THE UNIVERSITY OF CHICAGO

# ECOLOGICAL AND EVOLUTIONARY DYNAMICS OF SEASONAL INFLUENZA SHAPED BY HUMAN IMMUNITY

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
FRANK TIAN WEN

CHICAGO, ILLINOIS
AUGUST 2018

For my mother.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

I am fortunate to have been surrounded by many supportive people during my time in graduate school. My advisor, Sarah Cobey, and my committee members, Greg Dwyer, Anup Malani, and Mercedes Pascual have been excellent teachers. I must also thank Greg for first introducing me to the field of infectious disease ecology and evolution during my last year of undergraduate studies. I am also grateful to Sarah for being a tireless mentor throughout my graduate studies, unafraid to explore every hypothesis.

I am grateful for the friendship and support of my lab-mates Phil Arevalo, Ed Baskerville, Kangchon Kim, Colin Kyle, Sylvia Ranjeva, Rahul Subramanian, Marcos Vieira, and Daniel Zinder, our neighbors in the department, Xiangjun Du, Qixin He, Shai Pilosof, Pamela Martinez, Matt Smith, and Mauricio Santos-Vega, and our collaborators across the country, Trevor Bedford and Sidney Bell. In particular, Ed, Matt, Trevor, and Sidney have taught me programming techniques that have become embedded in my identity as a computational scientist.

I would not have set out on this path without the initial support of Hans and Karin Schreiber, and the other members of Schreiberia during my undergraduate years. Hans?s devotion to clinically relevant animal models imprinted on me strongly, and continues to remind me of what really matters in biomedical research.

My mother, Lin Tian, and my father, Guoyong Wen, have supported me from the beginning. They not only overcame much greater obstacles to walk this same path as I, but also paved the way for me. I feel that, as children, we are often unaware of our parents' tribulations until we encounter them ourselves. I have thus only recently begun to realize that I am really not so different from my parents, and hope to someday find what they have achieved.

Lastly, I thank my loving wife, Samantha, for her enduring patience during my seemingly indefinite graduate and medical education, and our cat Buffy, for maintaining order in our

household.

# ABSTRACT

A central goal in ecology and evolution is to explain the mechanisms that determine the fitness of species. In the context of infectious disease epidemiology, understanding pathogen fitness informs approaches to disease management, especially in the case of rapidly evolving pathogens such as seasonal influenza. This dissertation examines how human ecology shapes the ecology and evolution of seasonal influenza, and thereby affects the fitness of influenza viruses. We examine how human ecology shapes geographical patterns of influenza?s evolution. We them examine how public health interventions through vaccination may affect long term patterns of evolution. We find that vaccination can slow the evolution of rapidly-evolving pathogens, which potentially reduces disease burden by more than is appreciated using present models. However, we do not find consistent evidence for vaccine-driven selection in presently available surveillance data.

# CHAPTER 1

# INTRODUCTION

An important subject in ecology and evolution is to characterize fitness with the goal of understanding why some individuals and populations succeed while others fail. In the context of infectious diseases, two evolutionary concerns are often the focus of approaches to disease management. Where do pathogens come from and why do they originate there? How do pathogens adapt to interventions?

The evolutionary dynamics of obligate pathogen populations are rooted in competition among pathogens for susceptible hosts as a resource. For some pathogens, interactions with host immunity divide pathogen populations into groups based commonly referred to as strains [46, 85, 84, 86]. The precise definition of a strain depends on the combination of immunological interactions between host and pathogen. For some pathogens, such as the measles virus, host immunity does not distinguish among members of the pathogen population, and the evolutionary dynamics appear to be neutral with respect to host immunity [82]. Other pathogens, such as the influenza virus, elicit immunity that has homologous and heterologous components with respect to strains within the pathogen population. Heterologous immunity targets shared features of individual pathogens, which facilitates competition for susceptible hosts, potentially causing competitive exclusion [77]. Homologous immunity, which targets features that are specific to certain individual pathogens, facilitates selection for immunologically distinct rare variants through negative frequency-dependent selection, thus stabilizing coexistence [43].

The influenza virus has adopted a strategy of rapid mutation of antigenic sites to successfully persist in the human population. During and after infection, hosts develop immunity specific to individual viruses, primarily through antibodies directed at the head region of the hemagglutinin (HA) surface protein. In turn, mutations in antigenic sites allow the virus to evade host immunity against existing strains. The epidemic processes that emerge

from interactions between influenza and human immunity select for novel mutants, sometimes causing replacement of older circulating viruses [82]. The HA antigen is considered immunodominant because hosts preferentially develop antibodies that bind HA instead of other targets. Because of HA's immunodominance and the prevalence of anti-HA antibodies that only bind specific groups of viruses, individual influenza viruses have been classified into strains based on antibody reactivity against HA [69]. However, influenza viruses' appearance to the host immune system (i.e. their immunological phenotype) is not determined by specific antibodies against HA alone. The immune response against influenza also has heterologous components, in the form of antibodies that bind conserved sites on stalk region of HA and $CD8^+$ and $CD4^+$ T cell immunity directed against conserved epitopes in other viral proteins [188, 50, 161, 64, 178, 190]. How specific and generalized components of immunity collectively shape influenza's evolution remains largely unexplored, as much study has focused on specific immunity. Accounting for the contributions of specific and general immunity to a virus' immunological phenotype by is necessary to predict ecological and evolutionary dynamics resulting from host-pathogen interactions.

Pathogen fitness is also affected by non-immune mechanisms. The range of hosts that the pathogen is able to infect (i.e. host tropism) in part influences the relationship between pathogen survival and host behavior [59]. For example, among pathogen species without environmental reservoirs, survival depends on the host's survival and transmissive potential, which in turn depends on the modes of transmission. Seasonal influenza transmission appears to be partly mediated through direct contact and airborne particles [144, 145], which implies influenza transmission may be density-dependent, though indirect contact through fomites [9] would suggest less density-dependent transmission. For pathogens that also elicit specific immunity and have brief durations of infection, such as seasonal influenza, successful transmission between hosts is even more critical to success. Population dynamics of hosts and the generation of immunity potentially interact in complex ways to define the evolutionary

landscape of pathogen species.

This dissertation explores drivers of influenza's evolutionary success from two perspectives. I examine how human ecology influences the phylogeography of seasonal influenza in theory, and subsequently how immunological interventions through vaccination affects influenza's evolution in theory and in practice.

## 1.1 Overview of dissertation

### *1.1.1 Chapter 2. Explaining the geographical origins of seasonal influenza (H3N2)*

The geographic distributions of obligate pathogens is dependent on ecological properties of host populations, such as population structure and demography. Most antigenically novel and evolutionarily successful strains of seasonal influenza A (H3N2) originate in East, South, and Southeast Asia. To understand this pattern, we simulated the ecological and evolutionary dynamics of influenza in a host metapopulation representing the temperate north, tropics, and temperate south. Although seasonality and air traffic are frequently used to explain global migratory patterns of influenza, we find that other factors may have a comparable or greater impact. Notably, a region's basic reproductive number ($R_0$) strongly affects the antigenic evolution of its viral population and the probability that its strains will spread and fix globally: a 17-28% higher $R_0$ in one region can explain the observed patterns. Seasonality, in contrast, increases the probability that a tropical (less seasonal) population will export evolutionarily successful strains but alone does not predict that these strains will be antigenically advanced. The relative sizes of different host populations, their birth and death rates, and the region in which H3N2 first appears affect influenza's phylogeography in different but relatively minor ways. These results suggest general principles that dictate the spatial dynamics of antigenically evolving pathogens and offer predictions for how changes

in human ecology might affect influenza evolution.[1]

### *1.1.2 Chapter 3. Vaccination and the evolution of seasonal influenza*

Host populations that support a diverse population of pathogen strains can have differential strength and breadth of immunity among the resident strains, which causes selection for strains that hosts have less immunity against. For pathogens that rapidly generate strains with novel immunological phenotypes (i.e. antigenic strains), selection may accelerate the rate of phenotypic evolution. Such evolutionary effects could change the benefits that vaccines confer to vaccinated individuals and the host population (i.e. private and social benefits). To investigate vaccination's potential evolutionary impacts on a rapidly-evolving pathogen, we simulated the dynamics of an influenza-like pathogen in an annually vaccinated host population. On average, vaccination decreased the cumulative amount of antigenic evolution of the viral population and the incidence of disease. Vaccine-driven accelerated antigenic evolution only occurred in extremely rare cases (at low vaccination rates when the breadth of vaccine-induced immunity was much narrower than naturally-acquired immunity). To understand how the evolutionary effects of vaccination might affect its private and social benefits over multiple seasons, we fit linear panel models to simulated infection and vaccination histories. Including the evolutionary effects of vaccination lowered the private benefits but increased the social benefits compared to when evolutionary effects were ignored. Thus, in the long term, vaccines' private benefits may be lower and social benefits may be greater than predicted by current measurements of vaccine impact, which do not capture long-term evolutionary effects. These results suggest that conventional vaccines against seasonal influenza could greatly reduce the burden of disease by slowing antigenic evolution like universal vaccines. Furthermore, vaccination's evolutionary effects compound a collective

1. The results from chapter 2 were published previously [181]. Much of the chapter text appears as in the published manuscript.

action problem, highlighting the importance on social policies concerning vaccination.[2]

### *1.1.3 Chapter 4. Estimating vaccine-driven selection in seasonal influenza*

Vaccination could be an evolutionary pressure on seasonal influenza if vaccines reduce the transmission rates of some ("targeted") strains more than others. In theory, more vaccinated populations should have a lower prevalence of targeted strains compared to less vaccinated populations. We tested for vaccine-induced selection in influenza by comparing strain frequencies between more and less vaccinated human populations. We defined strains in three ways: first as influenza types and subtypes, next as lineages of type B, and finally as clades of influenza A/H3N2. We were unable to detect consistent spatial differences in the frequencies of subtypes and types or between the lineages of influenza B, suggesting that vaccines do not select strongly among these phylogenetic groups at regional scales. We did detect a significantly greater frequency of an H3N2 clade with known vaccine escape mutations in more vaccinated countries during the 2014-2015 season, which is consistent with vaccine-driven selection within the H3N2 subtype. Overall, we do not find consistent support for vaccine-driven selection in influenza. Further examination of the influenza vaccine's evolutionary effects will probably require improvements in epidemiological surveillance.

2. The results from chapter 3 were published previously [182]. Much of the chapter text appears as in the published manuscript.

# CHAPTER 2

# EXPLAINING THE GEOGRAPHICAL ORIGINS OF SEASONAL INFLUENZA (H3N2)

## 2.1 Introduction

Antigenic variants of seasonal influenza continuously emerge and escape human immunity in a process known as antigenic drift. These drifted strains are less easily recognized by host immunity and therefore have a transmission advantage. More antigenically advanced strains are also more likely to spread globally and successfully perpetuate the evolutionary lineage of subsequent variants.

Asia has long been recognized as a major source of not only new influenza subtypes but also new strains of seasonal influenza [147, 179, 53, 54]. Influenza A/H3N2, A/H1N1, and two B lineages currently circulate in the human population, with the H3N2 subtype causing the most disease [187]. Phylogeographic analyses show that East, South, and Southeast Asia contribute disproportionately to the evolution of seasonal H3N2, exporting most of the evolutionarily successful strains that eventually spread globally [140, 142, 6, 10, 12]. The trunk of H3N2's phylogeny traces the evolutionary path of the most successful lineage and was estimated to be located in Asia 87% of the time from 2000 to 2010 [12]. Additionally, strains of H3N2 isolated in E-SE Asia appear to be more antigenically advanced, with new antigenic variants emerging earlier in E-SE Asia than in the rest of the world [142, 13]. These observations suggest that ecological differences between regions, such as climate and human demography, affect the local antigenic evolution of H3N2, which in turn shapes its global migratory patterns. Here we ask what ecological factors might cause disproportionate contributions of particular host populations to the evolution of an influenza-like pathogen. This information may be immediately useful for viral forecasting. Over the long term, it could help predict changes in influenza's phylogeography and identify source populations to

improve global vaccination strategies.

The conspicuous role of Asia in H3N2's evolution has been attributed to the seasonal nature of influenza in temperate regions [179, 175, 140, 142, 10, 6]. Approximately 85% of Asia's population and 48% of the global population resides in a climatically tropical or subtropical region [44] where semiconnected host populations support asynchronous epidemics that enable regional persistence year-round [175, 142, 42]. Uninterrupted transmission might increase both the efficiency of selection and the probability of strain survival and global spread. By contrast, transmission bottlenecks from late spring through autumn in temperate populations necessarily limit local evolution and reduce opportunities for strain emigration [2, 193]. Smaller contributions from other tropical and subtropical regions might arise from the weaker connectivity of their host populations [10, 37, 106].

Although seasonality clearly affects temporal patterns of viral migration [6], a robust explanation for differences in regions' long-term contributions to the evolution of H3N2 would consider the effects of seasonal variation in transmission in light of other potentially influential differences among host populations, including:

*Host population size.* E-S-SE Asia alone contains more than half of the global population [171]. Larger host populations should sustain larger viral populations, and in the absence of other effects, they should contribute a proportionally larger fraction of strains that happen to spread globally. Additionally, if rare mutations limit the generation of antigenic variants, larger populations could contribute a disproportionate number of antigenically novel strains with high fitness.

*Host population turnover.* Birth rates have historically been higher in E-S-SE Asia than in most temperate populations [171]. Demographic rates influence the replenishment of susceptibles and loss of immune individuals, thereby modulating selection for antigenic change. Faster replenishment of susceptibles increases prevalence, and thus viral abundance and diversity, but weakens the fitness advantage of antigenic variants. A more immune population

imposes greater selection for antigenic change but supports a smaller, less diverse viral population. Thus, the rate of antigenic evolution may vary in a complex way with the rate of host population turnover [82].

*Initial conditions.* H3N2 first emerged in or near Hong Kong in 1968. The region in which a subtype emerges may effectively give the viral population a head start on evolution. The first epidemic will almost certainly occur in this region, and viruses here will be the first to experience selective pressure for antigenic change. If host migration rates are low and the founding viral population persists, this antigenic lead could be maintained or even grow in time.

*Transmission rates.* Differences in human behaviour can affect transmission rates. The transmission rate affects a strain's intrinsic reproductive number ($R_0$), the expected number of secondary cases caused by a single infection in an otherwise susceptible population. Differences in regional $R_0$ could affect evolution in at least two ways. Higher $R_0$ increases the equilibrium prevalence, increasing the probability that rare beneficial mutations will appear. In addition, the rate of antigenic drift increases with $R_0$ in models that include mutation as a diffusion-like process [111, 80, 101, 12]. A higher intrinsic reproductive number in one population could thus accelerate the emergence of novel mutants in that area.

To understand the potential effects of these five factors on the evolution of H3N2 in space, we simulated an influenza-like pathogen in a simplified representation of the global human metapopulation. The simulated metapopulation consisted of three connected host populations, representing the temperate north, tropics, and temperate south. Conceptually, the tropics in the model approximate Asia, where most of the population is tropical or subtropical [44] and epidemics are asynchronous, and exclude other less connected tropical and subtropical populations on other continents [10, 37, 106]. The two temperate populations approximate northern and southern populations where influenza is strongly seasonal. The model can also be generalized to represent three arbitrary populations by reducing season-

ality.

We analysed the effects of these factors on two key metrics of influenza's spatial evolutionary and antigenic dynamics. The first metric measures the proportion of the trunk of the phylogeny present in the tropics (figure 2.1*a*). The phylogenetic trunk represents the most evolutionarily successful lineage that goes on to seed all future outbreaks. The second metric measures the degree to which tropical strains are antigenically advanced (figure 2.1*b*). Phenotypically, antigenic dissimilarities can be quantified as distances in antigenic space using pairwise measures of cross reactivity [160, 13]. Our model uses an analogous measure of antigenic distances, allowing us to determine the relative antigenic advancement of strains from each region. We analysed these two metrics from simulations to test whether any of the five ecological factors could create spatial evolutionary patterns of a similar magnitude to the observed data.

## 2.2 Results

### *2.2.1 Influenza-like patterns*

We simulated an individual-based model that included ecological and evolutionary dynamics in a metapopulation with three demes [11]. By default, in one deme, transmission rates are constant throughout the year, and in the two others, transmission rates vary sinusoidally with opposing phases. Viral phenotypes occur as points in 2D Euclidean space, and mutation displaces phenotypes in this 2D space according to a fixed kernel [11]. This space is analogous to an antigenic map constructed from pairwise measurements of cross-reactivity between influenza strains using a hemagglutination inhibition (HI) assay [160, 13]. Susceptibility to infection is proportional to the distance in antigenic space between the challenging strain and the nearest strain in the host's infection history, giving distant or antigenically advanced strains greater transmissive advantage.

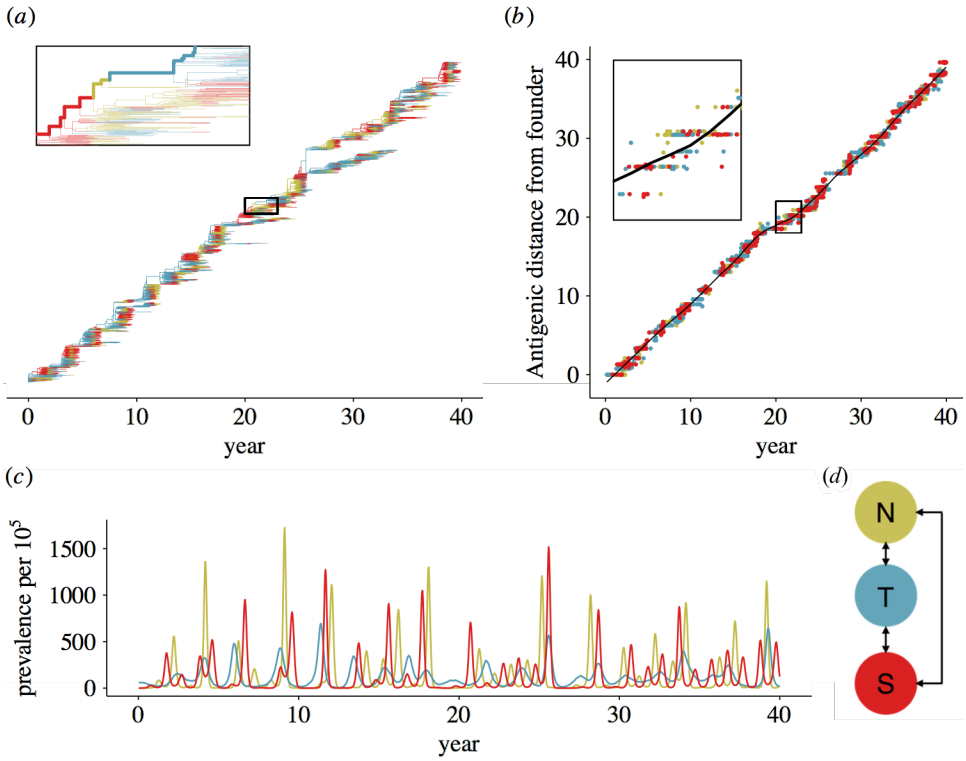


Figure 2.1: Representative output showing influenza-like behaviour from a sample simulation using the default parameters (table 2.2). Statistics reported here are based on 53 replicate simulations. (*a*) The phylogeny of the pathogen is reconstructed explicitly from the recorded ancestry of simulated strains. Branches are colored by region indicated in panel *d*. The trunk is determined by tracing the recorded ancestry of surviving strains at the end of the simulation. Side branches show lineages that go extinct. (*b*) Viruses evolve antigenically away from the founding strain in a canalized fashion. On average, the antigenic distance from the founding strain follows the trajectory indicated by the black LOESS spline fitted to viruses from all three regions. At any given point in time, strains above this line have drifted farther from the founder compared to average, and are thus considered antigenically leading. Conversely, strains below this line are considered antigenically lagging. Antigenic lead is calculated as the distance to the spline in antigenic units. (*c*) Prevalence of infection over time for each region. (*d*) Depiction of the totally connected model population, composed of the temperate north, tropics, and temperate south.

The model reproduces the characteristic ecological and evolutionary features of H3N2, except for the antigenic lead (table 2.1), under the default parameters (table 2.2). We restricted our analyses to simulations where the virus remained endemic and where the time to the most recent common ancestor (TMRCA) never exceeded 10 years during the 40 years of simulation. We chose this cutoff because in some simulations, the viral population developed

unrealistically deep branches. In excluding extinctions and excessive diversity (branching), we assume that H3N2's historical evolutionary patterns represent the virus' likeliest evolutionary dynamics. Of 100 replicate simulations, the viral population went extinct in 18 cases and exceeded the TMRCA threshold 29 times, leaving 53 simulations for analysis. The model tracks the ancestry of individual strains, allowing us to explicitly reconstruct the phylogeny of the virus and the geographic location of lineages. The phylogeny has the characteristically well-defined trunk with short branches of the H3N2 hemagglutinin (figure 2.1). This shape arises due repeated selective sweeps of antigenic variants, which reduces standing diversity; the average TMRCA across replicates was 3.72 years (SD = 0.26), comparable to empirical estimates of 3.89 years [12]. The antigenic distance from the founder increased linearly with time (figure 2.1), characteristic of H3N2's canalized antigenic evolution [11, 160]. The mean antigenic drift across replicate simulations was 0.97 antigenic units per year (SD = 0.11), comparable to observed rates of 1.01 antigenic units per year [13]. The mean annual incidence was 9.1% (SD = 0.8%). Reported annual incidence across all subtypes of seasonal influenza range from 9-15% [184]. Since we only modeled one lineage (e.g., the H3N2 subtype), the low estimate from the model is comparable to observed incidence.

