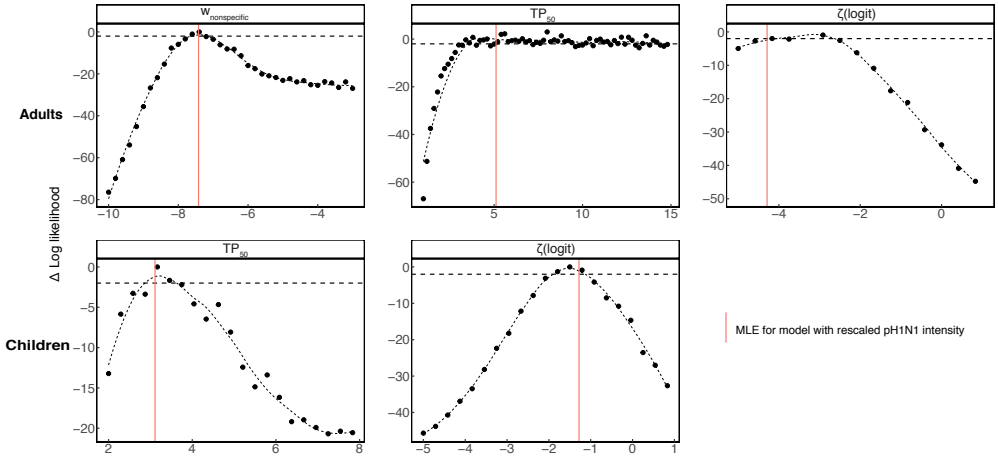

Figure 3.20: Likelihood profiles for the estimated parameters of the best-fit models in adults and children for pH1N1. The dashed horizontal line gives the threshold for statistical significance at a 95% level. The vertical lines denote the MLEs from the model with rescaled pH1N1 intensity during the 2009 pandemic.

3.8.2.4 The measurement error estimated from replicate titer measurements is consistent with literature estimates.

The sera from three visit dates were measured twice. In our models, we used the first titer measurement for each serum sample (the measurement recorded closest to the sampling

date). To approximate the measurement error, we calculated the difference in measured titer between the second and first replicates (Fig. 3.21). For detectable titer levels ($>$ 1:10), the standard deviation of the error distribution (sd = 1.2 log titer units) matches the measurement error that we fixed in the model according to estimates from the literature (Table 3.4). The negative central tendency of the difference between the second and first replicates among detectable titers (median = -0.98 log titer units log titer units) indicates that measured titer generally declines with time since sampling. Furthermore, in line with previous analyses [24], we find that the error distribution is zero-inflated for undetectable titers $<$1:10 (Fig. 3.21), justifying our use of a separate measurement error for undetectable titers. A previous study estimated the probability of 2-fold measurement error for undetectable titers [24]. We therefore calculated the corresponding error (σ = 0.74) in our normally distributed observation model that would yield the same probability of 2-fold measurement error for undetectable titer values. The observation model is non-invertible (Eq. 3.18). Therefore, while we use the measurement error to draw simulated observations from a normal distribution centered around the latent log titers, we cannot back-calculate the value of the latent titers from observed data. This is why we assign the initial baseline titer $h_{\text{baseline}_{i,s}(0)}$ from a possible two-fold range surrounding the lowest observed titer $h_{\text{obs}_{min_{i,s}}}$ (Eq. 3.16).

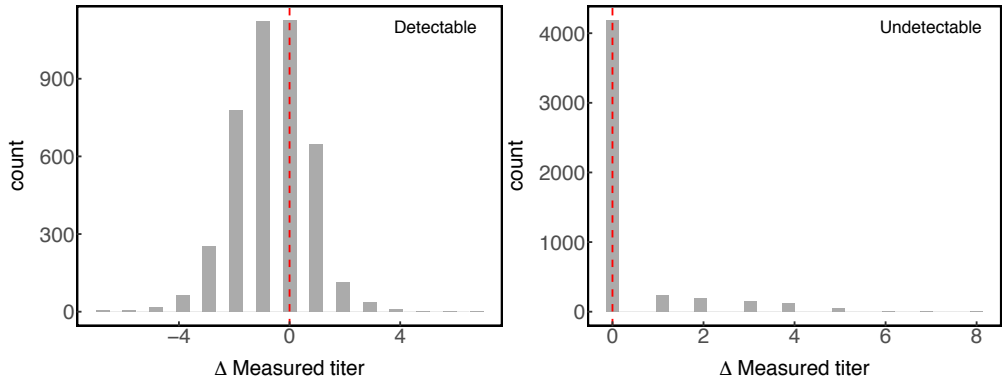

Figure 3.21: Distribution of difference between the second and first measured titer for sera that were tested twice, with distributions shown separately for detectable and undetectable titers based on the initial measurement. The vertical red line marks zero difference.

CHAPTER 4

CONCLUSION

In his "Homage to Santa Rosalia," Hutchinson [77] asked "Why are there so many types of animals?" This simple question underscores a century's worth of ecological theory about the dynamics of coexistence among competing species. Pathogen diversity poses an exquisitely intricate extension of theoretical questions about species coexistence in a variable environment. Pathogen competition occurs on a playing field marked by immense variation among human hosts. Host heterogeneity in population structure can alter the spatial and temporal scales of pathogen transmission [169]. Host heterogeneity in the immune response complicates the "pathogen-centric" predictions of most multi-strain models, which assume that individuals are invariant in their response to the same antigenic variant [30]. This dissertation provides coarse predictions about how the interplay between immune-mediated competition and host heterogeneity might explain the observed diversity of two viral communities: HPV and influenza A viruses. Both analyses highlight the importance of host heterogeneity as a driver of disease dynamics and pathogen diversity.

Chapter 2 investigated the factors that explain the diversity of over 200 low-prevalence HPV types. We tested for homologous immunity in HPV infection that might drive type coexistence by NFDS and promote antigenic niche partitioning. We found that, rather than inducing protective immunity, HPV infection strongly increases the risk of future infection by the same type. The infection risk for any type in a susceptible population is, on average, low and concentrated in high-risk individuals defined by demographic and behavioral risk factors. Thus, our results suggest that high HPV prevalence results from frequent reinfection or persistent infections within individuals.

Chapter 3 tested hypotheses about age-related differences in the dynamics of protection after infection with influenza A viruses. Because protection wanes by 50% over 2-5 years for both subtypes, multiple infections with the same subtype might change the

specificity of an individual's antibody response over time, yielding individual variation in the response to circulating influenza strains. We found that HI-mediated immunity alone best explains the dynamics in children, suggesting that protective homosubtypic antibody responses early in life are directed against highly mutable epitopes on the HA head. Non-HI-mediated immunity correlates with protection in adults, suggesting that the antibody repertoire evolves with age, which might be explained by repeated exposures focusing antibodies to more conserved epitopes. Broadly, our findings suggest that circulating influenza strains face different selective pressures in children and adults.

Our current understanding of host-pathogen interactions only scrapes the surface of the interplay between pathogens and the myriad aspects that define their human host environment. This dissertation used simplified mathematical abstractions to demonstrate the importance of different forms of host heterogeneity for pathogen dynamics and diversity. However, current limitations of biological insight offer exciting directions for future research. To conclude, I summarize the implications of the results of each chapter for future research and for vaccination and public health management.

4.1 Future directions and implications

Progress towards understanding HPV diversity will depend on better resolution of the ways in which HPV types establish niches within and among human hosts. Perhaps the biggest open question is how the dynamics of HPV infection might differ in women. For every HPV type, the most significant host-specific risk factor was the number of recent female sex partners, indicating that the factors that promote HPV type diversity in women are essential to understand diversity in men. The finding of higher seroprevalence [111] and lower prevalence of HPV types in women than men despite similar type-specific infection durations and distributions [59, 71] suggest that women generate some degree of homotypic immunity. However, serum antibodies in women appear to offer only weak protection against future infection [8] and episodic infection is common in women within

and across anatomic sites [60, 153], suggesting that immunity may be too weak to generate significant selective pressure, especially given frequent transmission to male partners [21]. However, the ecological niches of HPV types are certainly defined by properties of the epithelia of the female genitals and reproductive tract. One interesting question is whether HPV types are limited to the same epithelial environment in women and men, and how such overlap in tissue specificity might define the high-risk subpopulations for each type. For example, are HPV types that are more adapted to the highly keratinized surface of the circumcised penis in men also more adapted to the keratinized external genitalia of women? Likewise, are HPV types that are highly adapted to the cervical mucosa also preferentially transmitted at the anal mucosa among MSM males?

In addition to the fact that our analysis was limited to men, we identified differences between high-risk subpopulations among HPV types based on a limited set of host-specific risk factors, assuming a well-mixed population. Therefore, we provided a low-dimensional approximation of the complicated spectrum of factors that define high risk subpopulations. Consistent with many other studies [147], we identified variation in host susceptibility primarily through the effects of sexual partnerships. A promising direction for mechanistic models of HPV dynamics is to quantify the how finer-scale biological factors such as obesity [53], genetic variability in HLA genes [36], the composition of genital microbiomes [18], sex hormones [103], and the dynamics of epithelial turnover [148] may define type niches.