Although all three host populations were the same size, the tropical strains were on average more evolutionarily successful. The phylogenetic trunk traces the most evolutionarily successful lineage and was located in the tropics 77% (SD = 13%) of the time, comparable to the observed 87% of H3N2's trunk in E-S-SE Asia between 2000-2010 [12]. However, the default parametrization does not produce an antigenic lead in any population, despite the observed antigenic lead of Asian strains (table 2.1). Antigenic cartography shows that while H3N2 drifts on average at 1.01 antigenic units per year globally [13], Asian strains tend to be farther drifted at any given time, and the region is thus considered to lead antigenically [142, 13].

Table 2.1: Properties of the default model

| Statistic | Model mean ± SD | Observed (Ref) |
|---|---|---|
| Annual incidence | 0.091±0.0077 | 0.09 - 0.15 [184] |
| Antigenic drift rate (a.u. $yr^{-1}$) | 0.97±0.11 | 1.01 [13] |
| TMRCA (years) | 3.7±0.26 | 3.89 [12] |
| Frac. of trunk in the tropics | 0.61±0.13 | 0.87 [12] |
| Tropics antigenic lead (a.u.) | 0.0025±0.036 | 0.25 [142, 13] |

Table 2.2: Default parameters

| Parameter | Value | Reference |
|---|---|---|
| Intrinsic reproductive number ($R_0$) | 1.8 | [94, 19] |
| Duration of infection $\nu$ | 5 days | [30] |
| Population size $N$ | 45 million | (see SI) |
| Birth/death (turnover) rate $\gamma$ | 1/30 $year^{-1}$ | [171] |
| Mutation rate $\mu$ | $10^{-4}$ $day^{-1}$ | (see SI) |
| Mean mutation step size $\delta_{mean}$ | 0.6 antigenic units | (see SI) |
| SD mutation step size $\delta_{sd}$ | 0.3 antigenic units | (see SI) |
| Infection risk conversion $c$ | 0.07 | [11, 87, 135] |
| Migration rate $m$ | $10^{-3}$ $day^{-1}$ | (see SI) |
| Seasonal amplitude $\epsilon$ | 0.10 | [169] |

### *2.2.2 Seasonality*

We first varied the strength of seasonal forcing, holding other parameters at their default values. Seasonality by itself in the two temperate populations could not cause the tropics to produce more antigenically advanced strains; however, seasonality did cause the tropics to contribute a greater fraction of evolutionarily successful strains (figure 2.2). By linear regression, we estimate that the trunk would spend 87% of its time in the tropics (the same fraction that is observed in Asia [12]) with a seasonal transmission amplitude ($\epsilon$) of 0.19 (95% CI: 0.18, 0.20). Reduced seasonal forcing in the temperate populations equalized the fraction of the trunk in each population. In multivariate sensitivity analysis, the amplitude of seasonal transmission accounted for 33% of the variation in the tropical fraction of the trunk (supplementary information, figure 2.6, table 2.4). This result suggests that seasonal bottlenecks in temperate populations discourage seasonal strains from fixing globally, in agreement with

other models [2]. However, seasonality alone could not explain any variation in the tropic's antigenic lead (supplementary information, figure 2.6, table 2.5). We therefore hypothesized that ecological factors besides seasonality must contribute to regional differences in relative antigenic fitness.

### *2.2.3 Transmission rate in the tropics*

Increasing $R_0$ in the tropics relative to the temperate populations caused the tropics to produce strains that led antigenically while also preserving the tropics' contribution to the trunk (figure 2.3). Linear regression implies that a 28% (95% CI: 25%, 30%) increase in $R_0$ in the tropics causes the tropics to produce strains that are, on average, 0.25 antigenic units ahead of global mean, reproducing the observed antigenic lead in Asia [142, 13]. We also estimate that a 17% increase in $R_0$ (95% CI: 15%, 19%) causes the phylogenetic trunk to be located in the tropics 87% of the time, reproducing the observed fraction of the H3N2 trunk in Asia [12].

The effects of $R_0$ on antigenic lead were robust to changes in other ecological variables and over a range of baseline values of global $R_0$. When we varied the other parameters (table 2.2), relative $R_0$ in the tropics accounted for 77% of the variance in antigenic lead, making it the best predictor of antigenic lead in the tropics (supplementary information, figure 2.6, table 2.5). The fraction of the trunk in the tropics also increased with the relative $R_0$, although $R_0$ explained less of the variation in trunk proportion (41%), due to the effect of seasonality (supplementary information, figure 2.6, table 2.4).

Notably increased $R_0$ in one deme was sufficient by itself to make strains more evolutionarily successful and antigenically advanced. When we removed seasonality altogether to model three climatically identical populations, the population with the highest $R_0$ produced both the most antigenically leading and evolutionarily successful strains (figure 2.4). Thus, higher $R_0$ alone in one region can cause it to attain an antigenic lead and fraction of the

trunk as large as is observed in Asia.

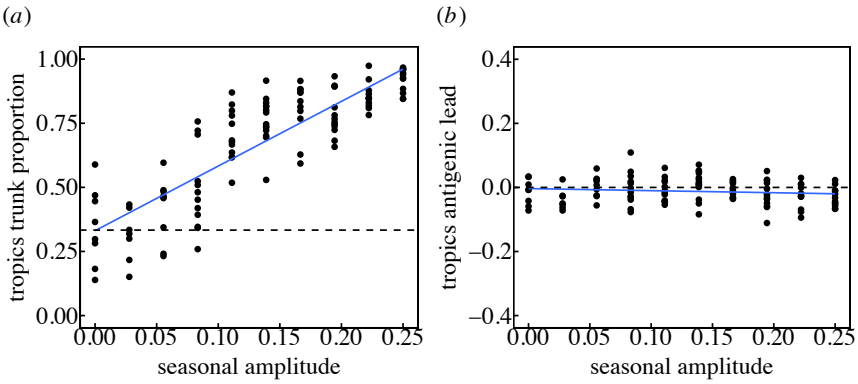


Figure 2.2: Seasonal amplitude $\epsilon$ in the temperate populations increases the tropics' contribution to the most evolutionarily successful lineage but alone does not affect regional differences in antigenic advancement. Transmission rates $\beta$ in the temperate north and south oscillate sinusoidally in opposite phase, with amplitude $\epsilon$. All other parameters remain at their default values (table 2.2). ($a$) Effects of seasonality on the fraction of the trunk in the tropics (Pearson's $r = 0.85$, $p < 0.001$; $R^2 = 0.72$). Each point shows the fraction of time that the phylogenetic trunk was located in the tropics during the course of one simulation. The dashed line represents the null hypothesis where tropical strains comprise one third of the phylogenetic trunk. ($b$) Effects on seasonality on the antigenic lead of the tropics (Pearson's $r = -0.12$, $p = 0.20$, $R^2 = 0.01$). Each point shows the average antigenic lead of tropical strains over time from one simulation. The dashed line represents the null hypothesis where tropical strains are neither antigenically ahead or behind. Blue lines represent linear least squares regression.

To better understand why increasing regional $R_0$ causes that region to produce more antigenically advanced strains, we examined the effect of $R_0$ on antigenic evolution in a single deme. Simulations showed that increasing $R_0$ increases the rate of antigenic drift (supplementary information, figure 2.7). To investigate further, we derived an analytic expression for the invasion fitness of a novel mutant in a population at the endemic equilibrium (supplementary information, equation S1). When the resident and mutant strains have the same intrinsic fitness ($R_0$), the growth rate of an antigenically distinct, invading mutant increases linearly with $R_0$ (supplementary information, figure 2.8). This linearity holds as long as the conversion between antigenic distance and host susceptibility (equation 2.3) is independent of $R_0$. As $R_0$ increases, not only do mutants invade faster, but the invasion

speed increases faster as a function of antigenic distance (supplementary information, figure 2.8).

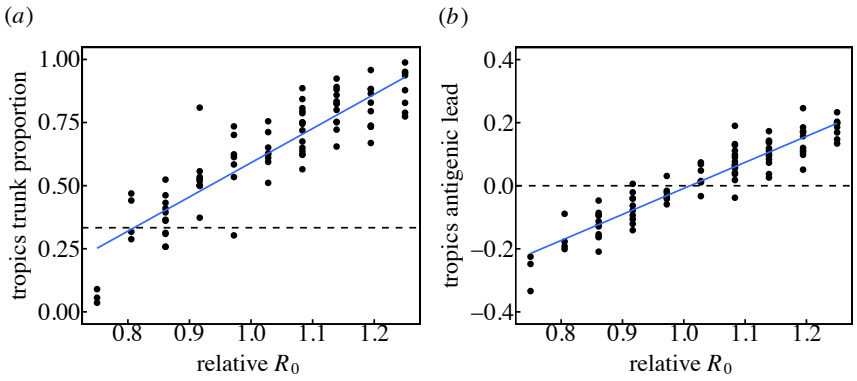


Figure 2.3: Increased $R_0$ in the tropics increases the tropics' contribution to the most evolutionarily successful lineage and the antigenic advancement of tropical strains. Relative $R_0$ is calculated as $R_0$ in the tropics divided by $R_0$ in the temperate regions. $R_0$ in the tropics was varied while $R_0$ in the temperate regions was kept at its default. Other parameters were also kept at their default values (table 2.2). (*a*) Effect of $R_0$ in the tropics on the fraction of the trunk in the tropics (Pearson's $r = 0.88$, $p < 0.001$; $R^2 = 0.78$). Each point shows the fraction of phylogenetic trunk located in the tropics during one simulation. The dashed line represents the null hypothesis where tropical strains comprise one third of the phylogenetic trunk. (*b*) Effect of $R_0$ in the tropics on the antigenic lead in the tropics (Pearson's $r = 0.93$, $p < 0.001$; $R^2 = 0.87$). Each point shows the average antigenic lead of tropical strains over time from one simulation. The dashed line represents the null hypothesis where tropical strains are neither antigenically ahead or behind. Blue lines represent linear least squares regression.

Although seasonality alone did not affect antigenic lead, the effects of $R_0$ on antigenic lead could be influenced by seasonality (figure 2.4). Introducing seasonality in the temperate populations reduced differences in antigenic phenotype between regions. When tropical strains were antigenically ahead of temperate strains (due to higher tropical $R_0$), introducing seasonality reduced the tropics' antigenic lead. When tropical strains were antigenically behind temperate strains (due to lower tropical $R_0$), introducing seasonality reduced the antigenic lag. Two factors explain the equalizing effect of seasonality on antigenic phenotype. First, higher contact rates during transmission peaks in the two temperate populations increase the rate of strain immigration from the tropics. Second, seasonal troughs in prevalence allow

tropical strains to invade more easily due to reduced competition with local strains.

### *2.2.4 Demographic rates, population size, and initial conditions*

Other ecological factors affected regional contributions to evolution but could not reproduce the observed patterns as well as differences in $R_0$ (supplementary information, figures 2.5, 2.6). Notably, strains were slightly more antigenically advanced in older populations (supplementary information, figure 2.5). When the rate of population turnover in the tropics was half that in the temperate regions, the tropics led by 0.04 antigenic units (SD = 0.03). Larger populations generally contributed more to the trunk, although there was much variation that population size alone did not explain (supplementary information, figures 2.5, 2.6 and table 2.4, 2.5). Initial conditions did not have a lasting effect (supplementary information, figure 2.9).

### *2.2.5 Implications for other influenza subtypes*

Influenza A/H1N1 and influenza B both evolve slowly compared to H3N2 and are suspected to have lower $R_0$ [12, 13]. Specifically, H1N1 drifts at a rate of 0.62 antigenic units per year, and the B/Victoria and Yamagata strains drift at 0.42 and 0.32 antigenic units per year respectively [13]. H1N1 and B viruses are also less apt to have Asian origins than H3N2 [12]. When we simulate with lower baseline $R_0$, we find that differences in $R_0$ between regions have a weaker influence on spatial patterns of evolution (supplementary information, figure 2.12). Based on the relationship between mean $R_0$ and antigenic drift (supplementary information, figure 2.7), we would expect seasonal H1N1, for example, to have an $R_0$ of 1.6. For this $R_0$, a 17% increase in $R_0$ causes the tropics to occupy only 79% (versus 87% for H3N2-like $R_0$ of 1.8) of the trunk, and a 28% increase in $R_0$ causes the tropics to lead by 0.20 (versus 0.25 for H3N2) antigenic units.

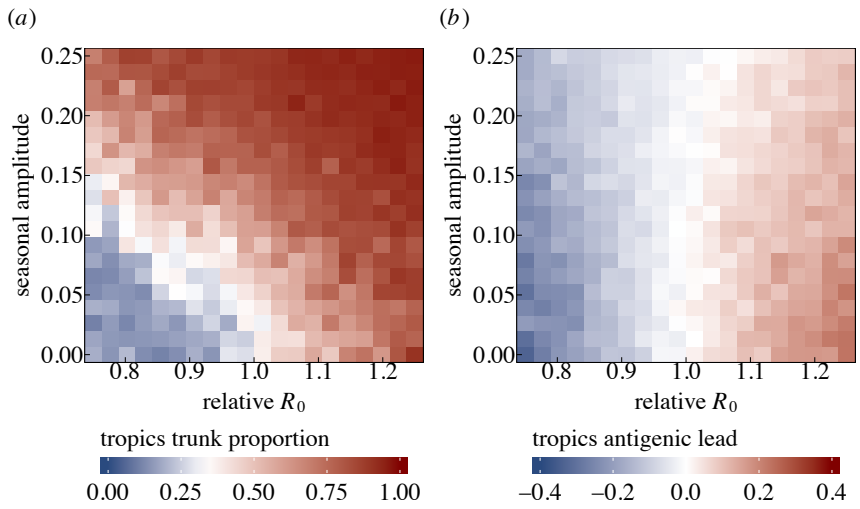


Figure 2.4: Seasonality in temperate populations has an equalizing effect on antigenic differences. Relative $R_0$ is calculated as $R_0$ in the tropics divided by $R_0$ in the temperate regions. ($a$) Effects of seasonality and $R_0$ on the fraction of the trunk in the tropics. Blue indicates that the phylogenetic trunk is located in the tropics less than 1/3 of the time, and red indicates that the trunk is the tropics more than 1/3 of the time. ($b$) Effects of seasonality and $R_0$ on antigenic lead in the tropics. Blue indicates that tropical strains are on average ahead antigenically relative to other global strains and red indicates that tropical strains are behind antigenically. Each square averages 1 to 17 replicate simulations.

## 2.3 Discussion

In our model, we find that the simplest explanation for why a host population produces more antigenically novel and evolutionarily successful strains than other populations is that its strains have a higher intrinsic fitness, or $R_0$. The strong effect of regional $R_0$ on spatial patterns of viral evolution is caused by the effect of $R_0$ on antigenic drift. Higher regional $R_0$ facilitates invasion of antigenically novel strains, resulting in faster antigenic drift. Seasonality reduces the rate at which temperate populations export strains that are evolutionarily successful, but seasonality alone cannot explain regional differences in the production of strains that are antigenically novel. Size and age can influence global patterns too, but to a lesser extent: larger populations export more strains that fix, and populations with slower replenishment of susceptibles increase the rate of antigenic evolution. These last two effects are sensitive to changes in seasonality and $R_0$. These results highlight the relationship be-

tween human ecology and influenza's phylogeography. Regions with high transmission rates may be expected to contribute disproportionately to influenza's evolution and may also be ideal targets for vaccine campaigns. Accordingly, changes in human ecology can be expected to alter influenza's phylogeography. These generalizations assume that H3N2 will evolve mostly as it has, with high strain turnover and limited genetic variation at any time, but more complex dynamics may be possible.

To make general predictions, we used a simple model. Although our three-deme metapopulation prevents us from replicating influenza's phylogeographic dynamics precisely, the model nonetheless reveals how ecological differences in populations create spatial patterns in the evolution of an influenza-like pathogen. Simulations with more complex metapopulation models showed the same trends as the simple three-deme model (figures 2.13, 2.14), suggesting that our results are robust to changes in metapopulation population structure.

These results immediately raise the question of whether there is evidence of regional variation in $R_0$. Low reporting rates and antigenic evolution make the $R_0$ of influenza difficult to measure with traditional methods, but we can conjecture from several lines of evidence. Low absolute humidity favors transmission via aerosol in experimental settings [144] and influences the timing of the influenza season in the United States [145]. Based on absolute humidity and aerosol transmission alone, these results suggest that $R_0$ of tropical and subtropical Asia would be lower than in temperate latitudes. However, in Vietnam the onset of influenza-like illness is associated with periods of high humidity [166]. This observation suggests that humidity is not the dominant driver of influenza transmission, at least in this region.

Contact rates also influence transmission [177]. Multiple studies have detected a significant effect of school closure on influenza spread [31, 89, 38], although this trend is not without exception [51]. Households also influence risk: after one household member is infected, the average risk of secondary infection in a household contact is 10% [170]. Differences

in classroom and household sizes may thus influence local transmission, and both are higher in, for instance, China and India than in Europe and the U.S. [172, 128]. Contact surveys report higher contact rates in Guangdong, China, than in European communities, whereas those in Vietnam are lower, although differences may arise from differences in survey design [141, 126, 91]. These surveys notably miss non-social, casual contacts (e.g., shared cafeterias and elevators) that might be important for influenza transmission.

Differences in local transmission rates may not scale: high rates of local transmission may be offset or attenuated by the structure of contact networks over larger areas. At the regional level, commuter and air passenger flows affect the spread of influenza epidemics, suggesting that adults are important to the long-range dispersal of the virus [175, 106]. The frequency of long-distance contacts differs between communities [141]. Although sensitivity of $R_0$ to network topology is well known theoretically [124, 3], there is a need to integrate the features of local and regional empirical transmission networks to infer large-scale differences $R_0$.

Empirical estimates of $R_0$ are in theory attainable from seroprevalence. Under a simplistic, single-strain $SIR$ model, which assumes random mixing and no maternal immunity, differences in $R_0$ should appear in differences in seropositivity by age. For instance, if $R_0$=1.8, approximately 5.1% of 2 year-olds would be seropositive, whereas 7.4% would be seropositive if $R_0$ were 20% higher. $R_0$ variation in this range could be detected by sampling as few as 1500 2 year-olds in each population. Detailed surveys of H3N2 seropositivity by age cohort exist for some European countries [22, 143] but show much faster increases in seropositivity with age than expected under the $SIR$ model: 100% of tested children are seropositive to H3N2 by age 7 in the Netherlands and by age 12 in Germany. This discrepancy between theory and data may be due to antigenic drift resulting in higher attack rates [12]. The spatial difference in seroprevalence may also reflect greater contact rates among school-aged children [126] and highlights the possibility that differences in exposure rates at

young ages do not reflect mean differences in the populations. Such effects may be reduced by examining seroprevalence at older ages, but these estimates must balance a tradeoff between minimizing age-related correlations in transmission rates and increasing sample sizes required to detect asymptotically small differences in seropositivity. Another potential approach to measuring $R_0$ is to refine estimates of annual incidence in different populations. Estimates of $R_0$ based on annual incidence would have to incorporate the histories of recent circulating strains, survey timing and titer dynamics, and vaccination in each population.

A greatly reduced birth rate confers a slight antigenic lead, but actual differences in birth rates between regions appear too small to explain Asia's observed lead. Current birth rates across most of Europe, China, and the United States are within 10% of each other [171]. Birth rates are almost twice as high in some SE Asian countries, including Cambodia, Laos, and the Philippines. The highest birth rates are found in Africa and the Middle East, and are three to four times higher than birth rates in the United States and China. Our model suggests that these regions should contribute relatively less to influenza's antigenic evolution, assuming the differences in population structure are not associated with higher $R_0$, and ignoring other differences. However, taking age-assortative mixing into account may negate this expectation, with younger populations having increased $R_0$ [3, 60] thus contributing more to antigenic evolution.

We expect these results to apply to other antigenically varying, fast-evolving pathogens, including other types of influenza. Enterovirus-71 circulates globally, and its VP1 capsid protein experiences continuous lineage replacement through time, similarly to H3N2 hemagglutinin [165]. Norovirus also demonstrates rapid antigenic evolution by amino acid replacements in its capsid protein [113]. We might expect that areas with high transmission contribute disproportionately to the antigenic evolution and global spread of these pathogens. In addition, when we simulate with lower $R_0$, we find that differences in $R_0$ between regions influence spatial patterns of antigenic variation less (supplementary information, figure 2.12).

This may explain why influenza A H1N1 and influenza B, which are suspected to have lower $R_0$ [12, 13], are less apt to have Asian origins than H3N2 [12].

## 2.4 Material and methods

We implemented an individual-based $SIR$ compartmental model of an influenza-like pathogen, originally described by Bedford et al. [11]. In this model, a global metapopulation is composed of three connected populations, representing tropics and temperate north and south. Individuals' compartments are updated using a $\tau$-leaping algorithm. Within a region $i$, the force of infection is given by

$$F_i(t) = \beta_i(t)\frac{I_i}{N_i} \tag{2.1}$$

where $I$ is the number of infected hosts. Between regions $i$ and $j$, the force of infection is given by

$$F_{ij}(t) = m\beta_j(t)\frac{I_i}{N_j} \tag{2.2}$$

where region $i$ is where the infection originates and region $j$ is the destination. Here, $m$ is a scaling factor for interregional transmission, and $\beta_j$ is the transmission rate of the destination region. Transmission rates in the seasonal north and south oscillate sinusoidally in opposite phase with amplitude $\epsilon$. After recovery from infection, a host acquires complete immunity to viruses with that specific antigenic phenotype. Hosts that clear infection accumulate an infection history that defines their immunity. In a contact event, the distances between the infecting viral phenotype and each phenotype in the susceptible host's immune history are calculated. The probability of infection after contact is proportional to the distance $d$ to the closest phenotype in the host's immune history. An individual's risk of infection by such a strain is

$$\text{Risk} = \min\{1, cd\} \tag{2.3}$$

where the proportionality constant for converting antigenic distance to a risk of infection $c = 0.07$ [11]; in other words, one unit of antigenic distance corresponds to 7% reduction in immunity. The linear relationship $c$ between antigenic distance and susceptibility derives from studies of vaccine efficacy [11, 87, 135].

Antigenic phenotypes are represented by points in a two-dimensional Euclidean antigenic space. One unit of antigenic distance in this space corresponds to a twofold dilution of antiserum in an HI assay [160]. The model is initialized at the endemic equilibrium with antigenically identical viruses. By default, all of the initial infections occur in the tropics. Mutational events occur at a rate $\mu$ mutations per day. When a virus mutates, it moves in a random radial direction with a gamma-distributed step size. This mutation rate, along with the mutation size parameters $(\delta_{\text{mean}}, \delta_{\text{sd}})$ determine the accessibility of more distant mutations in antigenic space. The radial direction of mutation is chosen from a uniform distribution.

Additional methods are described in the supplementary information.

## 2.5 Data accessibility

Code implementing the model is available at `https://github.com/cobeylab/antigen-phylogeography.git`. The complete code for reproducing these results is available at `https://github.com/cobeylab/influenza_phylogeography_manuscript.git`.

## 2.6 Supplementary information

### *2.6.1 Extended methods*

#### Selection of parameters

Parameter values were selected to be consistent with influenza's biology and to reproduce its major epidemiological and evolutionary patterns (table 1). The population size $N$ was chosen to minimize extinctions while also making efficient use of computational resources. The population birth/death rate $\gamma = 1/30$ year$^{-1}$ reflects the global crude birth rate estimates of 34 births per 1000 [171].

The proportionality constant $m$ for calculating the between-region contact rate was calculated from the number of international air travel passengers reported by the International Civil Aviation Organization divided by the global population [93].

We chose a baseline $R_0 = 1.8$. Estimates of $R_0$ from the first pandemic wave H3N2 in 1968 range from 1.06-2.06 [94], and estimates of $R_0$ for seasonal influenza range from 1.16 to 2.5, averaging approximately 1.8 [19].

The five-day duration of infection, $1/\nu$, is based on estimates from viral shedding [30]. The transmission rate, $\beta$, is calculated using the definition of $R_0$:

$$R_0 = \frac{\beta}{\nu + \gamma} \tag{2.4}$$

We chose the seasonal amplitude to ensure consistent troughs during the off-season in temperate populations while remaining within reasonable estimates of seasonal transmission rates [169].

Mutational parameters were selected to maximize the number of simulations where evolution was influenza-like (figure 2.10). Mutations occur at a rate of $10^{-4}$ mutations per day. This phenotypic mutation rate corresponds to 10 antigenic sites mutating at $10^{-5}$ mutations

per day [140, 11]. The distance of each mutation is sampled from a gamma distribution with parameters chosen to yield a mean step size of 0.6 and a standard deviation of 0.3 antigenic units. These values correspond to a reduction in immunity of 4.2% for an average mutation (SD = 2.1%). These mutation effect parameters give the gamma distribution an exponential-like shape, so that most mutations yield small differences in antigenic fitness, while occasionally mutations will yield greater differences. We chose $\mu$, $\delta_{\text{mean}}$, and $\delta_{\text{sd}}$ so that the simulations would exhibit influenza-like behaviour as consistently as possible (figure 2.10). Here, the criteria for influenza-like behavior included endemism, reduced genealogical diversity (TMRCA $<$ 10 years) [12], and a biologically plausible mean rate of antigenic drift (1.01 antigenic units per year) [13] and incidence (9-15%) [184] (table 1, figure 2.10).

## Calculation of antigenic lead and trunk proportion

We examined two metrics that describe influenza's evolutionary dynamics. For computational tractability, these metrics were calculated using a subset of strains sampled over course of the simulation. Strains were sampled proportionally to prevalence.

To calculate antigenic lead, we first calculated the antigenic distance of each sampled strain from the founding strain (figure 1$a$). We then fit a LOESS spline to these distances over time. The spline describes the expected antigenic drift of circulating lineages at any point in time. Strains above the spline have drifted farther than average and are considered antigenically leading. Strains below the spline have drifted less than average and are considered antigenically lagging. The antigenic lead in the tropics is calculated as the average antigenic distance to this spline for all sampled tropical strains.

To calculate the fraction of the trunk in each population, we first identified strains that comprise the trunk by tracing the lineage of strains that survive to the end of the simulation (figure 1$b$). Because multiple lineages may coexist at the end of the simulation, we excluded the last five years of strains from trunk calculations. The fraction of the trunk in the tropics

is calculated as the fraction of the time the trunk was composed of tropical strains.

## Univariate sensitivity analysis

In the univariate sensitivity analyses, we created regional differences in host ecology by varying each of the five ecological parameters individually ($R_0$, population turnover rate, seasonality, population size, and initial conditions) while keeping all other parameters at their default values (table 2). To test the effects of regional $R_0$, we changed $R_0$ only in the tropics. Similarly, we tested the effects of the rate of population turnover by varying it only in the tropics. To investigate seasonality, we varied the seasonal amplitude of the transmission rate in the temperate populations. (The transmission rate in the tropics was always constant over time.) To explore population size, we examined the ratio of tropical to temperate population sizes, keeping the global population constant. We initialized all simulations at the endemic equilibrium such that the total number of initial infecteds was constant. We then scaled the number of initial infecteds in the tropics while keeping the number in the two temperate demes the same. We ran twenty replicates for each unique combination of parameter values and discarded any simulations in which the virus went extinct or the TMRCA exceeded 10 years at any time in the 40-year simulation. The analyses for antigenic lead and trunk proportion were performed on the remaining simulations (figure 2.5).

## Multivariate sensitivity analysis

To test the robustness of the effects of individual parameters on the antigenic lead and the phylogenetic trunk, we simulated 500 points from a Latin hypercube with dimensions representing relative $R_0$, seasonality, relative population size, relative population turnover rate, and the fraction of initial infecteds in the tropics (figure 2.6). The ranges for each parameter (table 2.3) were chosen to remain within reasonable estimates. We simulated

twenty replicates for each of the 500 unique parameter combination and discarded simulations in which the virus went extinct or the TMRCA exceeded 10 years at any time in the 40-year simulation. We performed an ANOVA on the remaining 4119 influenza-like simulations to determine each parameter's contribution to the variance in antigenic lead and the fraction of the trunk in the tropics (table 2.4, 2.5).

### *2.6.2 Invasion analysis*

We assume that the host population supports a resident strain at the endemic equilibrium. We develop an expression for the fitness of an invading mutant strain to explain how the selection coefficient of the mutant changes with $R_0$.

Here, $S, I$, and $R$ represent the fraction of susceptible, infected, and recovered individuals. The birth rate $\gamma$ and the death rate are equal, so the population size is constant. All individuals are born into the susceptible class. Transmission occurs at rate $\beta$, and recovery occurs at rate $\nu$.

$$\begin{aligned}\frac{dS}{dt} &= \gamma(1-S) - \beta SI \\ \frac{dI}{dt} &= \beta SI - (\nu+\gamma)I \\ \frac{dR}{dt} &= \nu I - \gamma R\end{aligned}$$

We solve for the endemic equilibrium values of $S_{\text{eq}}$, $I_{\text{eq}}$, $R_{\text{eq}}$.

$$\begin{aligned}\frac{dI}{dt} = 0 &= \beta S_{\text{eq}} I_{\text{eq}} - (\nu+\gamma) I_{\text{eq}} \\ S_{\text{eq}} &= \frac{\nu+\gamma}{\beta} \equiv \frac{1}{R_0}\end{aligned}$$

$R_0$, the basic reproductive number, is defined as the number of secondary infections from

a single infected individual in a totally susceptible population. Continuing to solve for $I_{\text{eq}}$ and $R_{\text{eq}}$, we have

$$\begin{aligned}
\frac{dS}{dt} = 0 &= \gamma(1 - S_{\text{eq}}) - \beta S_{\text{eq}} I_{\text{eq}} \\
I_{\text{eq}} &= \frac{\gamma}{\beta}(R_0 - 1) \\
\frac{dR}{dt} = 0 &= \nu I_{\text{eq}} - \gamma R_{\text{eq}} \\
R_{\text{eq}} &= \frac{\nu}{\beta}(R_0 - 1)
\end{aligned}$$

To find the selection coefficient, we develop an expression for the effective reproductive number $R_e$ for both the resident and mutant strains. $R_e$ is the expected number of secondary infections from a single infected individual in a given population. We will use the relationship

$$R_e = SR_0$$

The mutant strain is $d$ antigenic units from the resident strain. The conversion factor between antigenic units and infection risk is notated by $c$. Thus, the susceptibility to the mutant is given by $\min\{cd, 1\}$, and immunity to the mutant is $\max\{1 - cd, 0\}$. For ease of notation, we assume $cd \leq 1$, and use $k = 1 - cd$.

The fraction of the population immune to the invading strain is denoted by $R'$. Note that the population is at the endemic equilibrium of the resident strain, and not the mutant.

$$\begin{aligned}
R' &= (1 - cd)R_{\text{eq}} \\
&= \frac{\nu}{\beta}(R_0 - 1)k
\end{aligned}$$

We start by allowing coinfection. The fraction of susceptibles to the mutant strain is

given by

$$\begin{aligned} S' &= 1 - R' - \frac{1}{N} \\ &= 1 - \frac{\nu k}{\beta}(R_0 - 1) - \frac{1}{N} \end{aligned}$$

For large $N$, we have

$$S' = 1 - \frac{\nu k}{\beta}(R_0 - 1)$$

As defined by our initial set of ODEs, the growth rates of the mutant and resident strains are

$$\begin{aligned} \frac{dI'}{dt} &= I'[\beta S' - (\nu + \gamma)] \\ \frac{dI}{dt} &= I_{\text{eq}}[\beta S_{\text{eq}} - (\nu + \gamma)] \end{aligned}$$

To get the selection coefficient, we take the difference between the growth rates:

$$\begin{aligned} s &= [\beta S' - (\nu + \gamma)] - [\beta S_{\text{eq}} - (\nu + \gamma)] \\ &= \beta - \gamma k(R_0 - 1) - \frac{\beta}{R_0} \end{aligned}$$

Recall that $\beta = (\nu + \gamma)R_0$

$$s = (\nu + \gamma)R_0 - \nu k(R_0 - 1) - (\nu + \gamma)$$

Simplifying,

$$s = (\nu cd + \gamma)(R_0 - 1) \tag{S1}$$

Now disallowing coinfection, we have

$$\begin{aligned} S' &= 1 - R' - I_{\text{eq}} - I' \\ &= 1 - \frac{\nu}{\beta}(R_0 - 1)k - \frac{\gamma}{\beta}(R_0 - 1) - \frac{1}{N} \end{aligned}$$

For large $N$,

$$S' = 1 - (R_0 - 1)[\frac{\nu k + \gamma}{\beta}]$$

Using the same arithmetic as in the case with coinfection, it follows that

$$s = \beta - (\nu k + \gamma)(R_0 - 1) - \frac{\beta}{R_0}$$

Simplifying,

$$s = (\nu cd)(R_0 - 1) \tag{S2}$$

In summary, the selection coefficient of an invading mutant strain increases linearly with

the $R_0$, which is shared by both strains. The slope of this relationship is proportional to the distance $d$ between the two strains in antigenic space (figure 2.8). Naturally, relationship between the selection coefficient on the distance $d$ between strains depends on the functional relationship between antigenic distance and immunity. However, the linear dependence of the selection coefficient on $R_0$ holds as long as the functional relationship between antigenic distance and immunity is independent of $R_0$.

### *2.6.3 Detecting differences in $R_0$*

In the $SIR$ model, the force of infection is

$$F = \beta I$$

where $\beta$ is the transmission rate and $I$ is the fraction of infecteds. At the endemic equilibrium, the cumulative fraction of seropositive individuals at a given age $a$ is

$$\begin{aligned} f(a) &= 1 - \exp(-\beta I_{\text{eq}} a) \\ &= 1 - \exp(-R_0(\nu + \gamma I_{\text{eq}} a) \end{aligned}$$

where $\nu$ is the recovery rate and $\gamma$ is the birth/death rate. $I_{\text{eq}}$, the fraction of infecteds at the endemic equilibrium, is given by

$$I_{\text{eq}} = \frac{\gamma}{\beta}(R_0 - 1)$$

Figure 2.11 shows the fraction of seropositive individuals by age for the baseline $R_0 = 1.8$

and a 20% higher $R_0 = 2.16$. The difference in the percentage of seropositive two-year-olds between the two groups is approximately 2.3%. The sample size in each group required to detect a difference $f_2(a) - f_1(a)$ with $\alpha$ confidence and $1 - \beta$ power is

$$N = \frac{f_1(1 - f_1) + f_2(1 - f_2)}{(f_1 - f_2)^2}(\Phi_{\alpha/2} + \Phi_\beta)^2$$

For legibility, $f_i(a)$ is written as $f_i$. To detect a 20% difference in $R_0$ between two populations with 0.05 significance and 0.80 power, we would require a sample of at least 1503 individuals in both groups.

## 2.7 Supplemental tables and figures

Table 2.3: Parameter ranges used in Latin hypercube sampling

| Parameter | Range |
|---|---|
| Relative $R_0$ | 0.8−1.2 |
| Seasonal amplitude ($\epsilon$) in temperate populations | 0.0−0.15 |
| Relative population size (N) | 0.5−2.0 |
| Relative turnover rate ($\gamma$) | 0.5−2.0 |
| Fraction of initial infecteds ($I_0$) in tropics | 0.0−1.0 |

Table 2.4: ANOVA of the fraction of trunk in tropics from multivariate sensitivity analysis

| Parameter | Df | Sum Sq | Frac of var | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|---|
| Relative $N$ | 1 | 5.04 | 0.017 | 5.04 | 316.48 | <0.0001 |
| Fraction $I_0$ in tropics | 1 | 0.62 | 0.002 | 0.62 | 38.94 | <0.0001 |
| Relative $R_0$ | 1 | 114.49 | 0.406 | 114.49 | 7193.84 | <0.0001 |
| Relative turnover | 1 | 2.37 | 0.008 | 2.37 | 148.97 | <0.0001 |
| Seasonal amplitude | 1 | 94.04 | 0.334 | 94.04 | 5908.50 | <0.0001 |
| Residuals | 4109 | 65.40 | 0.232 | 0.02 | | |
| Total | 4114 | 281.96 | 1.000 | | | |

Table 2.5: ANOVA of the tropics' antigenic lead from multivariate sensitivity analysis

| Parameter | Df | Sum Sq | Frac of var | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|---|
| Relative $N$ | 1 | 1.94 | 0.033 | 1.94 | 754.71 | <0.0001 |
| Fraction $I_0$ in tropics | 1 | 0.02 | <0.001 | 0.02 | 8.75 | 0.0031 |
| Relative $R_0$ | 1 | 44.55 | 0.766 | 44.55 | 17344.41 | <0.0001 |
| Relative turnover | 1 | 1.04 | 0.018 | 1.04 | 406.81 | <0.0001 |
| Seasonal amplitude | 1 | 0.02 | <0.001 | 0.02 | 9.53 | 0.0020 |
| Residuals | 4109 | 10.56 | 0.182 | 0.00 | | |
| Total | 4114 | 58.140 | 1.000 | | | |

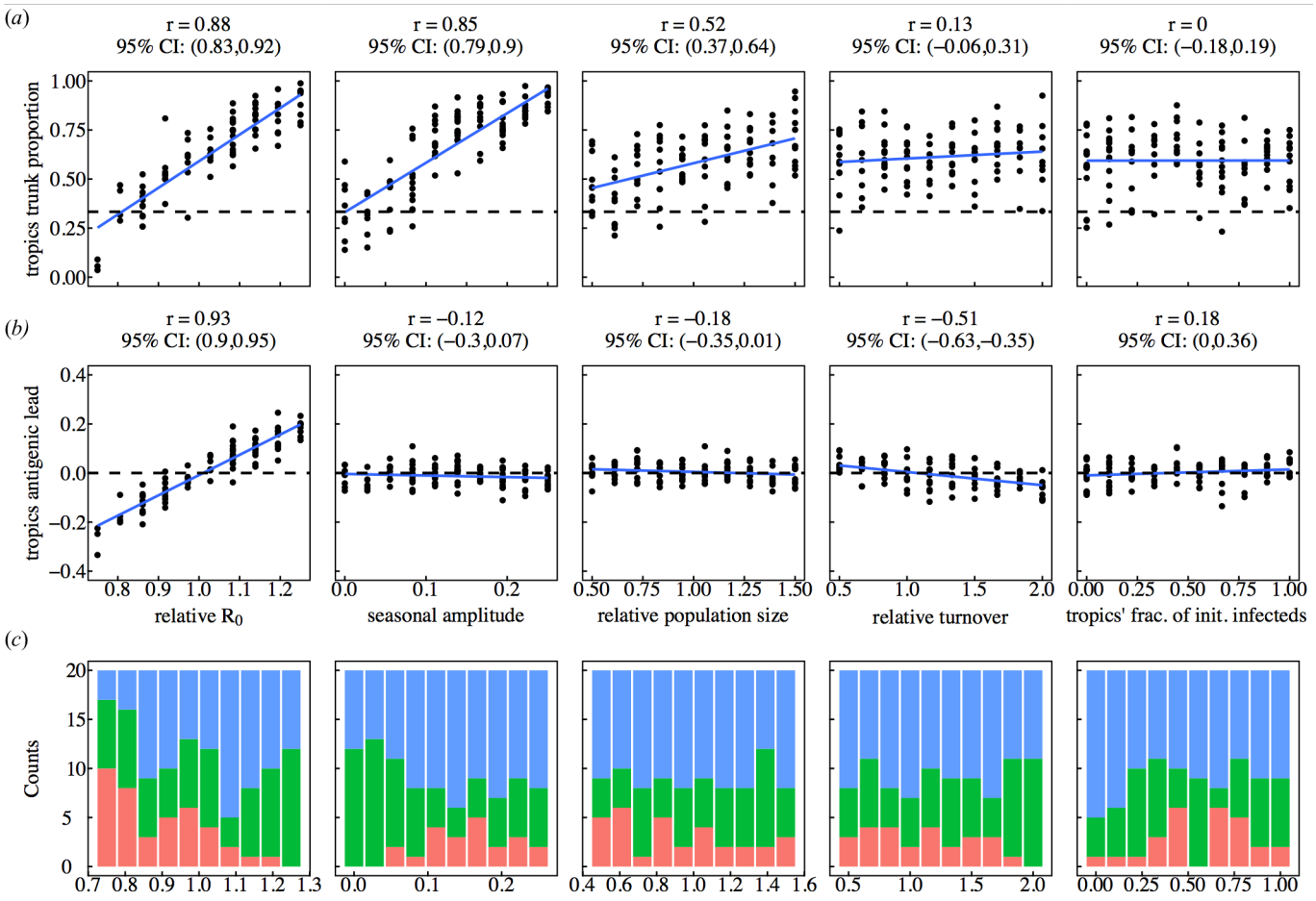


Figure 2.5: Univariate sensitivity analysis showing effects of individual parameters on ($a$) the antigenic lead and ($b$) the fraction of the phylogenetic trunk in the tropics. In each column of plots, only the parameter indicated on the x-axis is varying; all others are held constant at the default value. Each point represents the mean value over a single simulation. Blue lines indicate linear least squares regression. The dashed lines represent the null hypotheses where ($a$) the trunk is distributed equally among the three regions or ($b$) tropical strains are neither antigenically ahead or behind. ($c$) Number of simulations that went extinct (red), exceeded the TMRCA limit (green), or were suitable for analysis (blue).

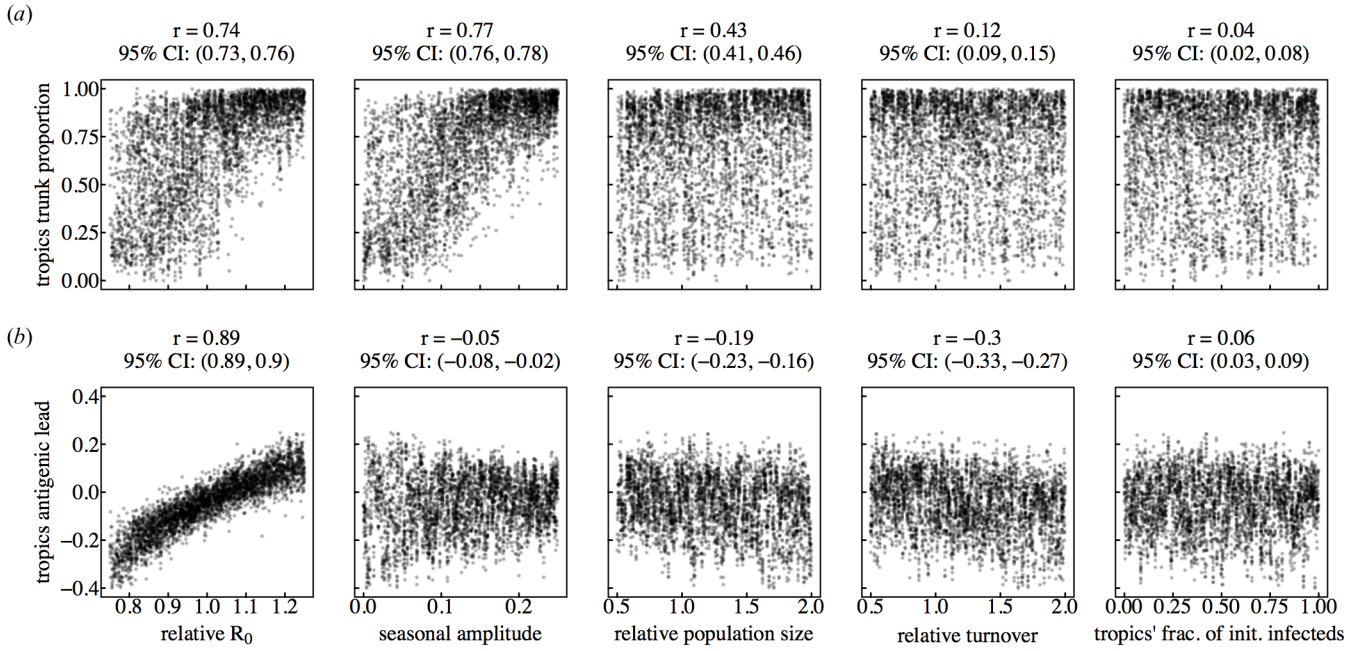


Figure 2.6: Multivariate sensitivity analysis showing effects of individual parameters on the $(a)$ antigenic lead and $(b)$ the fraction of the phylogenetic trunk in the tropics. Horizontal axes are projections of a Latin hypercube with dimensions corresponding to the five parameters indicated. Each point shows the mean value over a single simulation. The dashed lines represent the null hypotheses where $(a)$ the trunk is distributed equally among the three regions or $(b)$ tropical strains are neither antigenically ahead or behind. Pearson's correlation coefficients and associated 95% confidence intervals are indicated.