We found that HPV types rely on repeated or continuous infection within individuals to maintain their prevalence. However, future work must distinguish the possibility of viral latency and reactivation from auto-inoculation across infected sites. This pursuit will likely require higher-resolution molecular data about viral load and viral gene expression during infections within a large number of individuals over time. With such fine-scale data, epidemiological models could make more detailed predictions about the within-host dynamics of infection that allow HPV types to persist in certain individuals.

It appears that the the stable, ordered prevalence of HPV types in men cannot be explained by stabilizing competition through cross-immunity. Rather, it is possible that HPV diversity depends on multi-level meta-population dynamics, where dispersal is limited both within hosts by isolated, patchy infections across anatomical sites, and between hosts by sustained transmission within high-risk subpopulations. Initial HPV infections rely on opportune micro-lesions in basal epithelial cells, such that viral colonization of a within-host patch at any given exposure is relatively rare and infection events are partly stochastic [148]. Given that our results support evidence of slight type-specific differences in tissue tropism [106, 16], the prevalence of any type might depend on the frequency of its exposure by sexual transmission to the right epithelial environment. This reasoning is in line with the results, here and in many other studies [57, 56, 116, 117, 59], that sexual partnerships are the largest driving factors of HPV transmission. Once an infection is established, auto-inoculation might then sustain the type within individual hosts. While types may differ in fitness based on their ability to colonize or persist in host tissue [16], isolation to different types of hosts and different patches of epithelium within hosts might prevent competition and maintain the lowest-prevalence types at a population level.

The results of Chapter 2 suggest that the benefits of HPV vaccination might be stronger than previously expected. Because sterilizing immunity reduces infection rates, models that assume that infection induces protective immunity in men [42, 130, 165, 17] likely underestimates vaccine effectiveness. Furthermore, the effects of recurrent infection on subsequent risk suggest that vaccination may alter overall infection rates in ways that are not predicted from standard models of HPV dynamics [130, 42, 139]. Primarily, our results suggest that the best way to reduce HPV prevalence is to prevent the initial infection by vaccinating young boys. However, if indeed recurring HPV infection is driven mainly by auto-infection, then vaccinating previously infected individuals may also dramatically reduce HPV prevalence [142].

Understanding the dynamics of influenza diversity will require determining the spe-

cific factors that determine individual heterogeneity in immunity and shape the selective pressures acting on circulating strains. Our results suggest that influenza infection generates a highly variable short-term antibody response. We showed that stochastic variation in the magnitude of the short-term titer boost at each infection was important for the dynamics. However, several lines of evidence suggest that boosting should be consistent within individuals. Several models suggest that infection reactivates previous B cell responses in a reliable way based on exposure history [49, 100, 95]. Therefore, particular individuals might be predisposed to boost to new circulating epitopes. Predictable variation in boosting may also arise from genetic predisposition. For example, variation in MHC class II alleles determine the specificity of CD4+ helper T cells, which aid in B cell selection. Therefore, variation by HLA genotype could predispose hosts to have characteristic antibody responses [35]. A more mechanistic understanding of the factors that shape the diverse short-term immune response among individuals would allow for models with more precise functional forms for the boosting variability. Longer individual-level time series of serological observations and PCR-confirmed infections could also allow us to observe and quantify consistent differences between individuals.

Our results suggest that immunity in adults and children might generate different selective pressures on circulating strains. We showed that general forms of non-HI-mediated immunity are important for the dynamics of protection in adults, indicating that the response in adults does not targeting epitopes on the HA head. Instead, protection in adults might be driven by NA-mediated immunity [115], by stalk-directed anti-HA antibodies [133, 5], or by cellular immunity [96, 174, 62, 122]. The outcomes of age-related differences in the forms of protection for strain diversity will depend on the breadth and strength of these responses. One possibility is that the dynamics of influenza infection in adults are neutral with respect to antigenic HA variants if adults target completely conserved sites. Under this scenario, the immune response in adults could reduce fitness differences between competing strains, promoting strain diversity among older individu-

als and suggesting that selection by immunity in children limits diversity. This effect that has been demonstrated with gradually developing nonspecific immunity to pneumococcoal serotypes [29]. However, this prediction may not hold for flu, where age-specific differences in attack rates between seasons indicate that adults may indeed vary in their susceptibility to circulating variants [173].

Future work must identify new correlates of protection that encompass variation in immune responses. Sero-surveillance should rely on viral assays that capture responses to epitopes on the HA stalk and on NA to identify individuals that target more conserved sites. Furthermore, as cellular immunity may influence homosubtypic or heterosubtypic protection, genetic variation in susceptibility determined by polymorphisms in HLA genes may be another important predictor of individual variability [136]. Future models could integrate host genetic information with more comprehensive serology and clinical data to weight the contributions of heterogeneity in humoral and cellular responses in the dynamics of protection.

Patterns of social contacts drive community-level influenza transmission rates [94], and they may have important consequences for influenza diversity given our predictions about host heterogeneity in immune specificity. Therefore, future work should incorporate the dynamics of individual-level variation in immune responses within the context of population structure. For example, if different immune mechanisms drive protection in children and adults, then age-assortative mixing could concentrate the transmission of different strains within subpopulations of individuals that respond similarly to infection. Influenza strains could evolve on short timescales to escape immunity within subpopulations, contributing to broader patterns of antigenic evolution over time. As future work generates more insight into the specificity of individual immune responses, transmission models can synthesize more informed estimates about population immunity with social contact rates to explain and predict evolution.

The heterogeneity in immune specificity is important for influenza vaccine design.

The goal of vaccines is to permanently reduce incidence, but this goal is seldom realized due to immune escape. An ideal influenza vaccine would limit the diversity of the immunophenoytpe of a circulating strain, eliminating variability among hosts with respect to vaccine-induced immunity and thereby preventing escape [28]. Subpopulations of varying immune specificity may require targeted vaccines to achieve uniform vaccine-induced immunity at the population level. If homosubtypic and heterosubtypic competition act via cellular or innate immunity, then vaccines that mimic natural infection by inducing both humoral and cellular responses (e.g. live attenuated vaccines) could be effective against multiple subtypes.

4.2 Concluding anecdote

This thesis developed a unified perspective to understand the complex epidemiological patterns underying the diversity of two vastly different pathogen systems. HPV and influenza A viruses differ in their genetic diversity, in their interactions with host immunity, and in the subpopulations that they target. My work on these systems has led me to a somewhat paradoxical observation about the challenges in public health management for each system. Vaccines against HPV are highly effective [84] and our estimates about HPV dynamics suggest that vaccination may be even more beneficial than previously expected, especially among men. However, the benefits of HPV vaccines are limited by low vaccination rates driven in part by significant social stigma around sexually transmitted infections [75]. In influenza, by contrast, vaccination rates are relatively high in the USA and in other parts of the world [127], but the potential benefits of vaccination are limited by gaps in biological knowledge about influenza immunity that impede vaccine design. In many ways, this problem underscores the importance of coordination of scientists, public health officials, and health-care providers to both develop effective protection against disease and to ensure vaccination of high-risk individuals. I am enthusiastic to build my career working at the intersection of these fields.

REFERENCES

[1] Ali Akram and Robert D. Inman. Immunodominance: A pivotal principle in host response to viral infections. *Clinical Immunology*, 143(2):99–115, 5 2012.

[2] Ginesa Albero, Xavier Castellsagué, Hui-Yi Lin, William Fulp, Luisa L Villa, Eduardo Lazcano-Ponce, Mary Papenfuss, Martha Abrahamsen, Jorge Salmerón, Manuel Quiterio, Alan G Nyitray, Beibei Lu, F Xavier Bosch, and Anna R Giuliano. Male circumcision and the incidence and clearance of genital human papillomavirus (HPV) infection in men: the HPV Infection in men (HIM) cohort study. *BMC Infectious Diseases*, 14(1):75, 12 2014.

[3] R. M. Anderson and R. M. May. Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *Journal of Hygiene*, 94(03):365–436, 6 1985.

[4] Sarah F. Andrews, Yunping Huang, Kaval Kaur, Lyubov I. Popova, Irvin Y. Ho, Noel T. Pauli, Carole J. Henry Dunand, William M. Taylor, Samuel Lim, Min Huang, Xinyan Qu, Jane-Hwei Lee, Marlene Salgado-Ferrer, Florian Krammer, Peter Palese, Jens Wrammert, Rafi Ahmed, and Patrick C. Wilson. Immune history profoundly affects broadly protective B cell responses to influenza. *Science Translational Medicine*, 7(316):316–192, 12 2015.