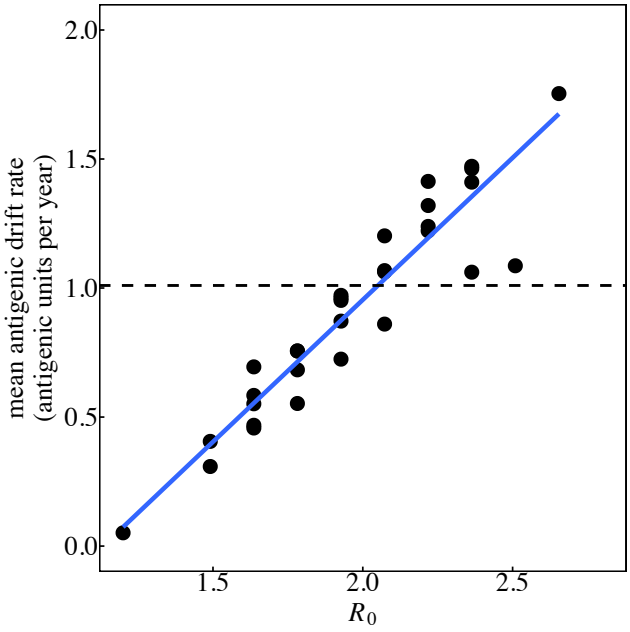


Figure 2.7: Effect of $R_0$ on antigenic drift in a single deme. Each point shows the mean antigenic drift rate from a single simulation. The blue line represents linear least squares regression, and the dashed line indicates the empirical estimate of the rate of antigenic drift for H3N2 [13]. Pearson's $r = 0.94$, $p < 0.001$; 95% CI: (0.88, 0.97).

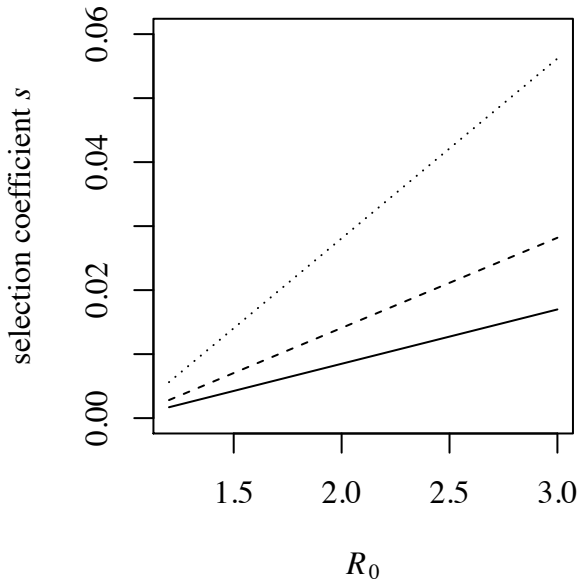


Figure 2.8: Relationship between $R_0$ and the selection coefficient for an invading strain with the resident at endemic equilibrium. The relationship for three different antigenic distances $d$ between the invading strain and the resident strain is shown: 0.6 (solid line), 1.0 (dashed line), or 2.0 (dotted line) antigenic units.

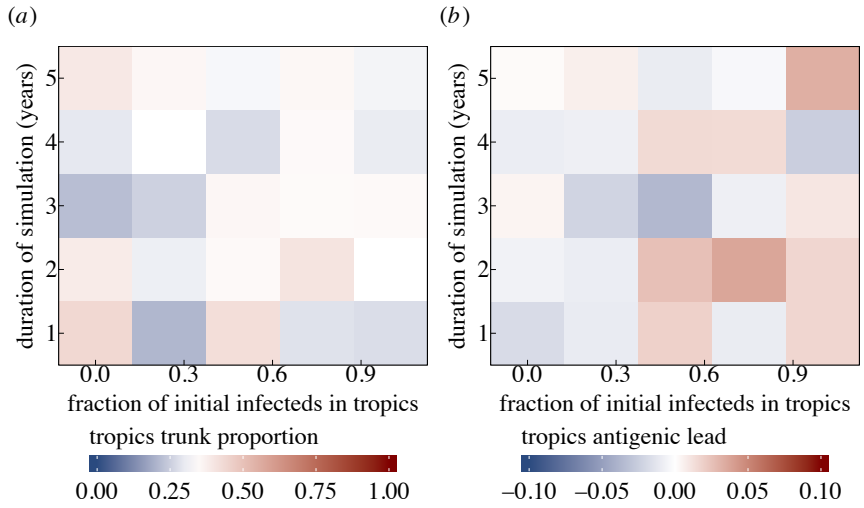


Figure 2.9: Effect of initial conditions on the antigenic lead and the fraction of the trunk in the tropics early in the simulation. Each square represents the average value for n=5 to 11 replicate simulations.

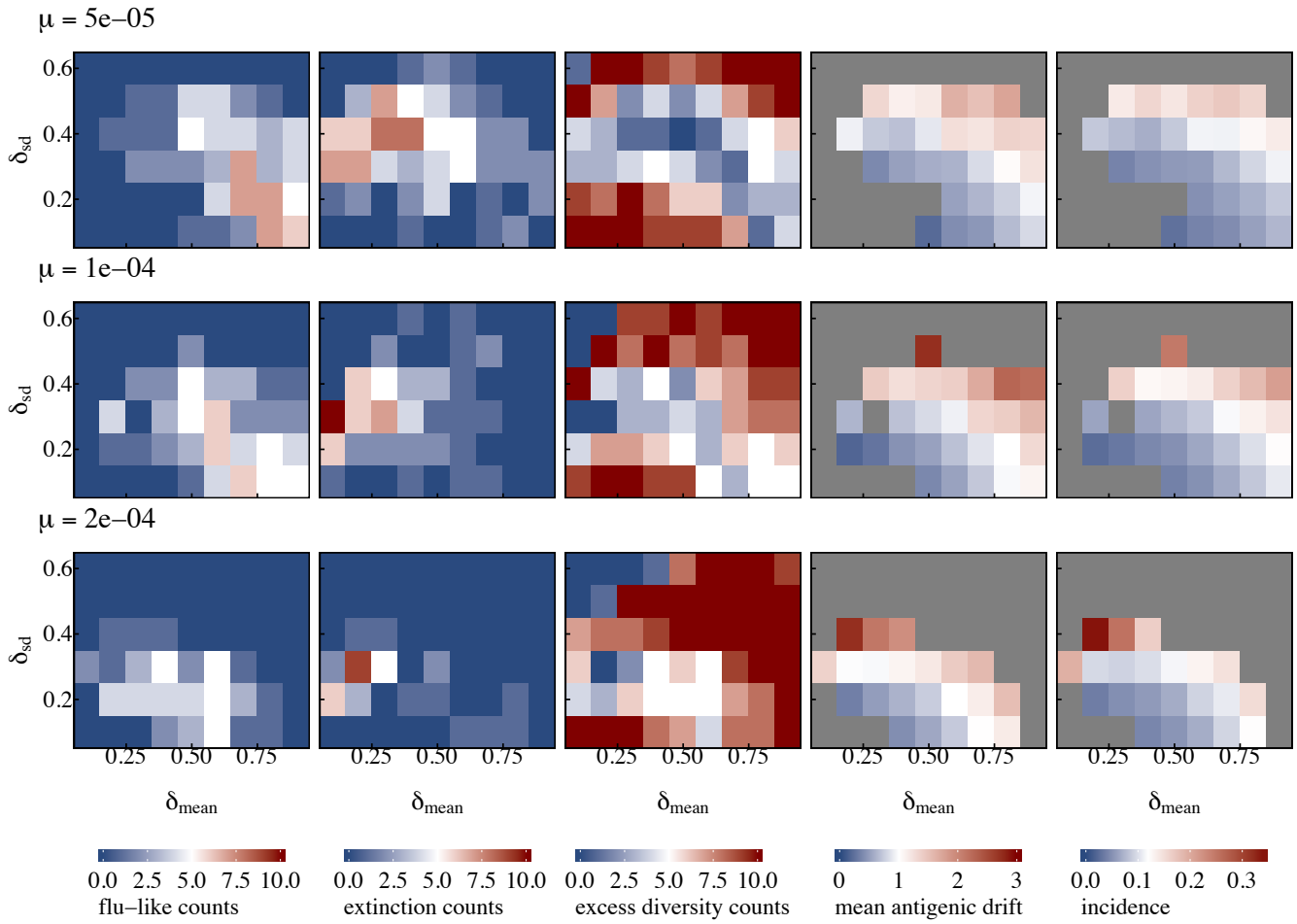


Figure 2.10: Sensitivity of influenza-like behaviour to changes in the mutational parameters, the mutation rate $\mu$, mean mutation size $\delta_{\text{mean}}$, and standard deviation of the mutation size $\delta_{\text{sd}}$. Within each plot, each square represents ten replicate simulations. Each row of plots shows results from simulations using different mutation rates $\mu$. The number of simulations where the virus went extinct is shown in the second column of plots, and the number of simulations where the viral population exceeded a TMRCA of 10 years is shown in the third column of plots. The remaining simulations are considered influenza-like and are shown in the first column of plots. The reported mean antigenic drift rates and prevalences are averaged over the influenza-like replicates. The color scales for mean antigenic drift and incidence are centered (white) at the observed values for H3N2 (table 1).

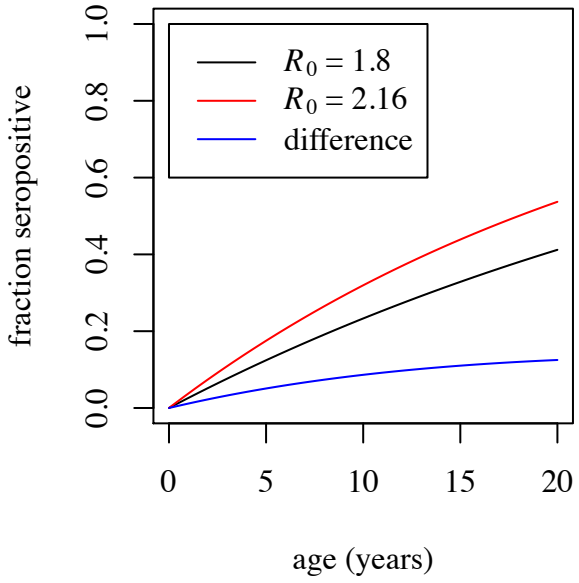


Figure 2.11: Theoretical increase in the fraction of seropositive individuals with age with $R_0 = 1.8$ and a 20% higher $R_0 = 2.16$.

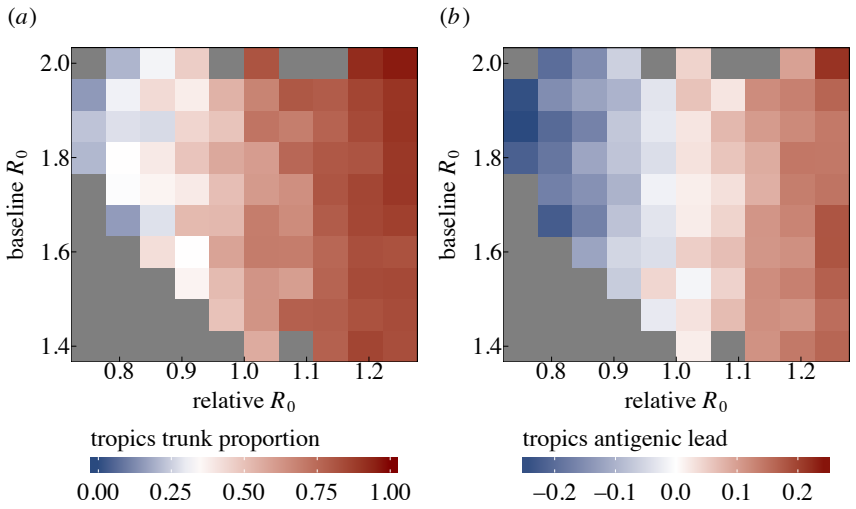


Figure 2.12: Lowering baseline $R_0$ decreases the effect of relative $R_0$ on the fraction of the trunk and antigenic lead in the tropics. ($a$) Effects of baseline and relative $R_0$ on the fraction of the trunk in the tropics. Blue indicates that the phylogenetic trunk is located in the tropics less than 1/3 of the time, and red indicates that the trunk is the tropics more than 1/3 of the time. ($b$) Effects of baseline $R_0$ and relative $R_0$ on antigenic lead in the tropics. Blue indicates that tropical strains are on average ahead antigenically relative to other strains, and red indicates that tropical strains are behind antigenically. Each square represents an average from 1 to 14 replicate simulations. Grey squares indicate parameter combinations where all of twenty attempted simulations either went extinct or exceeded the TMRCA threshold of 10 years.

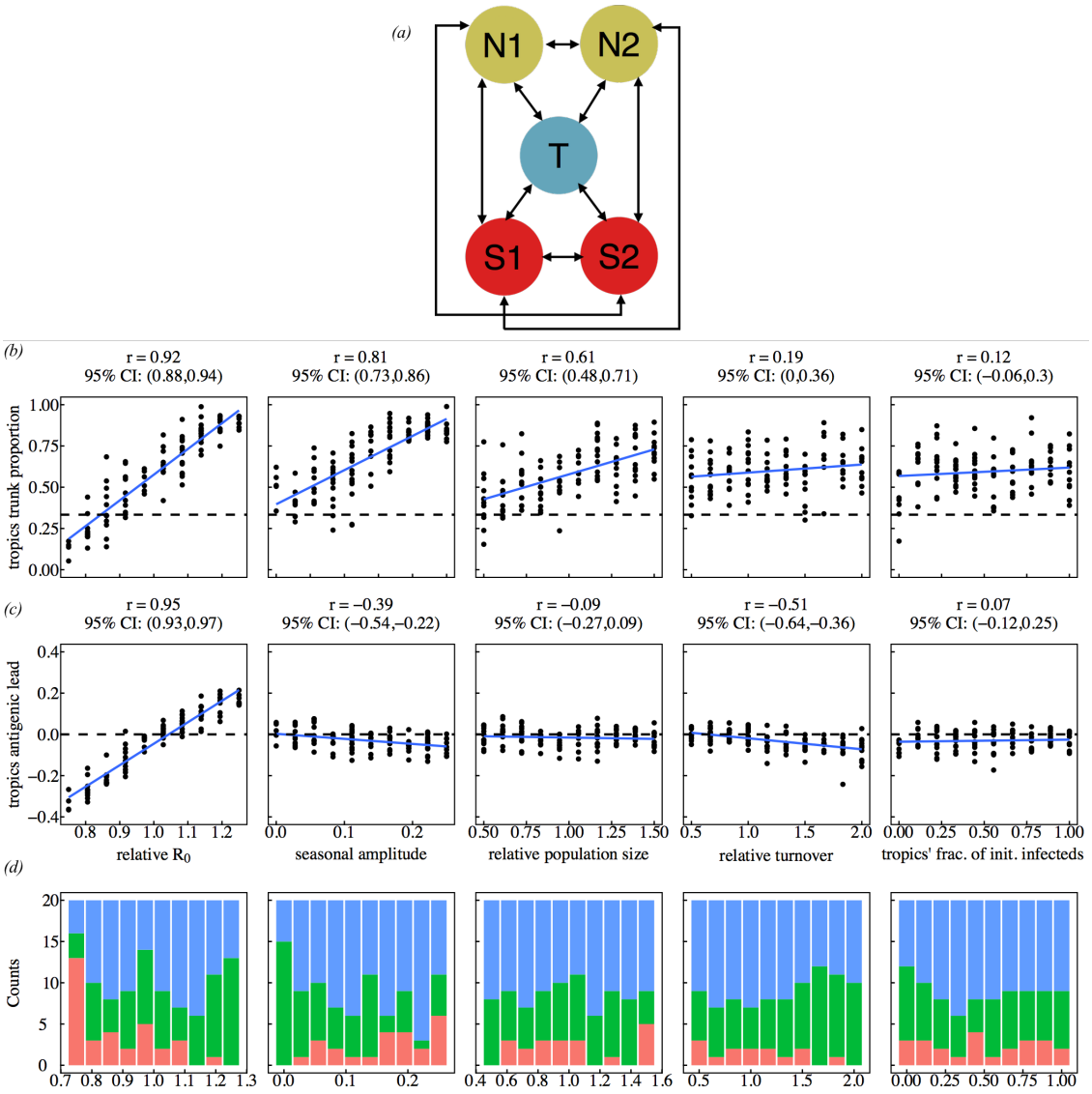


Figure 2.13: Univariate sensitivity analysis using a fully connected 5-deme model ($a$) showing the effects of individual parameters on ($b$) the antigenic lead and ($c$) the fraction of the phylogenetic trunk in the tropics. By default, the tropics have a population size that is twice as large as any single temperate deme. In each column of plots, only the parameter indicated on the x-axis is varying; all others are held constant at the default value. Each point represents the mean value over a single simulation. Blue lines indicate linear least squares regression. The dashed lines represent the null hypotheses where ($b$) the trunk is distributed proportionally to the default population size among the regions or ($c$) tropical strains are neither antigenically ahead or behind. ($d$) Number of simulations that went extinct (red), exceeded the TMRCA limit (green), or were suitable for analysis (blue).

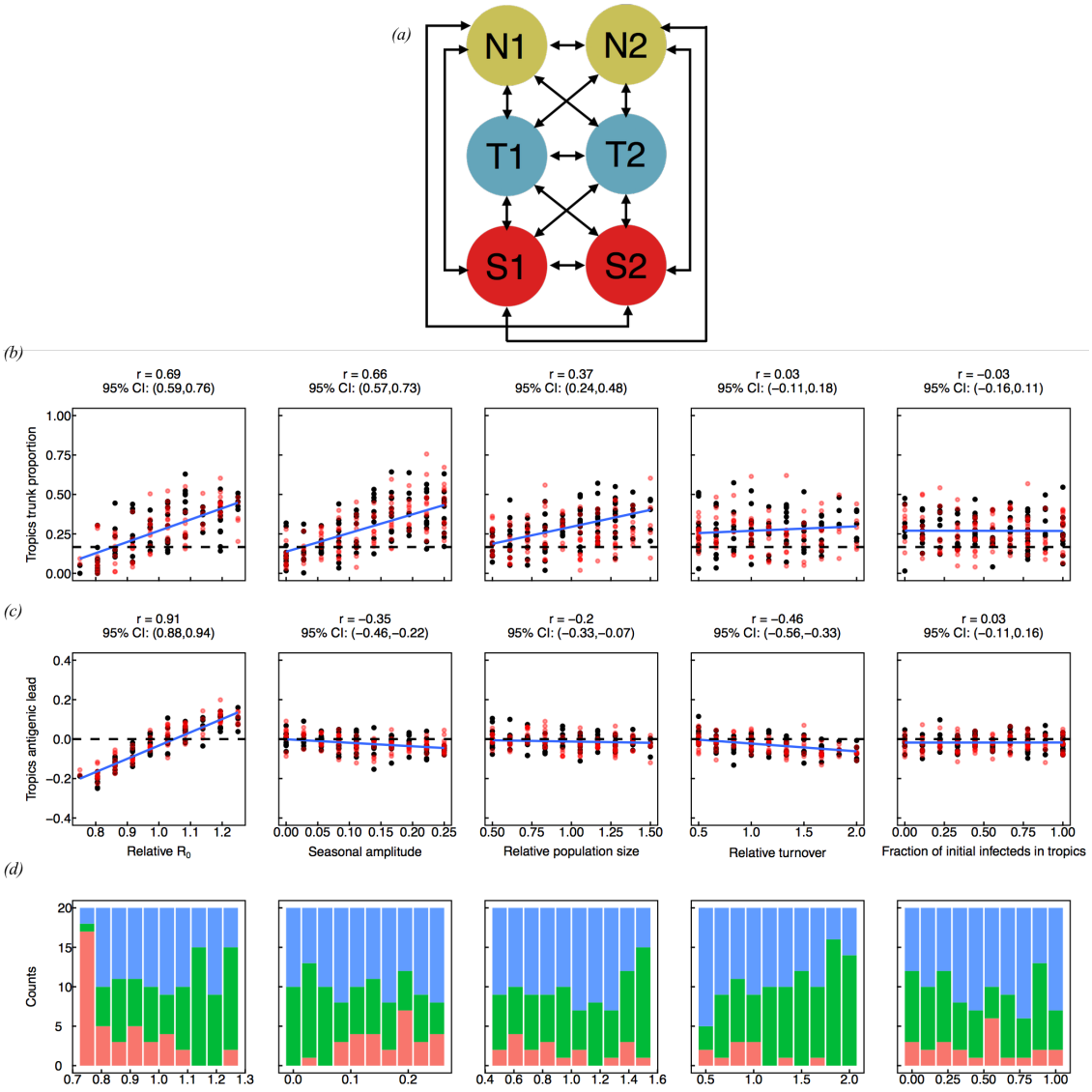


Figure 2.14: Univariate sensitivity analysis using a fully connected 6-deme model ($a$) showing the effects of individual parameters on ($b$) the antigenic lead and ($c$) the fraction of the phylogenetic trunk in each of the two tropical demes. By default, all demes have the same population size. In each column of plots, only the parameter indicated on the x-axis is varying; all others are held constant at the default value. Each point represents the mean value over a single simulation. Black points show results from one tropical deme and red points from the other. Blue lines indicate linear least squares regression to the combined data from both tropical demes. The dashed lines represent the null hypotheses where ($b$) the trunk is distributed proportionally to the default population size among the regions or ($c$) tropical strains are neither antigenically ahead or behind. ($d$) Number of simulations that went extinct (red), exceeded the TMRCA limit (green), or were suitable for analysis (blue).

# CHAPTER 3
# VACCINATION AND THE EVOLUTION OF SEASONAL INFLUENZA

## 3.1 Introduction

As seasonal influenza evolves from year to year, antigenic differences between previously and currently circulating strains contribute to low vaccine efficacy [87, 26, 15, 192] and high incidence of influenza illness [29, 26]. The influenza A/H3N2 subtype evolves faster than influenza A/H1N1 and B [13], and the vaccine is least effective against A/H3N2 on average compared to other circulating subtypes [16]. While vaccines regularly undergo reformulation to accommodate antigenic evolution, it is also theoretically possible for vaccines to affect antigenic evolution [100]. Traditional estimates of the public health benefits of influenza vaccines tend to focus on the benefits of vaccination in the current season and assume viral evolution is unchanged by the vaccine [7, 183, 123, 146]. Accounting for the potential evolutionary impact of vaccines, however, may alter assessments of their long-term value.

In theory, seasonal influenza vaccines might be able to slow antigenic evolution [80, 101, 5]. Universal vaccines, which confer immunity against all antigenic variants, are predicted to slow antigenic evolution by uniformly decreasing the fitness of all strains [5]. Conventional vaccines against seasonal influenza, which protect against some strains more than others and thereby confer narrower immunity, might have similar effects. First, by reducing the prevalence of infection, they reduce viral population size and thus the probability that antigenic escape mutants will arise. Second, although vaccination increases the growth rate of antigenically distant mutants relative to less distant mutants (which can lead to strain replacement in other pathogens [28, 114, 132, 105, 65, 117, 1, 138, 66, 73, 78]), it also increases the amount of immunity in the population. This increased immunity reduces the growth rate or invasion fitness of escape mutants, slowing the rate of strain replacement (SI 3.9.1, Eq.

3.25, Fig. 3.5). Finally, smaller viral population sizes increase the rate at which different strains go stochastically extinct, weakening selection for more antigenically diverged strains. However, vaccination might *accelerate* antigenic evolution if the vaccine is ineffective against some strains that compete with vaccine-targeted strains, leading to strain replacement or vaccine escape [23, 162].

Vaccination's potential evolutionary effects may change the private and social benefits of vaccination. Vaccination confers a private benefit to vaccinated individuals by directly reducing their risk of infection: the vaccine reduces the within-season rate of clinical laboratory-confirmed influenza infections in healthy adult recipients by 41% (95% CI 36-47%) [58]. Vaccination also confers a social benefit to the host population by reducing the burden of disease, although these effects are infrequently measured. Vaccinating children reduces the risk of influenza infection in unvaccinated household contacts by 30-40% [92, 139], in the local community by up to 5-82% [116], and in a metropolitan county by up to 59% [136]. The valuation of private and social benefits changes according to how much vaccination decreases the burden of disease. If vaccines slow antigenic evolution and thereby further decrease incidence, then the social benefit increases. However, the private benefit may fall as the lower infection risk reduces vaccines' marginal protective benefit. As the private benefit falls, additional incentives might be necessary to compensate for less frequent voluntary vaccination [25, 39]. A reduction in antigenic evolution from vaccination could also reduce the need to update vaccines as frequently.

Empirical estimates of the benefits of vaccination have so far been unable to measure the potential long-term evolutionary effects of vaccination. Most studies estimating the value of vaccination occur in temperate populations such as North America, Europe, and Oceania, which have high vaccine coverage but do not consistently contribute to influenza's long-term evolution [134, 10, 12, 131, 16]. By contrast, source populations that contribute more to influenza's evolution (e.g., China and India) have almost zero vaccination [134, 10, 12], and

few studies of vaccination occur there [164].

We consider here the consequences of an idealized vaccination strategy, where vaccination occurs in populations that shape influenza's long-term evolution. To assess the potential effects of vaccines on antigenic evolution, we simulated the evolutionary and epidemiological dynamics of an influenza-like pathogen. We evaluated how different rates of vaccination may slow antigenic evolution and in turn decrease the total burden of disease. We then quantified how the evolutionary effects change the relative magnitude of the private and social benefits of vaccination in the short and long term.

## 3.2 Results

### *3.2.1 Modeling approach and choice of parameters*

We adapted a model to simulate the transmission and evolution of an influenza-like pathogen over 20 years in a well-mixed population (Methods) [11]. Individuals infected with a strain of the virus can transmit their infection to susceptible individuals upon contact. The risk of infection given contact depends on the antigenic identities (phenotypes) of previous infections and the challenging strain. After recovering from infection, individuals acquire immunity against the infecting strain, whose antigenic phenotype is represented by a point in two-dimensional Euclidean space (Fig. 3.1A). Geometrically distributed mutations displace strains in this space (Table 3.1, Fig. 3.1D). This space is analogous to the main components after multidimensional scaling of pairwise measurements of cross-reactivity in hemagglutination inhibition (HI) assays, where one antigenic unit of distance represents a twofold dilution of antiserum [160, 13]. Each antigenic unit difference in distance between strains increases susceptibility by 7% (Fig. 3.1C) [135, 87, 11].

The model reproduces characteristic epidemiological and evolutionary patterns of the A/H3N2 subtype in the absence of vaccination (Fig. 3.1A,B). Unvaccinated populations

are best for model validation because they contribute most to the evolution of seasonal influenza in reality [134, 12]. We chose transmission and mutation parameters (Table 3.1) such that simulated epidemiological and evolutionary patterns most resembled qualitative patterns observed for H3N2 [181]. H3N2 has remained endemic in the human population since its emergence in 1968 and also has low standing genetic and antigenic diversity. Due to the stochastic nature of the simulations, the viral population goes extinct 18% of the time and becomes too diverse 29% of the time across replicate simulations. A viral population is considered too diverse when the time separating two co-circulating lineages (time to most recent common ancestor, or TMRCA) exceeds 10 years, since recent H3N2 HA lineages have coexisted for no more than 7 years. The remaining 53% of simulations that show qualitatively influenza-like dynamics reproduce epidemiological and evolutionary statistics of H3N2. The viral population has low genealogical diversity with an average TMRCA across replicates of 3.80 years (SD = 0.52), comparable to empirical estimates of 3.84 years [12]. The path of evolution in antigenic space is mostly constrained to one dimension (Fig. 3.1A), characteristic of H3N2's antigenic evolution [160, 13]. Antigenic evolution occurs at an average rate of 1.09 antigenic units per year (SD = 0.14), comparable to an observed rate of 1.01 antigenic units per year [13]. The mean annual incidence is 9.0% (SD = 1.0%). Annual incidence across all types of seasonal influenza ranges from 9-15% [184]. To confirm the accuracy of the model's transmission dynamics, we compared model outputs against analytic expectations without evolution (since analytic solutions for a model with evolution are intractable) (Figs. 3.6, 3.7, 3.8, and 3.9).

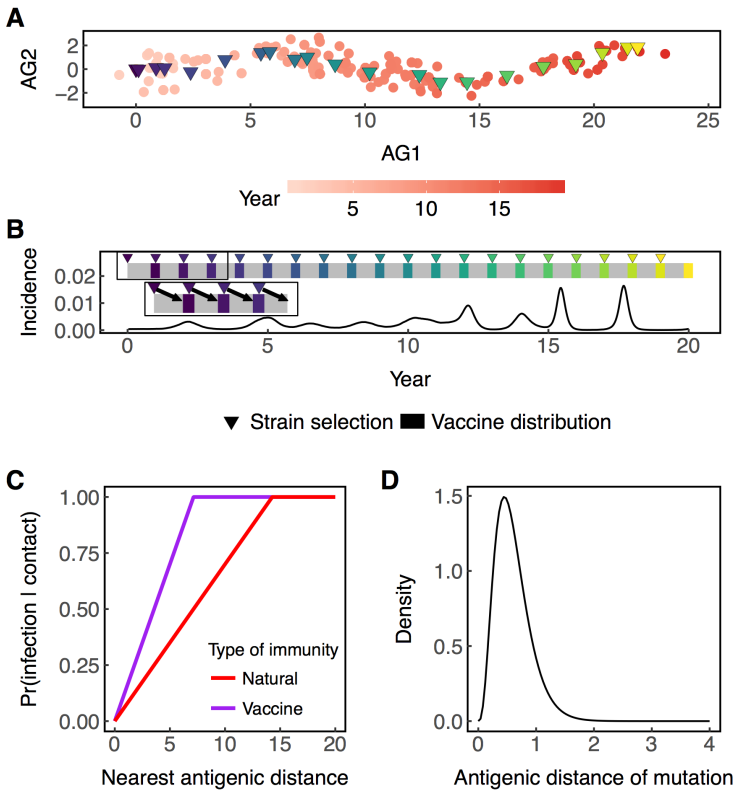


Figure 3.1: Properties of the model. (A) Antigenic phenotypes are represented as points in two-dimensional space (AG1 is antigenic dimension 1 and AG2 is antigenic dimension 2). Over time, new strains appear as old strains can no longer transmit to immune hosts. Viral evolution is mostly linear in antigenic space. The amount of evolution is calculated as the distance between the founding strain and the average phenotype of strains circulating at the end of the simulation. Vaccine strains (triangles) are chosen at the beginning of each year by averaging the antigenic phenotype of all circulating strains. Strains are colored according time. (B) Incidence per 10 days is shown. Cumulative incidence (not shown) is calculated as the sum of cases over the duration of the simulation. Vaccines are distributed beginning 300 days after strain selection for 120 days. Strain selection for the following year occurs during the distribution of the current vaccine (inset). (C) Upon contact, the risk of infection increases linearly with the distance between the infecting strain and the strain in the host's infection or vaccination history that minimizes the risk of infection (Eq. 3.3) (D) The sizes of antigenic mutations are chosen from a gamma distribution with mean and standard deviation $\delta_{\text{mean}}$ and $\delta_{\text{sd}}$. The radial directions (not pictured) of mutations are chosen from a random uniform distribution. In this example, vaccines confer half the breadth of immunity as natural immunity ($b = 0.5$).

To assess the potential effects of vaccination on antigenic evolution and disease burden, we introduced vaccination to the host population. At the beginning of each year, a vaccine

strain is selected with the average antigenic phenotype of circulating strains. In the United States, the seasonal influenza vaccine is typically distributed from September through February. Distribution usually peaks in October or November, 8-9 months after strain selection [32]. In the model, the vaccine is distributed 300 days after strain selection and for a period of 120 days. During distribution, individuals are randomly vaccinated at a constant daily rate (Eq. 3.2). Since individuals are randomly vaccinated each year, the fraction of vaccinated individuals over time. At a 5% annual vaccination rate, approximately 4.9% of individuals in the population are vaccinated every year (due to sampling with replacement in the model) and 48.4% of the population has been vaccinated at least once by the twentieth year (Fig. 3.10A). At this rate, vaccination effectively renders 26.0% of individuals immune when vaccination is in equilibrium with antigenic evolution (Fig. 3.10B). We also tested the effects of the breadth of cross-immunity conferred by vaccination. The vaccine's breadth $b$ is defined as the ratio of the vaccine-induced immunity to that of infection-induced (or "natural") immunity (Fig. 3.1). Vaccines with $b = 1$ have breadth identical to natural immunity, whereas vaccines with $b < 1$ ($b > 1$) have respectively smaller (larger) breadth compared to natural immunity.

We initially used two metrics to quantify the effects of vaccination on the evolution and epidemiology of the virus. First, because antigenic phenotypes evolve roughly linearly in two dimensions [160, 11, 13], we measured the cumulative amount of antigenic evolution by calculating the antigenic distance between the founding strain's antigenic phenotype and the average antigenic phenotype of strains circulating at the end of the simulation (Fig. 3.1). Second, we measured the burden of disease by calculating the cumulative incidence, or the total number of cases over the duration of the simulation divided by the population size (Fig. 3.1). In calculating the amount of antigenic evolution and incidence, we included simulations where the viral population remained endemic or went extinct. However, we excluded simulations where the viral population became too diverse (TMRCA $>$ 10 years)

because our measure of cumulative antigenic evolution is inadequate for branching viral populations.

Because vaccination may qualitatively alter evolutionary patterns of H3N2, we used an additional metric to asses evolutionary effects, namely the probability that viral populations would become too diverse (TMRCA $>$ 10 years) under different vaccination regimes. Viral populations that are too diverse have the potential to cause high morbidity because hosts are unlikely to have immunity against many antigenic variants. Influenza subtypes H1N1 and B evolve antigenically slower than H3N2 but have greater genetic diversity at any time [40, 52, 13, 12] Thus, we also examine whether vaccination, by affecting antigenic evolution, could also impact diversification.

To estimate the contribution of evolution to vaccination's epidemiological impact, we compared simulations in which vaccination could affect antigenic evolution to simulations where it could not. We generated the latter by first running simulations without vaccination and recording strain phenotypes and relative abundances at every time step to use as a reference. Then, in each time step of the simulations with vaccination, we replaced all infections with randomly selected contemporaneous strains from an unvaccinated reference simulation, matching the reference frequencies. In this way, temporal changes in strain frequencies were unaffected by vaccination.

### *3.2.2 Vaccination reduces the average amount of antigenic evolution and disease burden*

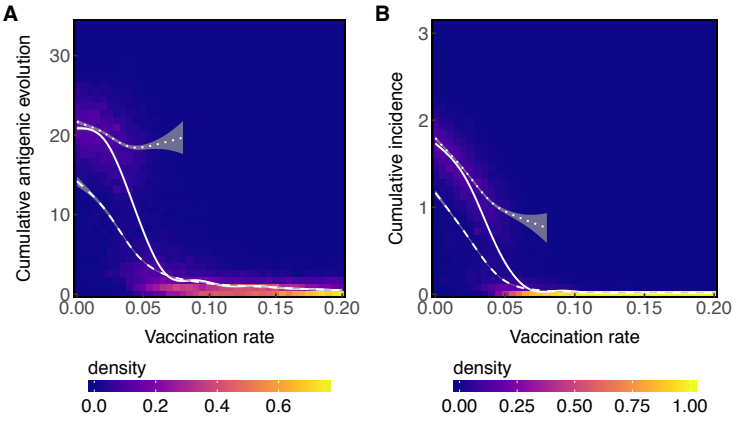


Figure 3.2: High vaccination rates decrease the average amount of (A) cumulative antigenic evolution and (B) cumulative incidence. The solid white lines show LOESS curves fit to cumulative antigenic evolution and incidence across all simulations. The dotted white lines show fits for simulations where the viral population survived until the end of the simulation. The dashed white lines show fits for simulations where the viral population went extinct. Shaded areas show 95% confidence intervals. Densities reflect 500 total simulations for each vaccination rate with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300 - 400$ simulations.

Vaccination reduces the average amount of antigenic evolution (Spearman's $\rho = -0.75$, $p < 0.001$) and incidence (Spearman's $\rho = -0.86$, $p < 0.001$, Fig. 3.2) when the breadth of vaccine-induced immunity is the same as that of infection. Without vaccination, the viral population evolves on average 21.5 (SD = 3.3) antigenic units and causes an average of 1.8 (SD = 0.2) cases per person over the 20-year simulation. By reducing susceptibility in the host population, vaccination decreases the number of cases and the average size of surviving mutations, thus slowing the rate of antigenic evolution. In turn, slower antigenic evolution further reduces the force of infection, often driving the virus extinct. Once extinct, the viral population can no longer evolve or cause new infections. Above a 10% annual vaccination rate, implying a 28% cumulative vaccination rate over 4 years, extinction occurs rapidly, typically within 2.3 years (SD = 0.6, Fig. 3.11). Eliminating the time interval

between strain selection and vaccine distribution reduces the amount of antigenic evolution (Wilcoxon rank-sum test, $p < 0.001$) and incidence (Wilcoxon rank-sum test, $p < 0.001$) even more (Fig. 3.12).

Increasing the vaccination rate also decreases the probability that the viral population becomes too diverse (TMRCA $>$ 10 years on average, Fig. 3.13). Thus, vaccination is unlikely to increase morbidity from diversifying viral populations.

We next examined how much these reductions could be attributed solely to the "ecological" effects of vaccination—the reduction in prevalence and increased extinction risk from enhanced herd immunity—versus the combined ecological and evolutionary impacts. Relative to the case where the evolutionary effects of vaccination are blocked, vaccination with evolutionary effects decreases both the rate of antigenic evolution and the burden of disease (Wilcoxon rank-sum test, $p < 0.001$), (Fig. 3.3). Also relative to the same baseline, eradication is achieved at a lower vaccination rate. At an 8.5% annual vaccination rate ($\sim$ 20% cumulative vaccine coverage within 5 years), vaccination eradicates the virus 100% of the time (within 3.3 years on average) when vaccines can affect evolution but only does so 68% of the time (within 5.6 years on average) when vaccines cannot affect evolution.

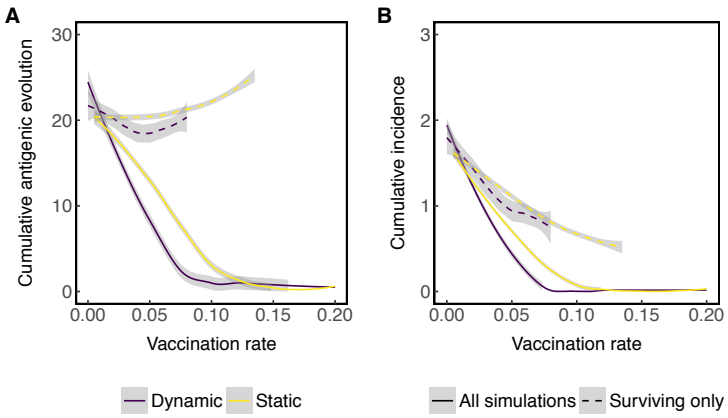


Figure 3.3: Vaccination further decreases incidence when vaccines can affect antigenic evolution compared to when they cannot. Purple lines represent simulations where vaccination can affect antigenic evolution. Yellow lines represent simulations where vaccination cannot affect antigenic evolution. The solid lines show LOESS fits to cumulative (A) antigenic evolution and (B) incidence across all simulations. The dotted lines show LOESS fits for simulations where the viral population does not go extinct. Shaded areas show 95% confidence intervals. Lines reflect 500 total simulations for each vaccination rate and evolutionary condition with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300 - 400$ simulations.

The breadth of vaccine-induced immunity and the delay between vaccine strain selection and distribution change the impact of vaccination. With narrower vaccines, higher vaccination rates are needed to achieve the same average reductions in cumulative antigenic evolution and incidence using broader vaccines (Fig. 3.14). Regardless of breadth, distributing vaccines immediately after strain selection helps vaccines achieve the same average reductions in evolution and incidence at lower vaccination rates (Fig. 3.16).

### *3.2.3 Vaccine-driven excessive evolution is rare*

We developed a test to determine whether vaccination causes excess evolution. We defined excess evolution as more than 21 antigenic units (the average amount of evolution without vaccination) over the duration of the simulation, or when the TMRCA exceeded 10 years. We counted the number of "excessively evolved" simulations for each vaccination rate and

breadth. If vaccination does not affect the rate of evolution, the frequency of excessively evolved simulations should be the same as in vaccine-free case (Fig. 3.18). In contrast, if vaccination increases the rate of evolution, the frequency of excessively evolved simulations should be greater than without vaccination.

Although viral populations that survive are associated with more evolution (Figs. 3.2, 3.3, 3.14), this apparent excess evolution is generally not caused by vaccination. Instead, these viral populations evolved just as much in the absence of vaccination, and only survive vaccination because they evolved unusually quickly. In these cases, more vaccination does not increase the rate of antigenic evolution, but instead drives slowly evolving viral populations extinct while occasionally allowing persistence of quickly evolving populations (Fig. 3.18). Thus, apparent increases in the amount of antigenic evolution among survivors generally reflect selection among simulations (not among viruses within a simulation) for fast-evolving populations, and these populations would appear at the same rate without vaccination.

We found that vaccine-driven excess evolution was only possible at low-intermediate immune breadth ($b = 0.2$ or 0.3) and at low vaccination rates (Fig. 3.17). Even when we detected statistically significant excess evolution, these outcomes were only 10% more common with vaccination relative to without. Based on this analysis, we conclude that vaccine-driven excessive evolution is rare for the influenza-like parameters considered.