[5] Sarah F Andrews, M Gordon Joyce, Michael J Chambers, Rebecca A Gillespie, Masaru Kanekiyo, Kwanyee Leung, Eun Sung Yang, Yaroslav Tsybovsky, Adam K Wheatley, Michelle C Crank, Jeffrey C Boyington, Madhu S Prabhakaran, Sandeep R Narpala, Xuejun Chen, Robert T Bailer, Grace Chen, Emily Coates, Peter D Kwong, Richard A Koup, John R Mascola, Barney S Graham, Julie E Ledgerwood, and Adrian B Mcdermott. Preferential induction of cross-group influenza A hemagglutinin stemspecific memory B cells after H7N9 immunization in humans. *Sci. Immunol*, 2(14), 2017.

[6] Bertran Auvert, Joelle Sobngwi-Tambekou, Ewalde Cutler, Marthi Nieuwoudt, Pascale Lissouba, Adrian Puren, and Dirk Taljaard. Effect of male circumcision on the prevalence of high-risk human papillomavirus in young men: results of a randomized controlled trial conducted in Orange Farm, South Africa. *Journal of Infectious Diseases*, 199(1):14–19, 1 2009.

[7] Oliver Balmer and Marcel Tanner. Prevalence and implications of multiple-strain infections. *The Lancet Infectious Diseases*, 11(11):868–878, 11 2011.

[8] Daniel C. Beachler, Gwendolyne Jenkins, Mahboobeh Safaeian, Aime R. Kreimer, and Nicolas Wentzensen. Natural Acquired Immunity Against Subsequent Genital Human Papillomavirus Infection: A Systematic Review and Meta-analysis. *Journal of Infectious Diseases*, 213(9):1444–1454, 5 2016.

[9] Steven Black, Uwe Nicolay, Timo Vesikari, Markus Knuf, Giuseppe Del Giudice, Giovanni Della Cioppa, Theodore Tsai, Ralf Clemens, and Rino Rappuoli. Hemagglutination Inhibition Antibody Titers as a Correlate of Protection for Inactivated Influenza Vaccines in Children. *The Pediatric Infectious Disease Journal*, 30(12):1081–1085, 12 2011.

[10] Jacob Bock Axelsen, Rami Yaari, Bryan T Grenfell, and Lewi Stone. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *PNAS*, 111(26):9538–9542, 2014.

[11] R Bodewes, G de Mutsert, F R M van der Klis, M Ventresca, S Wilks, D J Smith, M Koopmans, R A M Fouchier, A D M E Osterhaus, and G F Rimmelzwaan. Prevalence of antibodies against seasonal influenza A and B viruses in children in Netherlands. *Clinical and vaccine immunology : CVI*, 18(3):469–76, 3 2011.

[12] Rogier Bodewes, Albert D M E Osterhaus, and Guus F Rimmelzwaan. Targets for the induction of protective immunity against influenza a viruses. *Viruses*, 2(1):166–88, 1 2010.

[13] Ben Bolker. Likelihood and all that. In *Ecological Models and Data in R*, chapter 6, pages 169–221. Princeton University Press, 508 edition, 2007.

[14] Adrianus C M Boon, Gerrie de Mutsert, Debbie van Baarle, Derek J Smith, Alan S Lapedes, Ron A M Fouchier, Kees Sintnicolaas, Albert D M E Osterhaus, and Guus F Rimmelzwaan. Recognition of homo- and heterosubtypic variants of influenza A viruses by human CD8+ T lymphocytes. *Journal of immunology (Baltimore, Md. : 1950)*, 172(4):2453–60, 2 2004.

[15] F X Bosch, X Castellsagué, and S de Sanjosé. HPV and cervical cancer: screening or vaccination? *British Journal of Cancer*, 98(1):15–21, 2008.

[16] Ignacio G Bravo and Marta Félez-Sánchez. Papillomaviruses: viral evolution, cancer and evolutionary medicine. *Evolution, medicine, and public health*, 2015(1):32–51, 2015.

[17] M Brisson, Bénard, M Drolet, M-c Boily, H C Turner, Marc Brisson, lodie Bénard, Mlanie Drolet, Johannes A Bogaards, Iacopo Baussano, Simopekka Vänskä, Mark Jit, Marie-Claude Boily, Megan A Smith, Johannes Berkhof, Karen Canfell, Harrell W Chesson, Emily A Burger, Yoon H Choi, Birgitte Freiesleben De Blasio, Sake J De Vlas, Giorgio Guzzetta, Jan A C Hontelez, Johannes Horn, Martin R Jepsen, Jane J Kim, Fulvio Lazzarato, Suzette M Matthijsse, Rafael Mikolajczyk, Andrew Pavelyev, Matthew Pillsbury, Leigh Anne Shafer, Stephen P Tully, Hugo C Turner, Cara Usher, and Cathal Walsh. Population-level impact, herd immunity, and elimination after human papillomavirus vaccination: a systematic review and meta-analysis of predictions from transmission-dynamic models. *Funding Canadian Institutes of Health Research*, 1(16):8–17, 2016.

[18] R. M. Brotman, M. D. Shardell, P. Gajer, J. K. Tracy, J. M. Zenilman, J. Ravel, and P. E. Gravitt. Interplay Between the Temporal Dynamics of the Vaginal Microbiota and Human Papillomavirus Detection. *Journal of Infectious Diseases*, 210(11):1723–1733, 12 2014.

[19] Christopher B Buck, Patricia M Day, and Benes L Trus. The papillomavirus major capsid protein L1. *Virology*, 445(1-2):169–74, 10 2013.

[20] Caroline Buckee, Leon Danon, and Sunetra Gupta. Host community structure and the maintenance of pathogen diversity. *Proceedings. Biological sciences*, 274(1619):1715–21, 7 2007.

[21] A. N. Burchell, F. Coutlee, P.-P. Tellier, J. Hanley, and E. L. Franco. Genital Transmission of Human Papillomavirus in Recently Formed Heterosexual Couples. *Journal of Infectious Diseases*, 204(11):1723–1729, 12 2011.

[22] Fabrice Carrat, Elisabeta Vergu, Neil M. Ferguson, Magali Lemaitre, Simon Cauchemez, Steve Leach, and Alain-Jacques Valleron. Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies. *American Journal of Epidemiology*, 167(7):775–785, 4 2008.

[23] Andrew J. Caton, George G. Brownlee, Jonathan W. Yewdell, and Walter Gerhard. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell*, 31(2):417–427, 12 1982.

[24] Simon Cauchemez, Peter Horby, Annette Fox, Le Quynh Mai, Le Thi Thanh, Pham Quang Thai, Le Nguyen Minh Hoa, Nguyen Tran Hien, and Neil M. Ferguson. Influenza Infection Rates, Measurement Errors and the Interpretation of Paired Serology. *PLoS Pathogens*, 8(12):e1003061, 12 2012.

[25] A. K. Chaturvedi. Prevalence and Clustering Patterns of Human Papillomavirus Genotypes in Multiple Infections. *Cancer Epidemiology, Biomarkers & Prevention*, 14(10):2439–2445, 2005.

[26] Yao-Qing Chen, Teddy John Wohlbold, Nai-Ying Zheng, Min Huang, Yunping Huang, Karlynn E. Neu, Jiwon Lee, Hongquan Wan, Karla Thatcher Rojas, Ericka Kirkpatrick, Carole Henry, Anna-Karin E. Palm, Christopher T. Stamper, Linda Yu-Ling Lan, David J. Topham, John Treanor, Jens Wrammert, Rafi Ahmed, Maryna C. Eichelberger, George Georgiou, Florian Krammer, and Patrick C. Wilson. Influenza Infection in Humans Induces Broadly Cross-Reactive and Protective Neuraminidase-Reactive Antibodies. *Cell*, 173(2):417–429, 4 2018.

[27] Peter Chesson. Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, 31(1):343–366, 11 2000.

[28] Sarah Cobey. Pathogen evolution and the immunological niche. *Annals of the New York Academy of Sciences*, 1320(1):1–15, 7 2014.

[29] Sarah Cobey and Marc Lipsitch. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science (New York, N.Y.)*, 335(6074):1376–80, 3 2012.

[30] Sarah Cobey and Mercedes Pascual. Consequences of host heterogeneity, epitope immunodominance, and immune breadth for strain competition. *Journal of theoretical biology*, 270(1):80–7, 2 2011.

[31] Sarah Cobey, Patrick Wilson, and Frederick A Matsen. The evolution within us. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1676):20140235, 9 2015.

[32] Laurent Coudeville, Fabrice Bailleux, Benjamin Riche, Franoise Megas, Philippe Andre, and Ren Ecochard. Relationship between haemagglutination- inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model. *BMC Medical Research Methodology*, 10(18):18, 12 2010.