### *3.2.4 Ignoring the evolutionary effects of vaccination overestimates the private benefit and underestimates the social benefit of vaccination*

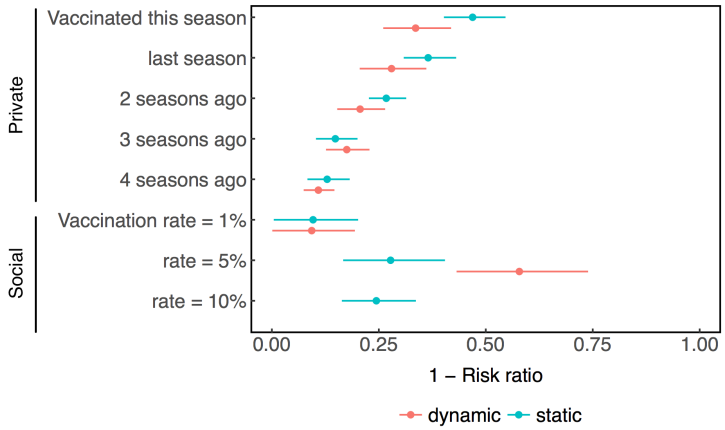


Figure 3.4: Comparison of the private and social benefits of vaccination when vaccination can or cannot affect antigenic evolution. Risk ratios are calculating using coefficients from a linear panel model fitted to the last 17 years of simulated hosts' infection and vaccination histories. Mean estimates and 95% confidence intervals are shown. Red lines represent simulations where vaccination can affect antigenic evolution (dynamic). Blue lines represent simulations where vaccination cannot affect antigenic evolution (static). The relative risk for a population with a 10% annual vaccination rate could not be calculated because all simulations were driven extinct within the first 3 years.

To quantify the private and social benefits of vaccination, we collected panel data consisting of individual hosts' vaccination and infection histories from simulations where vaccination could affect antigenic evolution and simulations where vaccination could not affect antigenic evolution. We then fit linear panel models to these data (Eq. 3.4). We measured the private benefit of vaccination as vaccine efficacy, or one minus the risk of infection having been vaccinated relative to the risk of infection having not been vaccinated (Eq. 3.5). To measure the social benefit, we used an analogous risk ratio. The social benefit is one minus the risk of infection in a population vaccinated at a given rate relative to the risk of infection in an unvaccinated population (Eq. 3.6). The social benefit reflects a reduction in the force of infection due to vaccination.

The social benefit of vaccination rises when vaccines can slow antigenic evolution compared to when evolutionary effects are omitted. The average risk of infection over the course of a season without vaccination is ∼10% (Table 3.3). When 5% of the host population is vaccinated annually, the average host is 60.5% less likely to become infected compared to a host in an unvaccinated population (Fig. 3.4, Table 3.3). However, when vaccination cannot affect antigenic evolution, the average host is only 27.7% less likely to become infected (Fig. 3.4, Table 3.3) at the same vaccination rate relative to a host in an unvaccinated population. The social benefits accounting for evolution at 10% vaccination rate could not be calculated because the virus was always eradicated quickly.

Since the evolutionary effects of vaccination further reduce the overall risk of infection in the population, individuals personally benefit less from getting vaccinated when vaccines affect antigenic evolution than when vaccines do not. The reduction in the private benefit due to evolutionary effects is a natural consequence of lower incidence: when the overall risk of infection is low, the marginal benefit of vaccination is lower than when incidence is high (Eq. 3.5). Individuals receiving the current vaccine are 36.0% less likely to become infected in the same season compared to unvaccinated individuals when vaccines can affect evolution (Fig. 3.4, Table 3.3). However, when vaccines cannot affect antigenic evolution, vaccinated individuals are 49.5% less likely to become infected (Fig. 3.4, Table 3.3). We observed similar patterns when the breadth of vaccine-induced immunity was half that of natural immunity (Table 3.4).

By slowing antigenic evolution, vaccination prolongs its own effectiveness. When vaccination cannot affect antigenic evolution, the private benefit decreases by 9.0% per passing year compared to only 5.6% per passing year when vaccines can affect evolution (Fig. 3.4, Table 3.3). Thus, evolutionary effects cause the private benefits of vaccination to decay slower with time. Consequently, ignoring the evolutionary effects of vaccines also undervalues the long-term private benefits relative to the short-term private benefits.

## 3.3 Discussion

We found that vaccination against seasonal influenza could hypothetically slow antigenic evolution and thereby reduce the disease burden beyond its immediate impact on transmission. Indeed, annual vaccination rates as low as 10%, which imply a 28% cumulative vaccine coverage after 4 years, can reliably eradicate the virus in simulation. This is a previously unrecognized potential benefit of widespread vaccination. At a 5% annual vaccination rate (16% cumulative coverage after 4 years), evolution increases the social benefits of vaccination by 30.4%, which in turn decreases the private benefits by 13.5% compared to when evolutionary effects are omitted. Thus, while the evolutionary effects of vaccination yield a large social benefit by reducing incidence, they reduce the private benefit to vaccinated individuals.

Though our simulations suggest that a 10% annual vaccination rate could eradicate influenza, this prediction may not appear realistic since up to 8% of the global population is vaccinated each year [134]. However, vaccination is almost exclusively concentrated in seasonal populations rather than in the populations that contribute most to influenza's evolution [134, 10, 12]. For instance, from the 2008-2009 season to the 2014-2015 season, seasonal vaccine coverage averaged 43.4% in the United States and 13.5% across European countries, but was <1% in China and India [133, 134]. Moreover, the same people tend to get vaccinated repeatedly, which lessens the accumulation of vaccine-induced immunity in the population over time. In the United States, up to 68.4% of vaccine recipients get vaccinated every year [72, 173]. Consecutive vaccinations may also reduce vaccine effectiveness by interacting with prior immune responses, although these effects are not well understood [129, 121, 149, 108]. Thus, the effective amount of vaccine-induced immunity in a population is potentially lower than vaccine coverage estimates would suggest, implying higher vaccination rates might be necessary for eradication.

The seasonal influenza vaccine is unlikely to cause excessive evolution, assuming that

the breadth of vaccine-induced immunity is similar to that of natural immunity. In simulations, vaccine-driven accelerated antigenic evolution only occurs when the breadth of vaccine-induced immunity is narrower than that of natural infection and then only at low vaccination rates. The relative breadths of vaccine-induced and and natural immunity are unclear. One difference is that although natural infection elicits antibodies that bind both the hemagglutinin and the neuraminidase (NA) antigens, inactivated vaccines may induce fewer antibodies to NA [41], suggesting that the breadth of vaccine-induced immunity could be narrower than that of natural immunity. Host immune history also affects the generation of immune responses [49, 194, 112, 55, 56], and by extension the breadths of vaccine-induced and natural immunity, in ways that are largely unexplored.

Although our simulations show vaccines typically slow evolution (and drive extinction) in a single, closed population (i.e., a global population), other models predict faster evolution or higher incidence under particular assumptions. Vaccination accelerates antigenic evolution when stochastic extinctions in small viral populations are ignored [23]. In contrast, stochastic extinctions in our agent based model weaken selection in small viral populations. Vaccines can also accelerate antigenic evolution locally when antigenically diverged strains can immigrate re-seed seasonal epidemics [162]. Our model simulates a closed global population where immigration is not a source of novel strains and extinct viral populations cannot be re-seeded. Finally, assuming that new strains do not appear by mutation, vaccination targeting a single strain potentially increases incidence when two competing strains co-circulate [191]. In our model, strains emerge dynamically by mutation, so the novel strains are less likely to appear when prevalence is low.

Improved understanding of the fine-scale evolutionary and immunological dynamics might shift predictions. For instance, the rate of vaccine-driven evolution is sensitive to transmission rates and the distribution of mutation sizes. We chose transmission and mutation parameters such that the simulated epidemiological and evolutionary dynamics match those

of H3N2 [11, 181]. However, in this model, increasing the mutation rate, skewing the distribution of mutation sizes toward large mutations, or increasing the transmission rate increases the rate of antigenic evolution and the tendency for viral populations to diversify [11, 181]. Such changes would also increase the probability that viral populations survive to evolve further or diversify especially under small amounts of vaccination (or vaccines with narrow breadth). Our model assumes that an individual's immune responses against multiple infections or vaccinations are independent, but immunity from prior infection or vaccination affects subsequent immune responses [159]. Consistent with this hypothesis, there is evidence that vaccination history [129, 121, 149] and recipient age (potentially a proxy for infection history) [122] affect vaccine efficacy.

Our results suggest that conventional seasonal influenza vaccines, already have the potential to slow antigenic evolution and eradicate seasonal influenza. In theory, universal vaccines that immunize against all strains necessarily slow antigenic evolution by not discriminating between antigenic variants [5]. Increasing seasonal vaccine coverage, especially in populations that contribute substantially to influenza's evolution, would help realize similar evolutionary benefits. However, as vaccination further reduces disease burden, people may require more incentives to get vaccinated [25, 39, 76].

## 3.4 Methods

### *3.4.1 Model overview*

We adapted an individual-based model of influenza's epidemiological and evolutionary dynamics [11] to include vaccination. In each time step of a tau-leaping algorithm, individuals can be born, can die, can become infected after contacting other hosts, can recover from infection, or can be vaccinated. Transmission occurs by mass action, with the force of infection given by

$$\lambda(t) = \beta \frac{I(t)}{N}, \tag{3.1}$$

where $I$ is the number of infected hosts. For computational efficiency, individuals cannot be coinfected.

Antigenic phenotypes are represented as points in 2-dimensional Euclidean space, analogous to antigenic maps produced using pairwise measurements of serum cross-reactivity [160, 13]. One antigenic unit corresponds to a two-fold antiserum dilution in a hemagglutination inhibition (HI) assay. At the beginning of the simulation, a single founding strain is introduced at the endemic equilibrium in the host population. When hosts recover from infection, they acquire lifelong immunity to the infecting strain. Upon contact with an infected host, the probability that the susceptible host becomes infected is proportional to the distance $d_{\mathrm{n}}$ between the infecting strain and the nearest strain in the susceptible host's infection history, with one unit of antigenic distance conferring a 7% absolute increase in risk (Eq. 3.3) [11, 87, 135].

Each infection mutates to a new antigenic phenotype at a rate $\mu$ mutations per day. The mutation's radial direction is drawn from a uniform distribution, and the size (distance) is drawn from a gamma distribution with mean $\delta_{\mathrm{mean}}$ and standard deviation $\delta_{\mathrm{sd}}$.

Vaccination occurs at rate $r$, breadth $b$ (relative to natural immunity), and lag $\theta$ (relative to the timing of strain selection). The vaccine strain is selected on the first day of each year. By default, the vaccine is distributed for 120 days. During the period of vaccine distribution, individuals are randomly vaccinated at a constant daily rate according to the specified annual vaccination rate.

$$r_{\mathrm{day}} = r_{\mathrm{annual}} \times \frac{1 \text{ year}}{365 \text{ days}} \tag{3.2}$$

By default, the breadth of vaccine-induced and natural immunity are equal. Thus, a

host's probability of infection upon contact is given by

$$\text{Risk} = P(\text{infection}|\text{contact}) = \min\{1, cd_{\text{n}}, \frac{cd_{\text{v}}}{b}\} \tag{3.3}$$

where $d_{\text{n}}$ is the distance between the infecting strain and the nearest strain in the host's infection history, and $d_{\text{v}}$ is the distance between the infecting strain and the nearest strain in the host's vaccination history (if the host is vaccinated) and $c = 0.07$ is a constant for converting antigenic distance to a risk of infection [11, 87, 135].

### *3.4.2 Simulation of vaccine-independent evolution*

We created a simulation where vaccination could not affect antigenic evolution, the "static" simulation. We first ran 500 simulations of the model without vaccination. For each simulation, we recorded the circulating strains and their relative abundances at each time step to use as reference viral populations. The evolution of these reference viral populations is unaffected by vaccination since they were obtained from simulations without vaccination.

To run the static simulation where vaccination could not affect antigenic evolution, we first randomly selected one of the reference viral populations. In each time step of the static simulation, the composition of the viral population was replaced with that of the reference viral population at the matched time step, scaled for prevalence. In this way, vaccination could still alter the overall viral abundance, but the rate of antigenic evolution was already previously set by the dynamics of the simulation without vaccination. Thus, vaccination was separated from the evolutionary process.

### *3.4.3 Estimating the private and social benefits of vaccination*

To generate panel data, we ran simulations at four annual vaccination rates $r$ (0%, 1%, 5%, and 10%) and recorded individual hosts' dates of infection and vaccination. We randomly

sampled 0.005% of individuals from the host population at the end of the simulation for analysis. We fit a linear panel model (equation 3.4) to the simulated longitudinal vaccination data from multiple simulations $j$. Observations are at host $i$ level in each time period $\tau$ (see Table 3.2 for hypothetical example). The dependent variable indicator variable $I_{ij\tau} = 1$ if a host is infected in the current season $\tau$, and 0 otherwise. The indicator $V_{ij\tau} = 1$ if a host is vaccinated in the current season. Analogously lags $V_{ij\tau-k}$ measure vaccination in period $\tau - k$. If the annual vaccination rate in the host population is, e.g., 5%, then $r_{5ij} = 1$. The regression is estimated as a linear probability model (with random effects) in order to simplify interpretation of reported coefficients. Standard errors are clustered at the simulation-level to account for correlation in outcomes across hosts in a simulation. The equation estimated is as follows.

$$\begin{aligned} I_{ij\tau} = \beta_0 + \beta_1 V_{ij\tau} + \beta_2 V_{ij\tau-1} + ... + \beta_5 V_{ij\tau-4} + \\ \beta_6 r_{1ij} + \beta_7 r_{5ij} + \beta_8 r_{10ij} + \epsilon_i + u_{j\tau} \end{aligned} \tag{3.4}$$

The fitted coefficients estimate the change in probability of infection given an individual's vaccination status (direct effects) and the host population's vaccination rate (indirect effects). For example, the coefficient $\beta_1$ estimates the absolute change in the probability of becoming infected in the current season for a host who has also been vaccinated in the current season. Likewise, $\beta_2, \beta_3, \beta_4,$ and $\beta_5$ estimate the respective changes in the risk of becoming infected in the current season given vaccination one, two, three, and four seasons ago. Collectively, $\beta_1, \cdots, \beta_5$ represent the direct benefits of vaccination. More formally, $\sum_{k=1}^{5} \beta_k$ is the impulse response to vaccination over 5 years and measures the total direct protective benefit of vaccination over time.

The coefficients $\beta_6, \beta_7$, and $\beta_8$ estimate the change in an individual's risk of infection in the current season when the population vaccination rate is 1%, 5%, or 10%, respectively.

Thus, $\beta_6, \beta_7$, and $\beta_8$ represent the indirect benefits of vaccination under different vaccination policies.

To estimate the private benefit (equivalent to vaccine efficacy), the absolute reduction in risk can be expressed in terms of a relative risk.

$$\text{Private} = \left[1 - \frac{P(I = 1|V = 1)}{P(I = 1|V = 0)}\right] \times 100\% \tag{3.5}$$

To estimate the social benefit (or a social vaccine efficacy) for a specific vaccination rate $R$, we calculate an analogous relative risk:

$$\text{Social} = \left[1 - \frac{P(I = 1|r = R)}{P(I = 1|r = 0)}\right] \times 100\% \tag{3.6}$$

## 3.5 Data and code availability

The source code of the model can be found at `https://github.com/cobeylab/antigen-vaccine`. All data and code used to generate the results in this manuscript are available at `https://github.com/cobeylab/vaccine-manuscript`.

## 3.6 Competing interests

We have no conflicts of interests to declare.

## 3.7 Author contributions

AM and SC conceived the study. FW performed the analysis and wrote the first draft of the paper. All of the authors contributed to and approved the final version.

## 3.8 Acknowledgements

This work was completed in part with resources provided by the University of Chicago Research Computing Center. FW and SC were supported by NIH grant DP2AI117921. FW was also supported by NIH grant T32GM007281. We thank Ed Baskerville for programming guidance and Mercedes Pascual for insightful comments.

## 3.9 Supplementary Information

### *3.9.1 Vaccination and the invasion fitness of mutants*

We use invasion analysis to understand how vaccination affects the invasion fitness of antigenically diverged strains by effectively reducing susceptibility. We develop an expression for the fitness of an invading mutant strain to explain how the antigenic selection gradient with vaccination.

Here, $S$, $I$, and $R$ represent the fraction of susceptible, infected, and recovered individuals. The birth rate $\nu$ and the death rate are equal, so the population size is constant. All individuals are born into the susceptible class. Transmission occurs at rate $\beta$, and recovery occurs at rate $\gamma$. We vaccinate some fraction $p$ of newborns. In practice, this approximates vaccination of young children, who are primarily responsible for influenza transmission. Vaccinated individuals move into the recovered class.

$$\frac{dS}{dt} = \nu(1-p) - \beta SI - \nu S \tag{3.7}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \nu I \tag{3.8}$$

$$\frac{dR}{dt} = \gamma I - \nu R + \nu p \tag{3.9}$$

The endemic equilibrium of $S_{\text{eq}}$, $I_{\text{eq}}$, and $R_{\text{eq}}$ is

$$S_{\text{eq}} = \frac{\gamma + \nu}{\beta} \equiv \frac{1}{R_0} \tag{3.10}$$

$$I_{\text{eq}} = \frac{\nu(R_0(1-p)-1)}{\beta} \tag{3.11}$$

$$R_{\text{eq}} = 1 - \frac{1}{R_0} - \frac{\nu(R_0(1-p)-1)}{\beta} \tag{3.12}$$

where $R_0$, the basic reproductive number, is the number of secondary infections from a single infected individual in a totally susceptible population.

The disease-free equilibrium (when $p > 1 - \frac{1}{R_0}$) is

$$S_{[I=0]} = 1 - p \tag{3.13}$$

$$I_{[I=0]} = 0 \tag{3.14}$$

$$R_{[I=0]} = p \tag{3.15}$$

We introduce a single invading mutant $I' = \frac{1}{N}$. To find the growth rate of the mutant, we develop an expression for the amount of immunity against the mutant strain. The single mutant has an antigenic phenotype $d$ antigenic units away from the resident. The conversion factor between antigenic units and infection risk is notated by $c$. Thus, the susceptibility to the mutant is given by $\min\{cd, 1\}$, and immunity to the mutant is $\max\{1 - cd, 0\}$. For convenience, we assume $cd \leq 1$.

We can decompose $R_{\text{eq}}$ into immunity conferred by recovery natural infection $R_{\text{n}}$ and immunity conferred by vaccination $R_{\text{v}}$:

$$R_{\text{n}} = 1 - \frac{1}{R_0} - \frac{\nu(R_0 - 1)}{\beta} \tag{3.16}$$

$$R_{\text{v}} = \frac{\nu R_0 p}{\beta} \tag{3.17}$$

$$R_{\text{eq}} = R_{\text{n}} + R_{\text{v}} \tag{3.18}$$

The fraction of the population immune to the invading strain from previous infection is denoted $R'$. Assuming that vaccines confer a breadth of immunity relative to natural immunity $b$,

$$R' = (1 - cd)R_{\mathrm{n}} + (1 - \frac{cd}{b})R_{\mathrm{v}} \tag{3.19}$$

Note that when the mutant and resident are identical ($d = 0$), the immunity to the invading strain is identical to the immunity against $R' = R_{\mathrm{eq}}$. Allowing for coinfection, the fraction susceptible to the invading strain is

$$S' = 1 - R' - \frac{1}{N} \tag{3.20}$$

$$= 1 - R' \tag{3.21}$$

for large $N$. When the vaccination rate exceeds $1 - \frac{1}{R_0}$, the resident is eradicated and $S'$ and $R'$ are calculated using the disease-free equilibrium.

The invasion fitness $s$ of the mutant relative to the endemic strain is the difference between the per-capita growth rates. Note that since the resident is in equilibrium, $dI/dt = 0$.

$$s = \frac{1}{I'}\frac{dI'}{dt} - \frac{1}{I}\frac{dI}{dt} = [\beta S' - (\gamma + \nu)] - 0 \tag{3.22}$$

$$= \beta S' - (\gamma + \nu) \tag{3.23}$$

The value of $s$ increases with greater distance between the mutant and resident, but decreases as more hosts become vaccinated (Fig. 3.5A). The expected $s$ can be used to determine the effect of vaccine coverage on the expected invasion fitness of the mutant $\frac{\partial \mathbf{E}(s)}{\partial p}$. $\mathbf{E}(s)$ is a function of the expected distance of a mutant $\mathbf{E}(d)$. In our model, we assume gamma distributed mutation sizes with a mean $\delta_{\mathrm{mean}}$ of 0.3 antigenic units and standard deviation $\delta_{\mathrm{sd}}$ of 0.6 antigenic units (Fig. 3.5C).

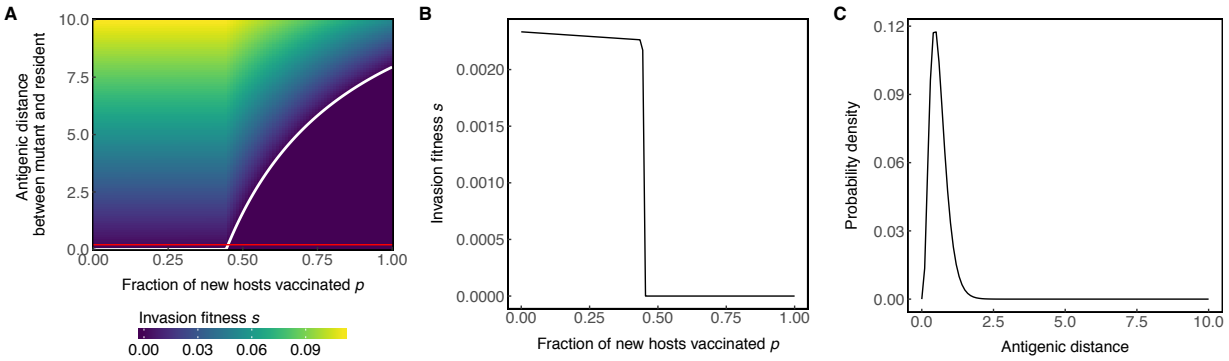


Figure 3.5: (A) High vaccination rates decrease the invasion fitness of mutant strains. For a given vaccination rate, the invasion fitness of a mutant increases with antigenic distance. However, the invasion fitness of a mutant at a given distance decreases as vaccine coverage increases. An example profile of invasion fitnesses is shown for $d = 0.2$ (the red line) in (B). Above the invasion threshold for the resident ($\rho > 1 - 1/R_0$), the mutant must be increasingly more distant to invade. The white curve shows the invasion threshold, where the invasion fitness for the mutant strain is zero. Mutants above the above the curve can invade, while mutants below the curve cannot. (C) Density of gamma distributed mutations with a $\delta_{\text{mean}} = 0.3$ and $\delta_{\text{sd}} = 0.6$.

We decompose $\frac{\partial \mathbf{E}(s)}{\partial p}$ to understand how vaccines affect susceptibility and resistance to change the invasion fitness of the mutant.

$$\frac{\partial \mathbf{E}(s)}{\partial p} = \left(\frac{\partial \mathbf{E}(s)}{\partial S'}\right)\left(\frac{\partial S'}{\partial R'}\right)\left(\frac{\partial R'}{\partial R_{\text{v}}}\right)\left(\frac{\partial R_{\text{v}}}{\partial p}\right) \tag{3.24}$$

$$= (\beta)(-1)(1 - \frac{c\mathbf{E}(d)}{b})(\frac{\nu R_0}{\beta}) \tag{3.25}$$

Since $1 - \frac{c\mathbf{E}(d)}{b} \geq 0$ (i.e. one cannot be more than 100% immune to infection), vaccination must decrease the expected invasion fitness of the mutant $\frac{\partial \mathbf{E}(s)}{\partial p} \leq 0$ , slowing evolution. This decrease is attributed to vaccination reducing susceptibility to the mutant by increasing immunity ($\frac{\partial S'}{\partial R'} \leq 0$ and $\frac{\partial R'}{\partial p} > 0$) against any mutant. Larger breadth of vaccine-induced immunity ($b$) also decreases the expected invasion fitness.

### *3.9.2 Model validation without antigenic evolution*

In the main text, we show general agreement between our simulations and observations of influenza's epidemiology and evolution using our parameterization. We further validate the epidemiological processes of our agent-based model by removing evolution and comparing output against analytic solutions to a model using deterministic ordinary differential equations. A simple analytic solution to a model with antigenic evolution is intractable.

Classical $SIR$ models include vaccination of newborns only. In a newborn-only vaccination model, the threshold eradication rate $p_t = 1 - 1/R_0 \equiv \frac{\gamma+\nu}{\beta}$. Here, we derive an eradication threshold vaccination rate for a model where all hosts are vaccinated at the same rate.

$$\frac{dS}{dt} = \nu - \nu S - \beta SI - pS \tag{3.26}$$
$$\frac{dI}{dt} = \beta SI - \gamma I - \nu I - pI \tag{3.27}$$
$$\frac{dR}{dt} = \gamma I - \nu R - pR \tag{3.28}$$
$$\frac{dV}{dt} = p - \nu V - pV \tag{3.29}$$

At equilibrium:

$$\frac{dI}{dt} = 0 = \beta S^* I^* - \gamma I^* - \nu I^* - pI^* \tag{3.30}$$
$$S^* = \frac{\gamma + \nu + p}{\beta} \equiv \frac{1}{R_0} \tag{3.31}$$

We find agreement between the simulated equilibrium fraction susceptible and the theoretical $S^*$ for a range of influenza-like values of $R_0$ (1.2-3.0) 3.6.

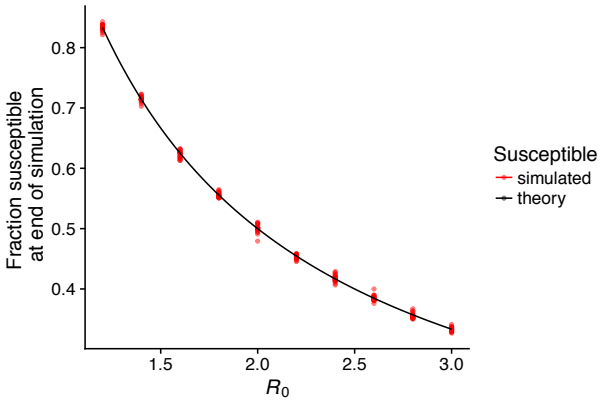


Figure 3.6: Simulated susceptible fraction at the end of 20 years without vaccination. The theoretical equilibrium fraction susceptible is given by $S^* = \frac{1}{R_0}$

We derive a general expression for the eradication threshold first by calculating $I^*$:

$$\frac{dS}{dt} = 0 = \nu - \nu S^* - \beta S^* I^* - pS^* \tag{3.32}$$

$$0 = \nu - S^*(\nu + \beta I^* + p) \tag{3.33}$$

$$\nu \frac{\beta}{\gamma + \nu + p} = \nu + p + \beta I^* \tag{3.34}$$

$$\nu \frac{\beta}{\gamma + \nu + p} - \nu - p = \beta I^* \tag{3.35}$$

$$I^* = \frac{\nu}{\beta}(R_v - 1) - \frac{p}{\beta} \tag{3.36}$$

The condition for the existence of a disease-free equilibrium is $I^* > 0$. We derive an eradication threshold $p_t$ for which $I^* = 0$:

$$I^* = \frac{\nu}{\beta}(R_v - 1) - \frac{p_t}{\beta} = 0 \tag{3.37}$$

$$\frac{\nu}{\beta}(R_v - 1) - \frac{p_t}{\beta} = 0 \tag{3.38}$$

$$\nu(R_v - 1) = p_t \tag{3.39}$$

$$\frac{\nu\beta}{\nu + \gamma + p} - \nu = p \tag{3.40}$$

$$\nu\beta - \nu(\nu + \gamma + p) = p^2 + (\gamma + \nu)p \tag{3.41}$$

$$\nu\beta - \nu(\nu + \gamma) = p^2 + (\gamma + 2\nu)p \tag{3.42}$$

$$0 = p^2 + (\gamma + 2\nu)p - \nu\beta + \nu(\nu + \gamma) \tag{3.43}$$

Since $p \geq 0$, we take the nonnegative root.

$$p = \frac{-(\gamma + 2\nu)}{2} + \frac{\sqrt{(\gamma + 2\nu)^2 - 4(\nu(\nu + \gamma) - \nu\beta)}}{2} \tag{3.44}$$

$$= \frac{-(\gamma + 2\nu)}{2} + \frac{\sqrt{\gamma^2 + 4\nu\gamma + 4\nu^2 - 4\nu^2 - 4\nu\gamma + 4\nu\beta)}}{2} \tag{3.45}$$

$$= \frac{-(\gamma + 2\nu)}{2} + \frac{\sqrt{\gamma^2 + 4\nu\beta}}{2} \tag{3.46}$$

Again, we find agreement between the simulated and theoretical eradication threshold vaccination rates over a range of influenza-like values of $R_0$ (Figs. 3.7, 3.8). Because we initialize the simulations at the endemic equilibrium *without* vaccination, some damped oscillation is to be expected, which may cause eradication at slightly lower vaccination rates than expected by theory (Fig. 3.9). For instance, at $R_0 = 1.8$, theory predicts eradication at $p = 0.0267$ day$^{-1}$, while simulation achieves extinction in 20/20 simulations within 20 years at $p = 0.024$ (Fig. 3.9).

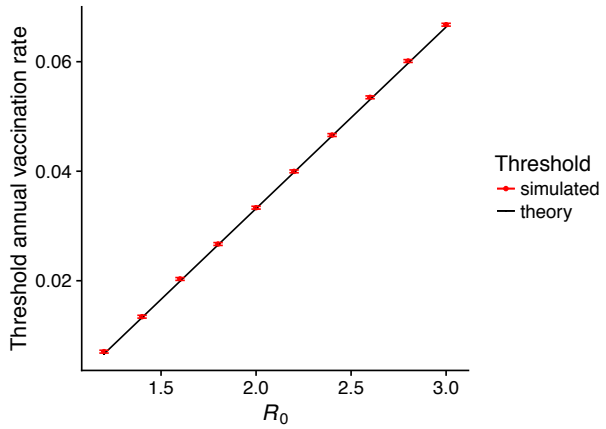


Figure 3.7: With vaccination, the simulated eradication thresholds agree with analytic predictions. The simulated threshold is the minimum vaccination rate where 40/40 simulations go extinct within 20 years. Error bars show the sampling resolution (Fig. 3.8). Simulations were initialized at the analytically derived equilibrium S, I, and R with vaccination (equation 3.46).

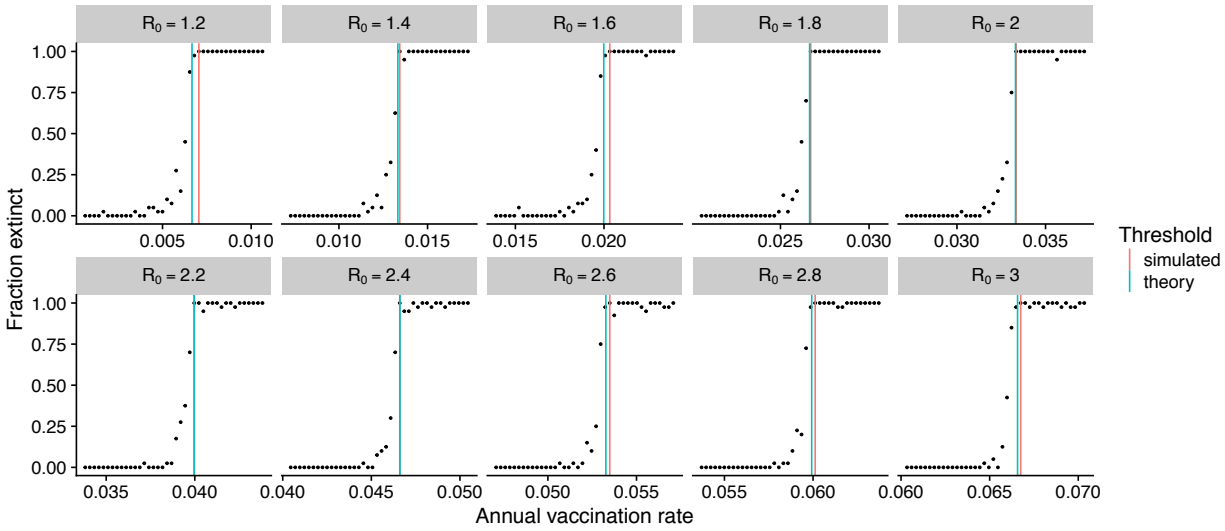


Figure 3.8: Estimation of simulated eradication thresholds without evolution, starting at the equilibrium $S$, $I$, and $R$ with vaccination. To generate response curves, we ran 40 replicate simulations for each combination of $R_0$ and vaccination rate and calculated the fraction of extinct simulations. The simulated eradication threshold is the minimum vaccination rate that causes 40/40 simulations to go extinct within 20 years. When the analytic equilibrium $I$ was nonnegative, we initialized the simulation with a single infection.

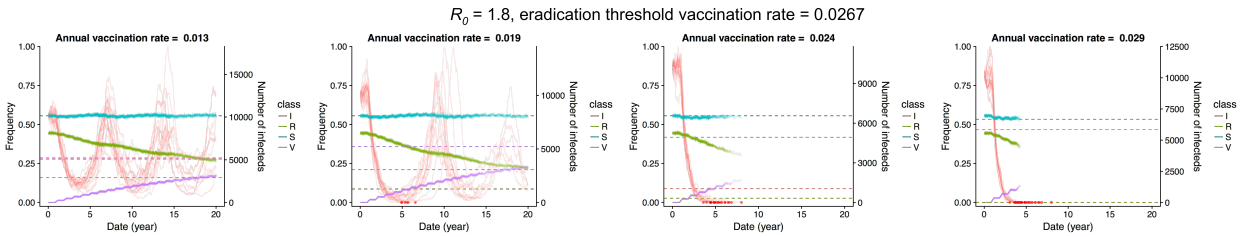


Figure 3.9: Simulated timeseries without evolution, starting at the endemic equilibrium *without* vaccination (i.e., $S_0 = 1/R_0 \equiv \frac{\gamma+\mu+p}{\beta}$, as in the manuscript, but in contrast to Appendix figures 2 and 3). Because the population starts away from the vaccinated equilibrium, the system experiences damped oscillations, which increase the probability of stochastic extinction. Thus, we observe extinction even when the vaccination rate is slightly below the expected eradication threshold. Vaccination remains pulsed in 9-month periods, as in the model. Frequencies of susceptible ($S$), infected ($I$), recovered ($R$), and vaccinated ($V$) individuals are shown for 20 replicate simulations. The left y-axis shows the frequencies of $S$ (blue), $R$ (green), and $V$ (purple). The right y-axis shows the number of infections (red). The dashed lines shows the expected equilibrium frequencies for each class.

## 3.10 Supplementary tables and figures

Table 3.1: Parameters

| Parameter | Value | Reference |
|---|---|---|
| Intrinsic reproductive number ($R_0$) | 1.8 | [94, 19] |
| Duration of infection $1/\gamma$ | 5 days | [30] |
| Population size $N$ | 50 million | (see text) |
| Birth/death (turnover) rate $\nu$ | $1/30$ year$^{-1}$ | [171] |
| Mutation rate $\mu$ | $10^{-4}$ day$^{-1}$ | (see text) |
| Mean mutation step size $\delta_{\text{mean}}$ | 0.6 antigenic units | (see text) |
| SD mutation step size $\delta_{\text{sd}}$ | 0.3 antigenic units | (see text) |
| Infection risk conversion $c$ | 0.07 | [11, 87, 135] |
| Duration of simulation | 20 years | |
| Annual vaccination rate $r$ | 0.0-0.2 year$^{-1}$ | |
| Breadth of vaccine-induced immunity $b$ | 100% | |
| Temporal lag between vaccine strain selection and distribution $\theta$ | 300 days | |

Table 3.2: Sample panel data

| Identifier | | | Data | | | | | | | | Interpretation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$ | $i$ | $j$ | $I_{ij\tau}$ | $V_{ij\tau-1}$ | $V_{ij\tau-2}$ | $V_{ij\tau-3}$ | $V_{ij\tau-4}$ | $r_{1ij}$ | $r_{5ij}$ | $r_{10ij}$ | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | The host was infected this season (1) and only vaccinated 2 seasons ago. The population vaccination rate is 5% |
| 1 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | Host not infected this season (1). Host vaccinated this season and 4 seasons ago. Population vaccination rate is 5% |
| ... | | | | | | | | | | | |
| 10 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | Host infected this season (10). Host vaccinated 4 seasons ago. Population vaccination rate is 10% |

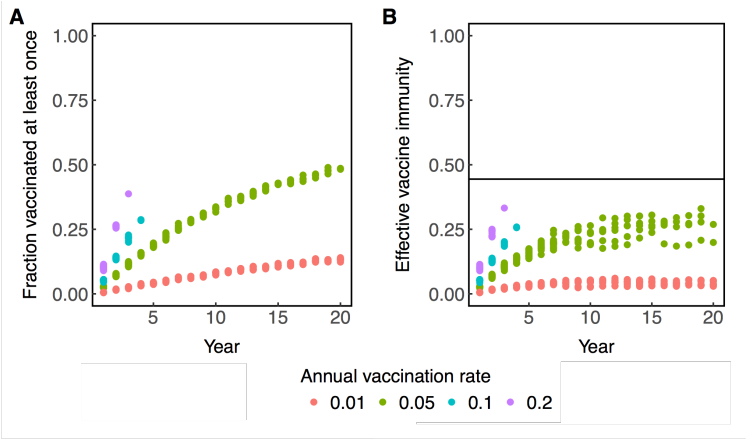


Figure 3.10: (A) Vaccine coverage and (B) effective vaccine-induced immunity over time calculated from simulations. (A) The fraction of individuals who have been vaccinated at least once accumulated over time and saturates at 50%. (B) The effective amount of vaccine-induced immunity in the population is calculated using the mean antigenic distance between circulating strains and the vaccinated hosts' vaccine strains. At any given time, the effective vaccine immunity is $\frac{1}{N}\sum_i^{Np} \min\{0, 1 - cd_{xv_i}\}$, where $N$ is the host population size, $p$ is the fraction of vaccinated, $v_i$ is the vaccine strain received by individual $i$, $x$ is the average circulating strain, $d$ is the antigenic distance between the strains, and $c$ is a constant that converts between antigenic distance and risk. The horizontal line indicates the theoretical eradication threshold in an antigenically homogenous population $1 - 1/R_0$.

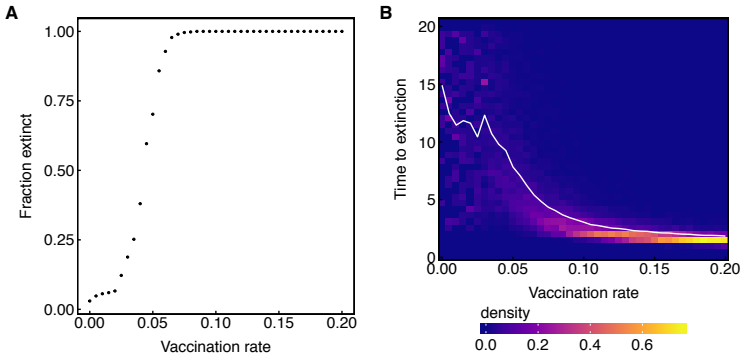


Figure 3.11: High vaccination rates increase the probability of extinction and shorten the average time to extinction. (A) Points show the fraction of simulations where the viral population went extinct before 20 years. (B) Density of times to extinction. The solid white line shows the average time to extinction across these simulations. Lines reflect 500 total simulations for each vaccination rate with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300 - 400$ simulations.

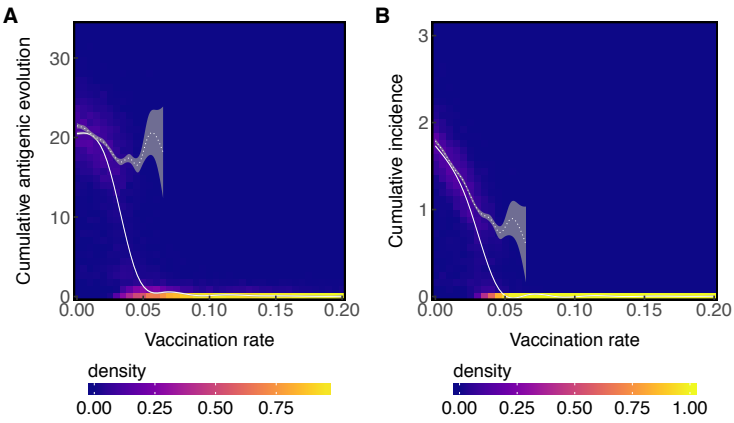


Figure 3.12: With no temporal lag between vaccine strain selection and distribution, increasing the vaccination rate quickly decreases the average amount of (A) cumulative antigenic evolution (A) and (B) incidence. The solid white line shows a LOESS fit to cumulative antigenic evolution and incidence across all simulations. The dotted white line shows a LOESS fit to cumulative antigenic evolution and incidence for simulations where the viral population did not go extinct. Shaded areas show 95% confidence intervals.

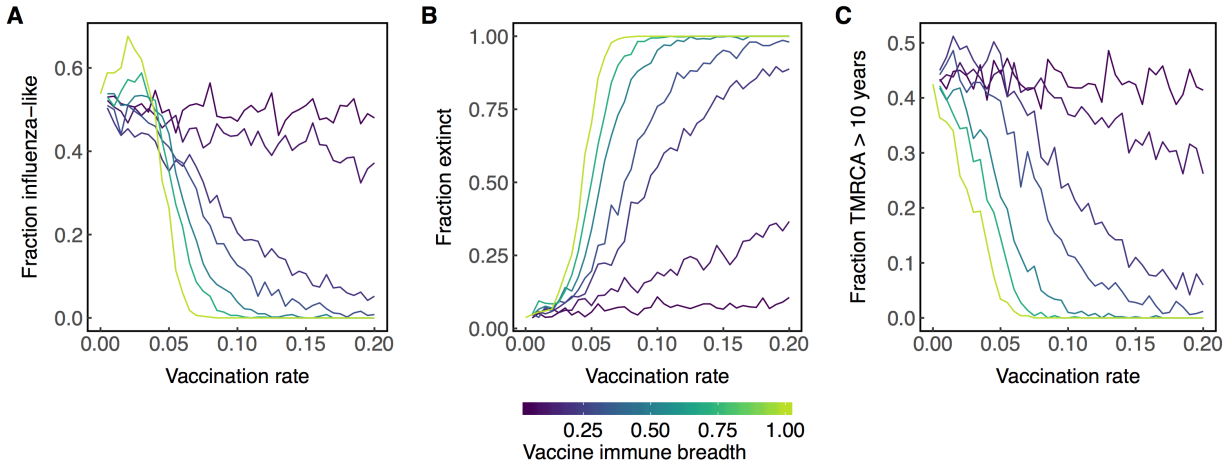


Figure 3.13: Increasing the vaccination rate increases the probability that the viral population will go extinct (B) and decreases the probability of exhibiting influenza-like dynamics (A) or excessive diversification (TMRCA $>$ 10 years) (C). Lines are colored according to the breadth of the vaccine. Data are collected from 500 replicate simulations per unique combination of vaccination rate and vaccine immune breadth with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300-400$ simulations per parameter combination.