[33] B. J. Cowling, R. A. P. M. Perera, V. J. Fang, K.-H. Chan, W. Wai, H. C. So, D. K. W. Chu, J. Y. Wong, E. Y. Shiu, S. Ng, D. K. M. Ip, J. S. M. Peiris, and G. M. Leung. Incidence of Influenza Virus Infections in Children in Hong Kong in a 3-Year Randomized Placebo-Controlled Vaccine Study, 2009-2012. *Clinical Infectious Diseases*, 59(4):517–524, 8 2014.

[34] Benjamin J Cowling, Kwok Hung Chan, Vicky J Fang, Lincoln L H Lau, Hau Chi So, Rita O P Fung, Edward S K Ma, Alfred S K Kwong, Chi-Wai Chan, Wendy W S Tsui, Ho-Yin Ngai, Daniel W S Chu, Paco W Y Lee, Ming-Chee Chiu, Gabriel M Leung, and Joseph S M Peiris. Comparative epidemiology of pandemic and seasonal influenza A in households. *The New England journal of medicine*, 362(23):2175–2184, 6 2010.

[35] Sherry R. Crowe, Shannon C. Miller, Deborah M. Brown, Pamela S. Adams, Richard W. Dutton, Allen G. Harmsen, Frances E. Lund, Troy D. Randall, Susan L. Swain, and David L. Woodland. Uneven distribution of MHC class II epitopes within the influenza virus. *Vaccine*, 24(4):457–467, 1 2006.

[36] Patrcia Savio de Araujo Souza, Laura Sichero, and Paulo Cesar Maciag. HPV variants and HLA polymorphisms: the role of variability on the risk of cervical cancer. *Future Oncology*, 5(3):359–370, 4 2009.

[37] Ethel-Michele de Villiers, Claude Fauquet, Thomas R Broker, Hans-Ulrich Bernard, and Harald zur Hausen. Classification of papillomaviruses. *Virology*, 324(1):17–27, 6 2004.

[38] A Doucet, N de Freitas, and NJ Gordon. Sequential Monte Carlo Methods in Practice. Series Statistics For Engineering and Information Science, 2001.

[39] E Draper, SL Bissett, R Howell-Jones, and D Edwards. Neutralization of non-vaccine human papillomavirus pseudoviruses from the A7 and A9 species groups by bivalent HPV vaccine sera. *Vaccine*, 29(47):8585–8590, 2011.

[40] Xiangjun Du, Aaron A King, Robert J Woods, and Mercedes Pascual. Evolution-informed forecasting of seasonal influenza A (H3N2). *Science translational medicine*, 9(413):eaan5325, 10 2017.

[41] Andrew J. Dunning. A model for immunological correlates of protection. *Statistics in Medicine*, 25(9):1485–1497, 5 2006.

[42] David P Durham, Martial L Ndeffo-Mbah, Laura A Skrip, Forrest K Jones, Chris T Bauch, and Alison P Galvani. National- and state-level impact and cost-effectiveness of nonavalent HPV vaccination in the United States. *Proceedings of the National Academy of Sciences*, 113(18):5107–5112, 2016.

[43] David P. Durham, Eric M. Poolman, Yoko Ibuka, Jeffrey P. Townsend, and Alison P. Galvani. Reevaluation of epidemiological data demonstrates that it is consistent with cross-immunity among human papillomavirus types. *Journal of Infectious Diseases*, 206(8):1291–1298, 2012.

[44] Ken T D Eames and Matt J Keeling. Contact tracing and disease control. *Proceedings. Biological sciences*, 270(1533):2565–71, 12 2003.

[45] Maryna C. Eichelberger and Hongquan Wan. Influenza Neuraminidase as a Vaccine Antigen. volume 386, pages 275–299. Springer Verlag, 6 2014.

[46] Neil M. Ferguson, Alison P. Galvani, and Robin M. Bush. Ecological and immunological determinants of influenza evolution. *Nature*, 422(6930):428–433, 3 2003.

[47] Arietta E. Fleming-Davies, Vanja Dukic, Viggo Andreasen, and Greg Dwyer. Effects of host heterogeneity on pathogen diversity and evolution. *Ecology Letters*, 18(11):1252–1261, 11 2015.

[48] J M Fonville, S H Wilks, S L James, A Fox, M Ventresca, M Aban, L Xue, T C Jones, N M H Le, Q T Pham, N D Tran, Y Wong, A Mosterin, L C Katzelnick, D Labonte, T T Le, G van der Net, E Skepner, C A Russell, T D Kaplan, G F Rimmelzwaan, N Masurel, J C de Jong, A Palache, W E P Beyer, Q M Le, T H Nguyen, H F L Wertheim, A C Hurt, A D M E Osterhaus, I G Barr, R A M Fouchier, P W Horby, and D J Smith. Antibody landscapes after influenza virus infection or vaccination. *Science (New York, N.Y.)*, 346(6212):996–1000, 11 2014.

[49] Judith M. Fonville, Pieter L. A. Fraaij, Gerrie de Mutsert, Samuel H. Wilks, Ruud van Beek, Ron A. M. Fouchier, and Guus F. Rimmelzwaan. Antigenic Maps of Influenza A(H3N2) Produced With Human Antisera Obtained After Primary Infection. *Journal of Infectious Diseases*, 213(1):31–38, 1 2016.

[50] Annette Fox, Le Quynh Mai, Le Thi Thanh, Marcel Wolbers, Nguyen Le Khanh Hang, Pham Quang Thai, Nguyen Thi Thu Yen, Le Nguyen Minh Hoa, Juliet E. Bryant, Tran Nhu Duong, Dang Dinh Thoang, Ian G. Barr, Heiman Wertheim, Jeremy Farrar, Nguyen Tran Hien, and Peter Horby. Hemagglutination inhibiting antibodies and protection against seasonal and pandemic influenza infection. *Journal of Infection*, 70(2):187–196, 2 2015.

[51] Thomas Francis and Jr. On the Doctrine of Original Antigenic Sin. *Proceedings of the American Philosophical Society*, 104:572–578.

[52] G Freeman, R A P M Perera, E Ngan, V J Fang, S Cauchemez, D K M Ip, J S M Peiris, and B J Cowling. Quantifying homologous and heterologous antibody titre rises after influenza virus infection. *Epidemiology and infection*, 144(11):2306–16, 8 2016.

[53] Tsung Chieh Fu, Long Fu Xi, Ayaka Hulbert, James P. Hughes, Qinghua Feng, Stephen M. Schwartz, Stephen E. Hawes, Laura A. Koutsky, and Rachel L. Winer. Short-term natural history of high-risk human papillomavirus infection in mid-adult women sampled monthly. *International Journal of Cancer*, 137(10):2432–2442, 2015.

[54] Alain Gagnon, Enrique Acosta, Stacey Hallman, Robert Bourbeau, Lisa Y Dillon, Nadine Ouellette, David J D Earn, D Ann Herring, Kris Inwood, Joaquin Madrenas, and Matthew S Miller. Pandemic Paradox: Early Life H2N2 Pandemic Influenza Infection Enhanced Susceptibility to Death during the 2009 H1N1 Pandemic. *mBio*, 9(1):02091–17, 1 2018.

[55] G. F. Gause. Experimental studies on the struggle for existence. I. Mixed population of two species of yeast. *Journal of Experimental Biology*, 9:389–402, 1932.

[56] Anna R Giuliano, Eduardo Lazcano, Luisa Lina Villa, Roberto Flores, Jorge Salmeron, Ji-Hyun Lee, Mary Papenfuss, Martha Abrahamsen, Maria Luiza Baggio, Roberto Silva, and Manuel Quiterio. Circumcision and sexual behavior: factors independently associated with human papillomavirus detection among men in the HIM study. *International Journal of Cancer*, 124(6):1251–7, 3 2009.

[57] Anna R Giuliano, Eduardo Lazcano-Ponce, Luisa L Villa, Roberto Flores, Jorge Salmeron, Ji-Hyun Lee, Mary R Papenfuss, Martha Abrahamsen, Emily Jolles, Carrie M Nielson, Maria Luisa Baggio, Roberto Silva, and Manuel Quiterio. The human papillomavirus infection in men study: human papillomavirus prevalence and type distribution among men residing in Brazil, Mexico, and the United States. *Cancer Epidemiology, Biomarkers & Prevention*, 17(8):2036–43, 8 2008.

[58] Anna R Giuliano, Ji-Hyun Lee, William Fulp, Luisa L Villa, Eduardo Lazcano, Mary R Papenfuss, Martha Abrahamsen, Jorge Salmeron, Gabriella M Anic, Dana E Rollison, and Danelle Smith. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. *Lancet*, 377(9769):932–40, 3 2011.

[59] Anna R. Giuliano, Alan G. Nyitray, Aim Ee, R Kreimer, Christine M Pierce Campbell, Marc T. Goodman, Staci L. Sudenga, Joseph Monsonego, and Silvia Franceschi. EUROGIN 2014 roadmap: Differences in human papillomavirus infection natural history, transmission and human papillomavirus-related cancer incidence by gender and anatomic site of infection. *International Journal of Cancer*, 136(12):2752–2760, 2015.

[60] Marc T Goodman, Yurii B Shvetsov, Katharine McDuffie, Lynne R Wilkens, Xuemei Zhu, Pamela J Thompson, Lily Ning, Jeffrey Killeen, Lori Kamemoto, and Brenda Y Hernandez. Sequential Acquisition of Human Papillomavirus Infection of the Anus and Cervix: The Hawaii HPV Cohort Study. 2010(9):1331–9, 5 2010.

[61] Katelyn M Gostic, Monique Ambrose, Michael Worobey, and James O Lloyd-Smith. Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. *Science (New York, N.Y.)*, 354(6313):722–726, 11 2016.

[62] Emma Grant, Chao Wu, Kok-Fei Chan, Sidonia Eckle, Mandvi Bharadwaj, Quan Ming Zou, Katherine Kedzierska, and Weisan Chen. Nucleoprotein of influenza A virus is a major target of immunodominant CD8+ T-cell responses. *Immunology and cell biology*, 91(2):184–94, 2 2013.

[63] P E Gravitt, C L Peyton, R J Apple, and C M Wheeler. Genotyping of 27 human papillomavirus types by using L1 consensus PCR products by a single-hybridization, reverse line blot detection method. *Journal of Clinical Microbiology*, 36(10):3020–7, 10 1998.

[64] Patti E Gravitt. Evidence and impact of human papillomavirus latency. *Open Virology Journal*, 6:198–203, 2012.

[65] Ronald H Gray, David Serwadda, Xiangrong Kong, Frederick Makumbi, Godfrey Kigozi, Patti E Gravitt, Stephen Watya, Fred Nalugoda, Victor Ssempijja, Aaron A R Tobian, Noah Kiwanuka, Lawrence H Moulton, Nelson K Sewankambo, Steven J Reynolds, Thomas C Quinn, Boaz Iga, Oliver Laeyendecker, Amy E Oliver, and Maria J Wawer. Male circumcision decreases acquisition and increases clearance of high-risk human papillomavirus in HIV-negative men: a randomized trial in Rakai, Uganda. *Journal of Infectious Diseases*, 201(10):1455–62, 5 2010.

[66] Kristie M Grebe, Jonathan W Yewdell, and Jack R Bennink. Heterosubtypic immunity to influenza A virus: where do we stand? *Microbes and infection*, 10(9):1024–9, 7 2008.

[67] Bryan T Grenfell, Oliver G Pybus, Julia R Gog, James L N Wood, Janet M Daly, Jenny A Mumford, and Edward C Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science (New York, N.Y.)*, 303(5656):327–32, 1 2004.

[68] S. Gupta and K.P. Day. A strain theory of malaria transmission. *Parasitology Today*, 10(12):476–481, 1 1994.

[69] S Gupta, N Ferguson, and R Anderson. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science (New York, N.Y.)*, 280(5365):912–5, 5 1998.

[70] Sunetra Gupta, Martin C.J. Maiden, Ian M. Feavers, Sean Nee, Robert M. May, and Roy M. Anderson. The maintenance of strain structure in populations of recombining infectious agents. *Nature Medicine*, 2(4):437–442, 4 1996.

[71] Jasmine J. Han, Thomas H. Beltran, John W. Song, John Klaric, and Y. Sammy Choi. Prevalence of Genital Human Papillomavirus Infection and Human Papillomavirus Vaccination Rates Among US Adult Men. *JAMA Oncology*, 3(6):810–816, 2017.

[72] Tomer Hertz, Christine M Oshansky, Philippa L Roddam, John P Devincenzo, Miguela A Caniza, Nebojsa Jojic, Simon Mallal, Elizabeth Phillips, Ian James, Elizabeth Halloran, Paul G Thomas, Lawrence Corey, and Peter C Doherty. HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection.

[73] J. HilleRisLambers, P.B. Adler, W.S. Harpole, J.M. Levine, and M.M. Mayfield. Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):227–248, 12 2012.

[74] D Hobson, R L Curry, A S Beare, and A Ward-Gardner. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. *The Journal of hygiene*, 70(4):767–77, 12 1972.

[75] Dawn M Holman, Vicki Benard, Katherine B Roland, Meg Watson, Nicole Liddon, and Shannon Stokley. Barriers to human papillomavirus vaccination among US adolescents: a systematic review of the literature. *JAMA pediatrics*, 168(1):76–82, 1 2014.

[76] Clifford M. Hurvich and Chih-Ling Tsai. Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76(2):297, 6 1989.

[77] G E Hutchinson. Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *Source: The American Naturalist*, 93(870):145–159.

[78] G Evelyn Hutchinson. Concluding Remarks. *Cold Spring Harb Symp Quant Biol*, 22:415–427, 1957.

[79] E L Ionides, C Breto, J Park, R A Smith, and A A King. Monte Carlo profile confidence intervals. 2017.

[80] Edward L Ionides, C Bretó, and Aaron A King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443, 12 2006.

[81] Robert M Jacobson, Diane E Grill, Ann L Oberg, Pritish K Tosh, Inna G Ovsyannikova, and Gregory A Poland. Profiles of influenza A/H1N1 vaccine response using hemagglutination-inhibition titers Robert. *Human Vaccines & Immunotherapeutics*, 11(4):961–969, 4 2015.

[82] Charles. Janeway. *Immunobiology 5 : the immune system in health and disease*. Garland Pub, 2001.

[83] K E Jensen, F M Davenport, A V Hennessey, and T Francis. Characterization of influenza antibodies by serum absorption. *The Journal of experimental medicine*, 104(2):199–209, 8 1956.

[84] Elmar A Joura, Suzanne M Garland, Jorma Paavonen, Daron G Ferris, Gonzalo Perez, Kevin A Ault, Warner K Huh, Heather L Sings, Margaret K. James, and Richard M Haupt. Effect of the human papillomavirus (HPV) quadrivalent vaccine in a subgroup of women with cervical and vulvar disease: retrospective pooled analysis of trial data. *British Medical Journal*, 344:e1401, 2012.

[85] Jessica a Kahn, Darron R Brown, Lili Ding, Lea E Widdice, Marcia L Shew, Susan Glynn, and David I Bernstein. Vaccine-type human papillomavirus and evidence of herd protection after vaccine introduction. *Pediatrics*, 130(2):249–56, 8 2012.

[86] Matt J Keeling and Rohani Pejman. Formulating the Deterministic SIR Model. In *Modeling Infectious Diseases in Humans and Animals*, chapter 2, pages 15–52. Princeton University Press, Princeton, NJ, 2 edition, 2008.

[87] Troy J. Kemp, Allan Hildesheim, Mahboobeh Safaeian, Joseph G. Dauner, Yuanji Pan, Carolina Porras, John T. Schiller, Douglas R. Lowy, Rolando Herrero, and Ligia A. Pinto. HPV16/18 L1 VLP vaccine induces cross-neutralizing antibodies that may mediate cross-protection. *Vaccine*, 29(11):2011–2014, 2011.

[88] Jin Hyang Kim, Ioanna Skountzou, Richard Compans, and Joshy Jacob. Original antigenic sin responses to influenza viruses. *Journal of immunology (Baltimore, Md. : 1950)*, 183(5):3294–301, 9 2009.

[89] J.W. Kirchner and B.A. Roy. *Evolutionary ecology research.*, volume 4. Evolutionary Ecology, 1999.

[90] Istvan Z. Kiss, Darren M. Green, and Rowland R. Kao. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Mathematical Biosciences*, 203(1):124–136, 9 2006.

[91] K. Koelle, S. Cobey, B. Grenfell, and M. Pascual. Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans. *Science*, 314(5807):1898–1903, 12 2006.

[92] J.H.C.M. Kreijtz, R.A.M. Fouchier, and G.F. Rimmelzwaan. Immune responses to influenza virus infection. *Virus Research*, 162(1-2):19–30, 12 2011.

[93] Adam Kucharski, Justin Lessler, Derek Cummings, and Steven Riley. Timescales of influenza A/H3N2 antibody dynamics. *bioRxiv*, page 183111, 11 2017.

[94] Adam J. Kucharski, Kin O. Kwok, Vivian W. I. Wei, Benjamin J. Cowling, Jonathan M. Read, Justin Lessler, Derek A. Cummings, and Steven Riley. The Contribution of Social Behaviour to the Transmission of Influenza A in a Human Population. *PLoS Pathogens*, 10(6):e1004206, 6 2014.

[95] Adam J. Kucharski, Justin Lessler, Jonathan M. Read, Huachen Zhu, Chao Qiang Jiang, Yi Guan, Derek A. T. Cummings, and Steven Riley. Estimating the Life Course of Influenza A(H3N2) Antibody Responses from Cross-Sectional Data. *PLOS Biology*, 13(3):e1002082, 3 2015.

[96] Nicole L. La Gruta and Stephen J. Turner. T cell mediated immunity to influenza: mechanisms of viral control. *Trends in Immunology*, 35(8):396–402, 8 2014.

[97] Nathaniel D Lambert, Inna G Ovsyannikova, V Shane Pankratz, Robert M Jacobson, and Gregory A Poland. Understanding the immune response to seasonal influenza vaccination in older adults: a systems biology approach. *Expert review of vaccines*, 11(8):985–94, 8 2012.

[98] J Lee, V Taneja, and R Vassallo. Cigarette smoking and inflammation: cellular and molecular mechanisms. *Journal of Dental Research*, 91(2):142–149, 2 2012.

[99] P. Lemieux-Mellouki, M. Drolet, J. Brisson, E. L. Franco, M.-C. Boily, I. Baussano, and M. Brisson. Assortative mixing as a source of bias in epidemiological studies of sexually transmitted infections: the case of smoking and human papillomavirus. *Epidemiology and Infection*, 144(07):1–10, 5 2015.

[100] Justin Lessler, Steven Riley, Jonathan M. Read, Shuying Wang, Huachen Zhu, Gavin J. D. Smith, Yi Guan, Chao Qiang Jiang, and Derek A. T. Cummings. Evidence for Antigenic Seniority in Influenza A (H3N2) Antibody Responses in Southern China. *PLoS Pathogens*, 8(7):e1002802, 7 2012.

[101] Marc Lipsitch, Caroline Colijn, Ted Cohen, William P Hanage, and Christophe Fraser. No coexistence for free: neutral null models for multistrain pathogens. *Epidemics*, 1(1):2–13, 3 2009.

[102] Marc Lipsitch and Justin J O'Hagan. Patterns of antigenic diversity and the mechanisms that maintain them. *Journal of The Royal Society Interface*, 4(16):787–802, 2007.

[103] Su-Hsun Liu, Rebecca M. Brotman, Jonathan M. Zenilman, Patti E. Gravitt, and Derek A. T. Cummings. Menstrual Cycle and Detectable Human Papillomavirus in Reproductive-age Women: A Time Series Study. *The Journal of Infectious Diseases*, 208(9):1404–1415, 11 2013.

[104] B. Lu, R. P. Viscidi, Y. Wu, J.-H. Lee, A. G. Nyitray, L. L. Villa, E. Lazcano-Ponce, R. J. C. da Silva, M. L. Baggio, M. Quiterio, J. Salmeron, D. C. Smith, M. E. Abrahamsen, M. R. Papenfuss, H. G. Stockwell, and A. R. Giuliano. Prevalent Serum

Antibody Is Not a Marker of Immune Protection against Acquisition of Oncogenic HPV16 in Men. *Cancer Research*, 72(3):676–685, 2 2012.

[105] Beibei Lu, Yougui Wu, Carrie Nielson, Roberto Flores, Martha Abrahamsen, Mary Papenfuss, Robin Harris, and Anna Giuliano. Factors Associated with Acquisition and Clearance of Human Papillomavirus Infection in a Cohort of US Men: A Prospective Study. *Journal of Infectious Diseases*, 199(3):362–371, 2 2009.

[106] Yingfei Ma, Ramana Madupu, Ulas Karaoz, Carlos W Nossa, Liying Yang, Shibu Yooseph, Patrick S Yachimski, Eoin L Brodie, Karen E Nelson, and Zhiheng Pei. Human papillomavirus community in healthy persons, defined by metagenomics analysis of human microbiome project shotgun sequencing data sets. *Journal of Virology*, 88(9):4786–97, 5 2014.

[107] M. J. Mackinnon and K. Marsh. The Selection Landscape of Malaria Parasites. *Science*, 328(5980):866–871, 2010.

[108] Gareth Adam Maglennon, Pauline McIntosh, and John Doorbar. Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology*, 414(2), 6 2011.

[109] Tala Malagón, Philippe Lemieux-Mellouki, Jean-Franois Laprise, and Marc Brisson. Bias Due to Correlation Between Times-at-Risk for Infection in Epidemiologic Studies Measuring Biological Interactions Between Sexually Transmitted Infections: A Case Study Using Human Papillomavirus Type Interactions. *American Journal of Epidemiology*, 184(12):873–883, 12 2016.

[110] Lauri E. Markowitz, Susan Hariri, Carol Lin, Eileen F. Dunne, Martin Steinau, Geraldine McQuillan, and Elizabeth R. Unger. Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003-2010. *Journal of Infectious Diseases*, 208(3):385–393, 8 2013.

[111] Laurie Markowitz, Maya Sternberg, Eileen F. Dunne, Geraldine McQuillan, and Elizabeth R Unger. Seroprevalence of Human Papillomavirus Types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003-2004. *Journal of Infectious Diseases*, 200(7):1059–1067, 10 2009.

[112] V. Marmara, A. Cook, and A. Kleczkowski. Estimation of force of infection based on different epidemiological proxies: 2009/2010 Influenza epidemic in Malta. *Epidemics*, 9:52–61, 12 2014.

[113] Suzette M. Matthijsse, Jan A C Hontelez, Steffie K. Naber, Joost van Rosmalen, Kirsten Rozemeijer, Corine Penning, Roel Bakker, Marjolein van Ballegooijen, Inge M C M de Kok, and Sake J. de Vlas. The estimated impact of natural immunity on the effectiveness of human papillomavirus vaccination. *Vaccine*, 33(41):5357–5364, 2015.

[114] Matthew S Miller, Thomas J Gardner, Florian Krammer, Lauren C Aguado, Domenico Tortorella, Christopher F Basler, and Peter Palese. Neutralizing antibodies against previously encountered influenza virus strains increase over time: a longitudinal analysis. *Science translational medicine*, 5(198):198ra107, 8 2013.

[115] Arnold S. Monto, Joshua G. Petrie, Rachel T. Cross, Emileigh Johnson, Merry Liu, Weimin Zhong, Min Levine, Jacqueline M. Katz, and Suzanne E. Ohmit. Antibody to Influenza Virus Neuraminidase: An Independent Correlate of Protection. *Journal of Infectious Diseases*, 212(8):1191–1199, 10 2015.

[116] Anna-Barbara Moscicki. HPV Vaccines: today and in the Future. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, 43(4 Suppl):26–40, 10 2008.

[117] Anna-Barbara Mosciki, Mark Schiffman, Ann Burchell, Ginsea Alberdo, Anna R. Giuliano, Mark T Marc T. Goodman, Susanne K. Kjaer, Joel Palefsky, AB Anna-Barbara Moscicki, Mark Schiffman, Ann Burchell, Ginesa Albero, Anna R. Giuliano, Mark T Marc T. Goodman, Susanne K. Kjaer, and Joel Palefsky. Updating the natural history of human papillomavirus and anogenital cancers. *Vaccine*, 30(Supplement 5):F24–F33, 11 2012.

[118] Carmen La Murall, Chris T Bauch, and Troy Day. Could the human papillomavirus vaccines drive virulence evolution? *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1798):20141069, 1 2015.

[119] Carmen La Murall, Kevin S. McCann, and Chris T. Bauch. Revising ecological assumptions about Human papillomavirus interactions and type replacement. *Journal of Theoretical Biology*, 350:98–109, 2014.

[120] Raffael Nachbagauer, Angela Choi, Ruvim Izikson, Manon M. Cox, Peter Palese, and Florian Krammer. Age Dependence and Isotype Specificity of Influenza Virus Hemagglutinin Stalk-Reactive Antibodies in Humans. *mBio*, 7(1):01996–15, 2016.

[121] Mayumi Nakagawa, DanielP. Stites, Sandeep Patel, Sepideh Farhat, Mark Scott, NancyK. Hills, JoelM. Palefsky, and AnnaBarbara Moscicki. Persistence of Human Papillomavirus Type 16 Infection Is Associated with Lack of Cytotoxic T Lymphocyte Response to the E6 Antigens. *Journal of Infectious Diseases*, 182(2):595–598, 8 2000.

[122] Jennifer L. Nayak, Katherine A. Richards, Francisco A. Chaves, and Andrea J. Sant. Analyses of the Specificity of CD4 T Cells During the Primary Immune Response to Influenza Virus Reveals Dramatic MHC-Linked Asymmetries in Reactivity to Individual Viral Proteins. *Viral Immunology*, 23(2):169–180, 4 2010.

[123] Sophia Ng, Dennis K. M. Ip, Vicky J. Fang, Kwok-Hung Chan, Susan S. Chiu, Gabriel M. Leung, J. S. Malik Peiris, and Benjamin J. Cowling. The Effect of Age and Recent Influenza Vaccination History on the Immunogenicity and Efficacy of

200910 Seasonal Trivalent Inactivated Influenza Vaccination in Children. *PLoS ONE*, 8(3):e59077, 3 2013.

[124] Huan H. Nguyen, Zina Moldoveanu, Miroslav J. Novak, Frederik W. van Ginkel, Elisabeth Ban, Hiroshi Kiyono, Jerry R. McGhee, and Jiri Mestecky. Heterosubtypic Immunity to Lethal Influenza A Virus Infection Is Associated with Virus-Specific CD8+Cytotoxic T Lymphocyte Responses Induced in Mucosa-Associated Tissues. *Virology*, 254(1):50–60, 2 1999.

[125] Alan G. Nyitray, Mihyun Chang, Luisa L. Villa, Roberto J. Carvalho da Silva, Maria Luiza Baggio, Martha Abrahamsen, Mary Papenfuss, Manuel Quiterio, Jorge Salmerón, Eduardo Lazcano-Ponce, and Anna R. Giuliano. The Natural History of Genital Human Papillomavirus Among HIV-Negative Men Having Sex With Men and Men Having Sex With Women. *Journal of Infectious Diseases*, 212(2):202–212, 7 2015.

[126] Alan G. Nyitray, Jongphil Kim, Chiu-Hsieh Hsieh Hsu, Mary Papenfuss, Luisa Villa, Eduardo Lazcano-Ponce, and Anna R. Giuliano. Test-retest reliability of a sexual behavior interview for men residing in Brazil, Mexico, and the United States. *American Journal of Epidemiology*, 170(8):965–974, 10 2009.

[127] OECD. Influenza vaccination rates. Technical report, OECD, 2018.

[128] SE Ohmit, JG Petrie, and RT Cross. Influenza hemagglutination-inhibition antibody titer as a correlate of vaccine-induced protection. *The Journal of Infectious Diseases*, 204(12):1879–1885, 2011.

[129] Sven-Eric Olsson, Susanne K Kjaer, Kristjn Sigurdsson, Ole-Erik Iversen, Mauricio Hernandez-Avila, Cosette M Wheeler, Gonzalo Perez, Darron R Brown, Laura A Koutsky, Eng Hseon Tay, Patricia García, Kevin A Ault, Suzanne M Garland, Sepp Leodolter, Grace W K Tang, Daron G Ferris, Jorma Paavonen, Matti Lehtinen, Marc Steben, F Xavier Bosch, Joakim Dillner, Elmar A Joura, Slawomir Majewski, Nubia Muñoz, Evan R Myers, Luisa L Villa, Frank J Taddeo, Christine Roberts, Amha Tadesse, Janine Bryan, Roger Maansson, Scott Vuocolo, Teresa M Hesley, Alfred Saah, Eliav Barr, and Richard M Haupt. Evaluation of quadrivalent HPV 6/11/16/18 vaccine efficacy against cervical and anogenital disease in subjects with serological evidence of prior vaccine type HPV infection. *Human vaccines*, 5(10):696–704, 10 2009.

[130] Paul A. Orlando, Robert A. Gatenby, Anna R. Giuliano, and Joel S. Brown. Evolutionary ecology of human papillomavirus: Trade-offs, coexistence, and origins of high-risk and low-risk types. *Journal of Infectious Diseases*, 205(2):272–279, 1 2012.

[131] Erik E. Osnas and Andrew P. Dobson. Evolution of virulence in heterogeneous host communities under multiple trade-offs. *Evolution*, 66(2):391–401, 2 2012.

[132] S. J. Pamnani, S. L. Sudenga, R. Viscidi, D. E. Rollison, B. N. Torres, D. J. Ingles, M. Abrahamsen, L. L. Villa, E. Lazcano-Ponce, J. Salmeron, M. Quiterio, Y. Huang,

A. Borenstein, and A. R. Giuliano. Impact of Serum Antibodies to HPV Serotypes 6, 11, 16, and 18 to Risks of Subsequent Genital HPV Infections in Men: The HIM Study. *Cancer Research*, 76(20):6066–6075, 10 2016.

[133] Jae-Keun Park, Alison Han, Lindsay Czajkowski, Susan Reed, Rani Athota, Tyler Bristol, Luz Angela Rosas, Adriana Cervantes-Medina, Jeffery K Taubenberger, and Matthew J Memoli. Evaluation of Preexisting Anti-Hemagglutinin Stalk Antibody as a Correlate of Protection in a Healthy Volunteer Challenge with Influenza A/H1N1pdm Virus. *mBio*, 9(1):02284–17, 1 2018.

[134] Jeffrey M Partridge, James P Hughes, Qinghua Feng, Rachel L Winer, Bethany A Weaver, Long-Fu Xi, Michael E Stern, Shu-Kuang Lee, Sandra F O'Reilly, Stephen E Hawes, Nancy B Kiviat, and Laura A Koutsky. Genital human papillomavirus infection in men: incidence and risk factors in a cohort of university students. *Journal of Infectious Diseases*, 196(8):1128–1136, 10 2007.

[135] Jo-ann S Passmore and Anna-lise Williamson. Host Immune Responses Associated with Clearance or Persistence of Human Papillomavirus Infections. *Current Obstetrics and Gynecology Reports*, 5(3):177–188, 2016.

[136] Bridget S Penman, Ben Ashby, Caroline O Buckee, and Sunetra Gupta. Pathogen selection drives nonoverlapping associations between HLA loci. *Proceedings of the National Academy of Sciences of the United States of America*, 110(48):19645–50, 11 2013.

[137] JG Petrie, SE Ohmit, and E Johnson. Efficacy studies of influenza vaccines: effect of end points used and characteristics of vaccine failures. *Journal of Infectious Diseases*, 203(9):1309–1315, 2011.

[138] S. J. Piersma, E. S. Jordanova, M. I.E. van Poelgeest, K. M.C. Kwappenberg, J. M. van der Hulst, J. W. Drijfhout, C. J.M. Melief, G. G. Kenter, G. J. Fleuren, R. Offringa, and S. H. van der Burg. High Number of Intraepithelial CD8+ Tumor-Infiltrating Lymphocytes Is Associated with the Absence of Lymph Node Metastases in Patients with Large Early-Stage Cervical Cancer. *Cancer Research*, 67(1):354–361, 1 2007.

[139] Eric M. Poolman, Elamin H. Elbasha, and Alison P. Galvani. Vaccination and the evolutionary ecology of human papillomavirus. *Vaccine*, 26(Supplement 3):25–30, 2008.

[140] Timothy J Powell, Tara Strutt, Joyce Reome, Joseph A Hollenbaugh, Alan D Roberts, David L Woodland, Susan L Swain, and Richard W Dutton. Priming with cold-adapted influenza A does not prevent infection but elicits long-lived protection against supralethal challenge with heterosubtypic virus. *Journal of immunology (Baltimore, Md. : 1950)*, 178(2):1030–8, 1 2007.

[141] Jung Pu Hsu, Xiahong Zhao, Mark I-Cheng Chen, Alex R Cook, Vernon Lee, Wei Yen Lim, Linda Tan, Ian G Barr, Lili Jiang, Chyi Lin Tan, Meng Chee Phoon, Lin Cui, Raymond Lin, Yee Sin Leo, and Vincent T Chow. Rate of decline of antibody

titers to pandemic influenza A (H1N1-2009) by hemagglutination inhibition and virus microneutralization assays in a cohort of seroconverting adults in Singapore. *BMC Infectious Diseases*, 14(1):414, 2014.

[142] Sylvia L Ranjeva, Edward B Baskerville, Vanja Dukic, Luisa L Villa, Eduardo Lazcano-Ponce, Anna R Giuliano, Greg Dwyer, and Sarah Cobey. Recurring infection with ecologically distinct HPV types can explain high prevalence and diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(51):13573–13578, 12 2017.

[143] Jonathan M Read and Matt J Keeling. Disease evolution on networks: the role of contact structure. *Proceedings. Biological sciences*, 270(1516):699–708, 4 2003.

[144] Roland R. Regoes, Martin A. Nowak, and Sebastian Bonhoeffer. Evolution of virulence in a heterogeneous host population. *Evolution*, 54(1):64–71, 2 2000.

[145] E. O. Romero-Severson, E. Volz, J. S. Koopman, T. Leitner, and E. L. Ionides. Practice of Epidemiology Dynamic Variation in Sexual Contact Rates in a Cohort of HIV-Negative Gay Men. *American Journal of Epidemiology*, 182(3):255–262, 8 2015.

[146] Anne F. Rositch, Michael G. Hudgens, Danielle M. Backes, Stephen Moses, Kawango Agot, Edith Nyagaya, Peter J F Snijders, Chris J L M Meijer, Robert C. Bailey, and Jennifer S. Smith. Vaccine-relevant human papillomavirus (HPV) infections and future acquisition of high-risk HPV types in men. *Journal of Infectious Diseases*, 206(5):669–677, 2012.

[147] Marc D Ryser, Patti E Gravitt, and Evan R Myers. Mechanistic Mathematical Models: An Underused Platform for HPV Research. *Papillomavirus Research*, 3:46–49, 2017.

[148] Marc D. Ryser, Evan R. Myers, and Rick Durrett. HPV Clearance and the Neglected Role of Stochasticity. *PLOS Computational Biology*, 11(3):e1004113, 2015.

[149] Mahboobeh Safaeian, Carolina Porras, Mark Schiffman, Ana Cecilia Rodriguez, Sholom Wacholder, Paula Gonzalez, Wim Quint, Leen-Jan van Doorn, Mark E Sherman, Valrie Xhenseval, Rolando Herrero, Allan Hildesheim, and for the Costa Rican Vaccine Trial Costa Rican Vaccine Trial Group. Epidemiological study of anti-HPV16/18 seropositivity and subsequent risk of HPV16 and -18 infections. *Journal of the National Cancer Institute*, 102(21):1653–62, 11 2010.

[150] A Sauerbrei, T Langenhan, A Brandstädt, R Schmidt-Ott, A Krumbholz, H Girschick, H Huppertz, P Kaiser, J Liese, A Streng, T Niehues, J Peters, A Sauerbrey, H Schroten, T Tenenbaum, S Wirth, and P Wutzler. Prevalence of antibodies against influenza A and B viruses in children in Germany, 2008 to 2010. *Eurosurveillance*, 19(5):20687, 2 2014.

[151] Andreas Sauerbrei, R. Schmidt-Ott, H. Hoyer, and P. Wutzler. Seroprevalence of influenza A and B in German infants and adolescents. *Medical Microbiology and Immunology*, 198(2):93–101, 5 2009.

[152] M Scott, M Nakagawa, and A B Moscicki. Cell-mediated immune response to human papillomavirus infection. *Clinical and Diagnostic Laboratory Immunology*, 8(2):209–20, 3 2001.

[153] Marcia L. Shew, Aaron C. Ermel, Yan Tong, Wanzhu Tu, Brahim Qadadri, and Darron R. Brown. Episodic detection of human papillomavirus within a longitudinal cohort of young women. *Journal of Medical Virology*, 87(12):2122–2129, 12 2015.

[154] C Sonnex, S Strauss, and J J Gray. Detection of human papillomavirus DNA on the fingers of patients with genital warts. *Sexually Transmitted Infections*, 75(5):317–9, 10 1999.

[155] Saranya Sridhar, Shaima Begom, Alison Bermingham, Katja Hoschler, Walt Adamson, William Carman, Thomas Bean, Wendy Barclay, Jonathan J Deeks, and Ajit Lalvani. Cellular immune correlates of protection against symptomatic pandemic influenza. *Nature Medicine*, 19(10):1305–1312, 10 2013.

[156] Margaret Stanley. Immune responses to human papillomavirus. *Vaccine*, 24(Supplement 1):16–22, 2006.

[157] H. D. Strickler, R. D. Burk, M. Fazzari, K. Anastos, H. Minkoff, L. S. Massad, C. Hall, M. Bacon, A. M. Levine, D. H. Watts, M. J. Silverberg, X. Xue, N. F. Schlecht, S. Melnick, and J. M. Palefsky. Natural History and Possible Reactivation of Human Papillomavirus in Human Immunodeficiency Virus-Positive Women. *Journal of the National Cancer Institute*, 97(8):577–586, 4 2005.

[158] K M Sullivan, A S Monto, and I M Longini. Estimates of the US health impact of influenza. *American journal of public health*, 83(12):1712–6, 12 1993.

[159] E I Svare, S K Kjaer, A M Worm, A Østerlind, C J L M Meijer, and A J C van den Brule. Risk factors for genital HPV DNA in men resemble those found in women: a study of male attendees at a Danish STD clinic. *Sexually Transmitted Infections*, 78(3):215–218, 2002.

[160] Barbara S Taylor, Magdalena E Sobieszczyk, Francine E McCutchan, and Scott M Hammer. The challenge of HIV-1 subtype diversity. *New England Journal of Medicine*, 358(15):1590–602, 4 2008.

[161] Regan N. Theiler, Sherry L. Farr, John M. Karon, Pangaja Paramsothy, Raphael Viscidi, Ann Duerr, Susan Cu-Uvin, Jack Sobel, Keerti Shah, Robert S. Klein, and Denise J. Jamieson. High-Risk Human Papillomavirus Reactivation in Human Immunodeficiency Virus Infected Women. *Obstetrics & Gynecology*, 115(6):1150–1158, 6 2010.

[162] David Tilman. Resources: A Graphical-Mechanistic Approach to Competition and Predation. *The American Naturalist*, 116(3):362–393, 9 1980.

[163] Jerome I Tokars, Sonja J Olsen, and Carrie Reed. Seasonal Incidence of Symptomatic Influenza in the United States. *Clinical Infectious Diseases*, 66(10):1511–1518, 5 2018.

[164] Tim K. Tsang, Simon Cauchemez, Ranawaka A. P. M. Perera, Guy Freeman, Vicky J. Fang, Dennis K. M. Ip, Gabriel M. Leung, Joseph Sriyal Malik Peiris, and Benjamin J. Cowling. Association Between Antibody Titers and Protection Against Influenza Virus Infection Within Households. *The Journal of Infectious Diseases*, 210(5):684–692, 9 2014.

[165] Nicolas Van de Velde, Marc Brisson, and Marie Claude Boily. Understanding differences in predictions of HPV vaccine effectiveness: A comparative model-based analysis. *Vaccine*, 28(33):5473–5484, 2010.

[166] Hillard Weinstock, Stuart Berman, and Willard Cates. Sexually Transmitted Diseases Among American Youth: Incidence and Prevalence Estimates, 2000. *Perspectives on Sexual and Reproductive Health*, 36(1):6–10, 1 2004.

[167] Cosette M. Wheeler, Xavier Castellsagué, Suzanne M. Garland, Anne Szarewski, Jorma Paavonen, Paulo Naud, Jorge Salmerón, Song Nan Chow, Dan Apter, Henry Kitchener, Jlio C. Teixeira, S. Rachel Skinner, Unnop Jaisamrarn, Genara Limson, Barbara Romanowski, Fred Y. Aoki, Tino F. Schwarz, Willy A J Poppe, F. Xavier Bosch, Diane M. Harper, Warner Huh, Karin Hardt, Toufik Zahaf, Dominique Descamps, Frank Struyf, Gary Dubin, and Matti Lehtinen. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncology*, 13(1):100–110, 2012.

[168] WHO. Global Influenza Surveillance and Response System, "Flu Net", 2017.

[169] Paul S. Wikramaratna, Adam Kucharski, Sunetra Gupta, Viggo Andreasen, Angela R. McLean, and Julia R. Gog. Five challenges in modelling interacting strain dynamics. *Epidemics*, 10:31–34, 3 2015.

[170] D. C. Wiley, I. A. Wilson, and J. J. Skehel. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature*, 289(5796):373–378, 1 1981.

[171] Tom M Wilkinson, Chris K F Li, Cecilia S C Chui, Arthur K Y Huang, Molly Perkins, Julia C Liebner, Rob Lambkin-Williams, Anthony Gilbert, John Oxford, Ben Nicholas, Karl J Staples, Tao Dong, Daniel C Douek, Andrew J McMichael, and Xiao-Ning Xu. Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. *Nature Medicine*, 18(2):274–280, 2 2012.

[172] Rachel L. Winer, James P. Hughes, Qinghua Feng, Sandra O'Reilly, Nancy B. Kiviat, King K. Holmes, and Laura A. Koutsky. Condom Use and the Risk of Genital Human Papillomavirus Infection in Young Women. *New England Journal of Medicine*, 354(25):2645–2654, 6 2006.

[173] Colin J Worby, Sandra S Chaves, Jacco Wallinga, Marc Lipsitch, Lyn Finelli, and Edward Goldstein. On the relative role of different age groups in influenza epidemics. *Epidemics*, 13:10–16, 12 2015.

[174] C. Wu, D. Zanker, S. Valkenburg, B. Tan, K. Kedzierska, Q. M. Zou, P. C. Doherty, and W. Chen. Systematic identification of immunodominant CD8+ T-cell responses to influenza A virus in HLA-A2 individuals. *Proceedings of the National Academy of Sciences*, 108(22):9178–9183, 5 2011.

[175] Hsiang-Yu Yuan, Marc Baguelin, Kin O. Kwok, Nimalan Arinaminpathy, Edwin van Leeuwen, and Steven Riley. The impact of stratified immunity on the transmission dynamics of influenza. *Epidemics*, 20(2017):84–93, 3 2017.

[176] Xiahong Zhao, Yilin Ning, Mark Chen, and Alex R. Cook. Individual and population trajectories of influenza antibody titers over multiple seasons in tropical Singapore. *American Journal of Epidemiology*, 6 2017.

[177] Xiahong Zhao, Karen Siegel, — Mark, I-Cheng Chen, and Alex R Cook. Rethinking thresholds for serological evidence of influenza virus infection. *Influenza Other Respi Viruses*, 11:202–210, 2017.