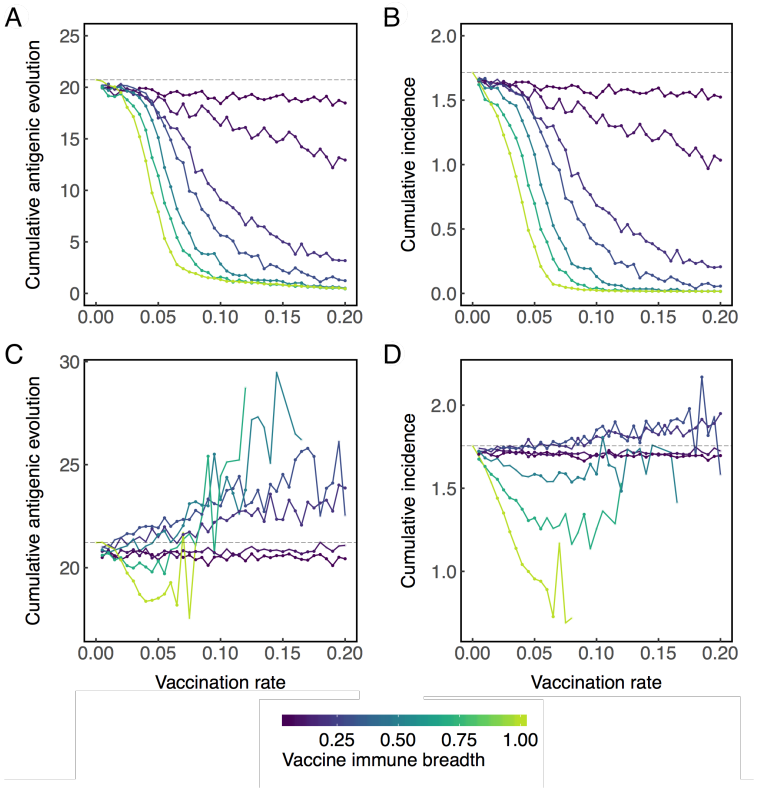


Figure 3.14: Across all simulations (A&B), vaccination decreases the average (A) cumulative antigenic evolution and (B) incidence regardless of breadth. In the subset of simulations where the viral population does not go extinct (C&D), vaccines with narrow breadth are associated with greater average antigenic evolution (C) and incidence (D), but these increases are not necessarily caused by vaccination (see Fig. 3.17. Lines are colored according to the breadth of vaccine-induced immunity. The grey dashed lines indicate the average amount of antigenic evolution (A,C) or incidence (B,D) without vaccination. Points indicate significant decrease (below the dashed line) or increase (above the dashed line) compared to no vaccination according to a Wilcoxon rank-sum test ($p < 0.05$) performed on at least 5 replicate simulations. Complete data are shown in figures/chapter3 3.15 and 3.18 Data are collected from 500 replicate simulations per unique combination of vaccination rate and vaccine immune breadth with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300 - 400$ simulations per parameter combination.

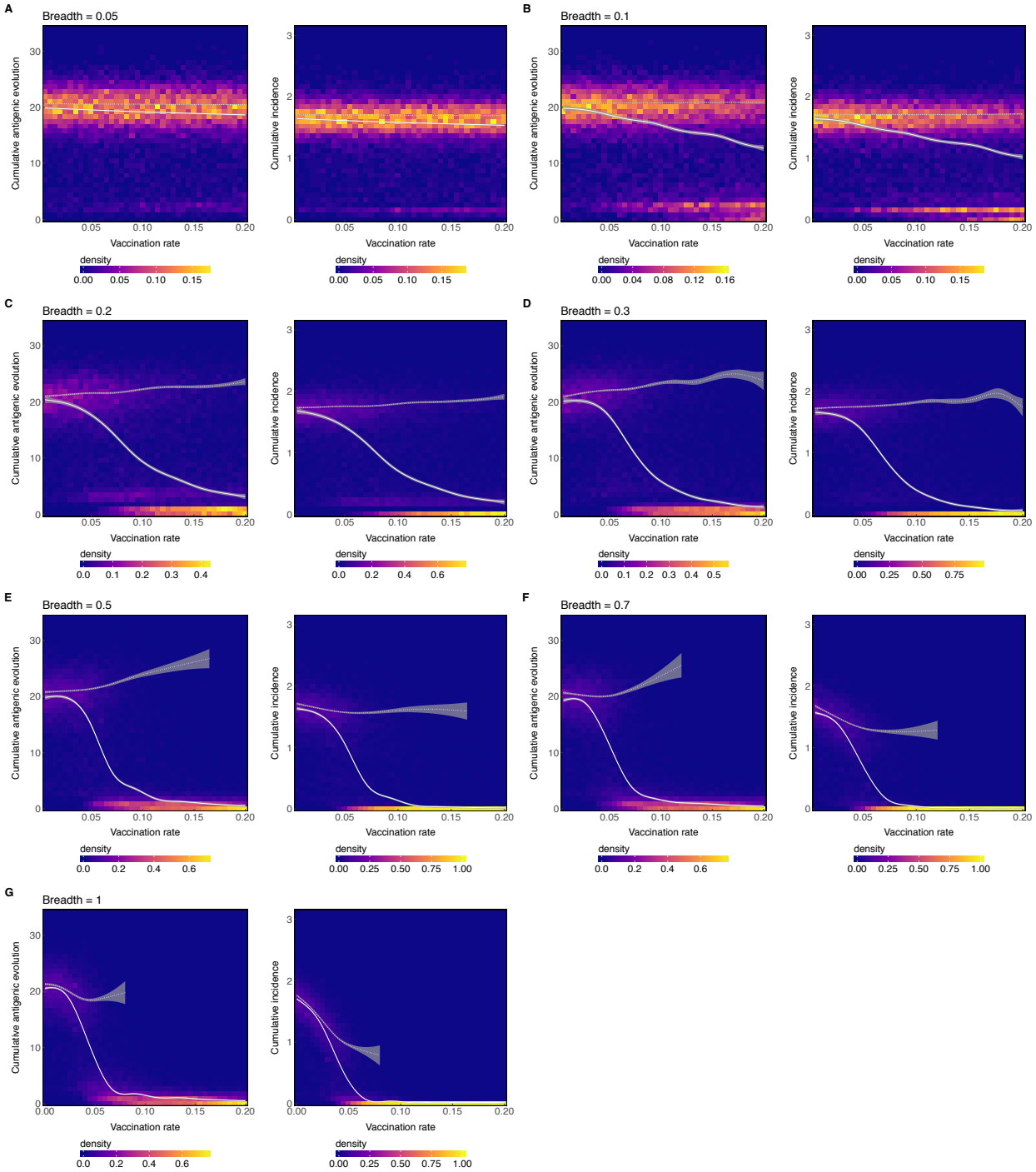


Figure 3.15: Density plots of complete simulation data corresponding to Figure 3.14. The solid white lines show a LOESS fit to cumulative antigenic evolution or incidence across all simulations. The dotted white lines show a LOESS fit to cumulative antigenic evolution or incidence for simulations where the viral population did not go extinct. Shaded areas show 95% confidence intervals Data are collected from 500 replicate simulations per unique combination of vaccination rate and vaccine immune breadth with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300-400$ simulations per parameter combination.

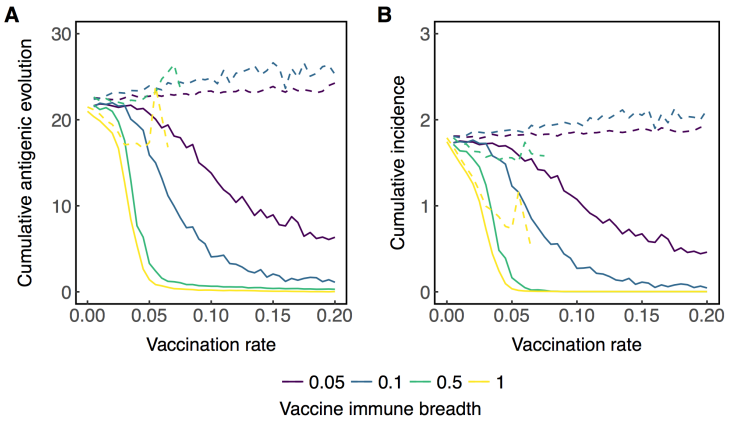


Figure 3.16: With no temporal lag between vaccine strain selection and distribution, lower vaccination rates are needed to achieve the same reductions in (A) cumulative antigenic evolution and (B) cumulative incidence compared to when vaccines are distributed 300 days after strain selection (Fig. 3.14). The solid lines show averages across all simulations, while dotted lines show averages over simulations where the viral population did not go extinct. Lines are colored according to the breadth of vaccine-induced immunity. Data are collected from 500 replicate simulations per unique combination of vaccination rate and vaccine immune breadth with excessively diverse simulations (TMRCA $> 10$ years) excluded, leaving $\sim 300 - 400$ simulations per parameter combination.

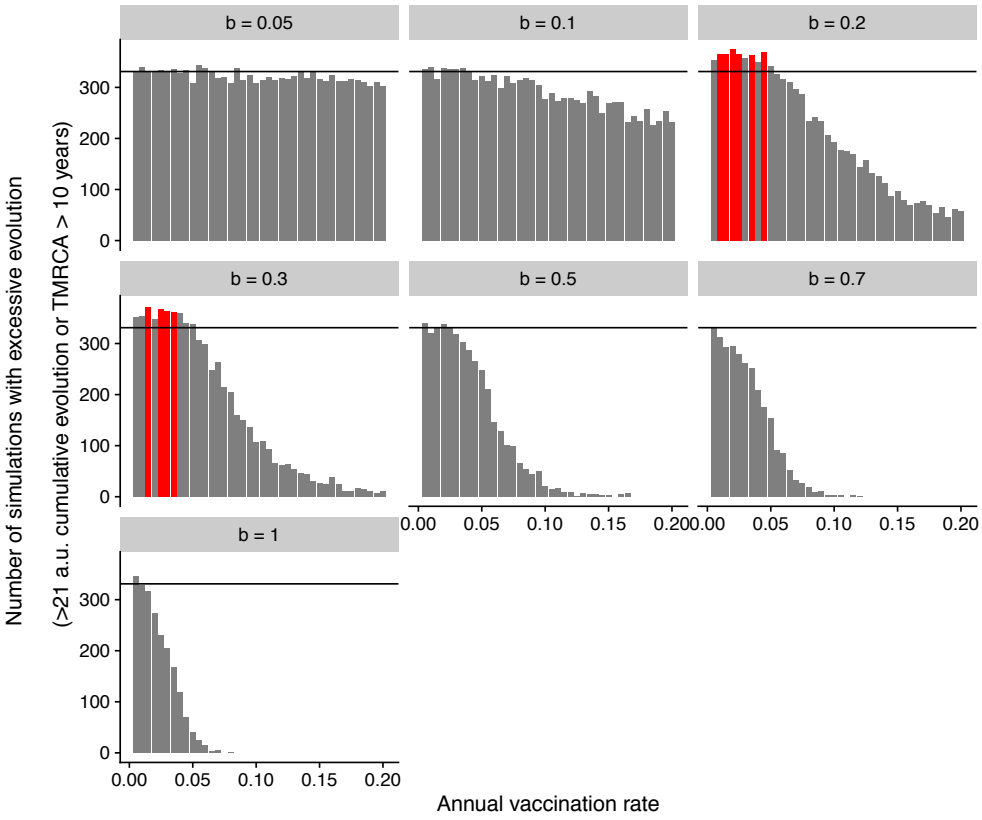


Figure 3.17: Vaccination almost always reduces the rate of antigenic evolution. The subplots show the number of simulations (out of 1000 replicates for each unique combination of parameters) that demonstrate excessive evolution for each vaccination rate and breadth $b$. Here, excessive evolution is defined by either more than 21 antigenic units of cumulative evolution or a TMRCA $> 10$ years. Black lines show the number of simulations that evolve excessively without vaccination (the null expectation if vaccines do not drive faster evolution). Red bars show significantly more counts of excessive evolution compared to unvaccinated simulations ($p < 0.05$, Pearson's $\chi^2$ test).

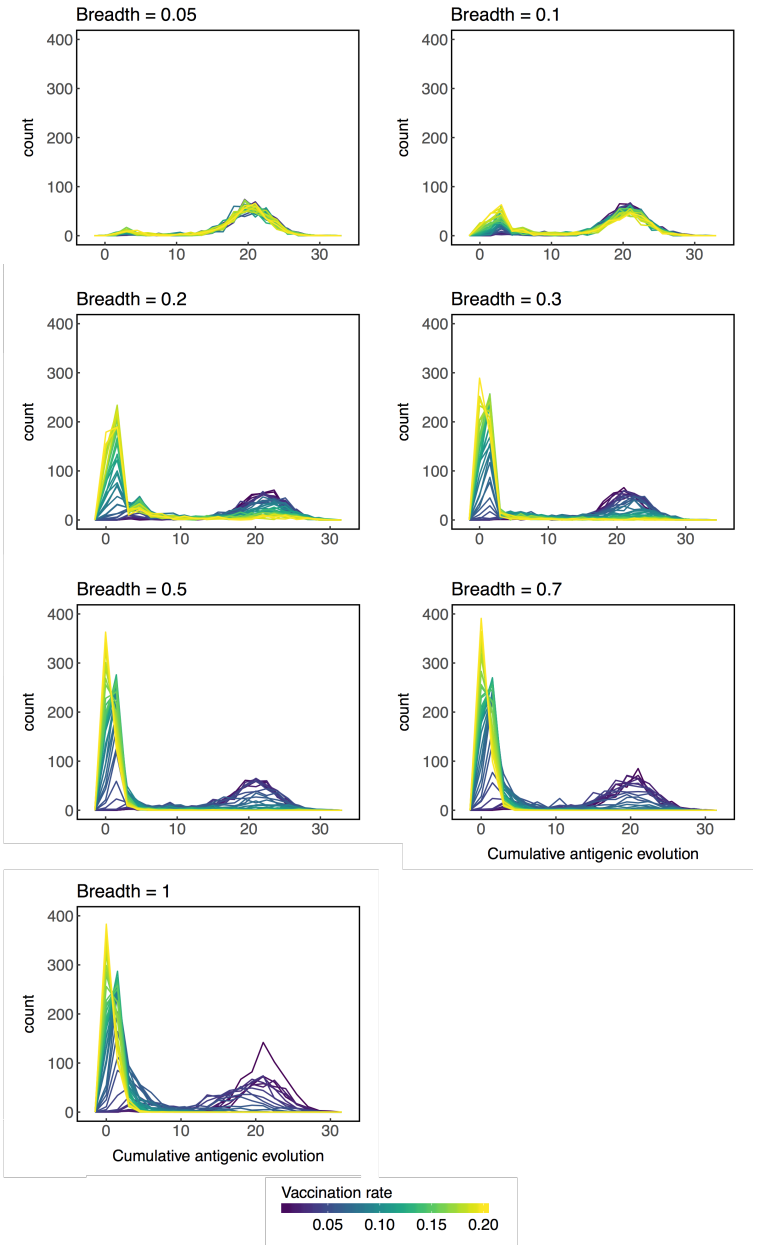


Figure 3.18: The distributions of cumulative antigenic evolution are profiles along each vaccination rate shown in figure 3.15. Data are collected from 500 replicate simulations per unique combination of vaccination rate and vaccine immune breadth with excessively diverse simulations (TMRCA $>$ 10 years) excluded, leaving $\sim 300 - 400$ simulations per parameter combination.

Table 3.3: Private and social benefits of vaccination. In the static model, vaccination cannot affect antigenic evolution. In the dynamic model, vaccination can affect antigenic evolution. Statistics are computed using a linear panel model on longitudinal panel data of simulated hosts' infection and vaccination histories. Robust standard errors shown in brackets are clustered by simulation.

| | Probability of infection in the current season ($\tau$) | |
|---|---|---|
| | Static ($\times 10^{-2}$) | Dynamic ($\times 10^{-2}$) |
| Constant | 9.91*** | 9.94*** |
| | [0.35] | [0.23] |
| Vaccinated in current season ($\tau$) | -4.65*** | -3.34*** |
| | [0.20] | [0.32] |
| Vaccinated 1 season ago ($\tau$-1) | -3.62*** | -2.78*** |
| | [0.18] | [0.33] |
| Vaccinated 2 seasons ago ($\tau$-2) | -2.65*** | -2.05*** |
| | [0.13] | [0.24] |
| Vaccinated 3 seasons ago ($\tau$-3) | -1.47*** | -1.74*** |
| | [0.19] | [0.22] |
| Vaccinated 3 seasons ago ($\tau$-4) | -1.28*** | -1.08*** |
| | [0.20] | [0.16] |
| Vaccination rate = 1% | -0.95** | -0.93** |
| | [0.46] | [0.47] |
| Vaccination rate = 5% | -2.75*** | -5.75*** |
| | [0.50] | [0.65] |
| Vaccination rate = 10% | -2.42*** | |
| | [0.35] | |
| Observations | 1,627,500 | 987,500 |
| Number of hosts | 140,000 | 87,500 |
| Vaccine efficacy (%) | **46.95** | **33.58** |
| Vaccine efficacy (%, social $r = 1\%$) | **9.28** | **9.36** |
| Vaccine efficacy (%, social $r = 5\%$) | **27.4** | **57.8** |
| Vaccine efficacy (%, social $r = 10\%$) | **24.1** | – |

Table 3.4: Private and social benefits of vaccination for a vaccine that provides half the immune breadth of natural immunity ($b = 0.5$). In the static model, vaccination cannot affect antigenic evolution. In the dynamic model, vaccination can affect antigenic evolution. Statistics are computed using a linear panel model on longitudinal panel data of simulated hosts' infection and vaccination histories. Robust standard errors shown in brackets are clustered by simulation.

| | Probability of infection in the current season ($\tau$) | |
|---|---|---|
| | Static ($\times 10^{-2}$) | Dynamic ($\times 10^{-2}$) |
| Constant | 9.63*** | 9.84*** |
| | [0.25 ] | [0.44] |
| Vaccinated this season ($\tau$) | -3.48*** | -3.22*** |
| | [0.19 ] | [0.22] |
| Vaccinated 1 seasons ago ($\tau$-1) | -2.00*** | -1.72*** |
| | [0.16 ] | [0.22] |
| Vaccinated 2 seasons ago ($\tau$-2) | -0.88*** | -0.82*** |
| | [0.14 ] | [0.19] |
| Vaccinated 3 seasons ago ($\tau$-3) | -0.08 | 0.26 |
| | [0.15 ] | [0.19] |
| Vaccinated 4 seasons ago ($\tau$-4) | 0.19 | 0.27 |
| | [0.19 ] | [0.20] |
| Vaccination rate = 1% | 0.68 | -0.20 |
| | [0.44 ] | [0.53] |
| Vaccination rate = 5% | -1.50*** | -0.34 |
| | [0.41 ] | [0.50] |
| Vaccination rate = 10% | -0.91 | -4.85*** |
| | [0.88] | [1.11] |
| Observations | 1,727,500 | 927,500 |
| Number of hosts | 155,000 | 82,500 |
| Vaccine efficacy (%, private) | **36.10** | **32.68** |
| Vaccine efficacy (%, social $r = 1\%$) | **-9.24** | **2.03** |
| Vaccine efficacy (%, social $r = 5\%$) | **1.34** | **3.05** |
| Vaccine efficacy (%, social $r = 10\%$) | **7.27** | **49.3** |

# CHAPTER 4

# DETECTING VACCINE-DRIVEN STRAIN REPLACEMENT OF SEASONAL INFLUENZA

## 4.1 Introduction

Vaccination against seasonal influenza is intended to reduce the incidence of disease. Vaccines that protect at least a little against all circulating influenza viruses should reduce prevalence directly by preventing infection in vaccine recipients and indirectly by preventing infection in potential contacts. In randomized controlled trials (RCTs), the trivalent inactivated vaccine directly reduced the risk of clinical infection by 22% (95% CI: 11%-41%) in healthy children [98] and 41% (95% CI: 36%-47%) in healthy adults [58]. In households and communities, vaccinating children indirectly reduced the risk of influenza infection in unvaccinated individuals by 5%-82% [92, 139, 116, 136]. Since annual vaccination coverage in the United States is nearly 76.3% in children aged 6-23 months and 43.3% in adults [33], the effective vaccination coverage (after taking efficacy against clinical infections into account) may be approximately 17% for both age groups. If we assume the vaccine is equally effective against all strains and that protection against clinical infection also protects against transmission, then current vaccination could be expected to reduce prevalence by 38% (equation 4.9). An obvious place to look for an effect of the seasonal vaccine is thus in prevalence, but the prevalence of influenza is not precisely estimated anywhere [98, 58]. Because the effectiveness of the influenza vaccine does appear to differ between types, subtypes, and clades of influenza, the indirect epidemiological effects of vaccines might be more detectable as changes in the relative abundances of influenza "strains."

Differences in vaccine effectiveness (VE) against circulating strains could lead to selection. In theory, vaccines that reduce the transmission of some strains more than others should increase the prevalence of the non-targeted strains relative to the targeted strains [118, 120,

115, 114]. Such vaccine-driven selection has been observed in several pathogens [138, 66, 114, 180, 28, 1, 78], including H5N2 in chickens [105], but it has not yet been reported for seasonal influenza in humans. Studies suggest that seasonal influenza vaccines prevent clinical infection against some strains more than others. RCTs in adults from 2005-2006 to 2008-2009 suggested lower average efficacy over time against H3N2 compared to H1N1, but similar efficacy against H3N2 and B (Tables 4.4, 4.5) [8, 17, 119, 125, 168]. More recent estimates based on test-negative design (TND) from 2009-2010 to 2016-2017 show lower VE against H3N2 compared to both H1N1 and B on average over time (Fig. 4.6, 4.7) [67, 163, 107, 150, 151, 158, 156, 157, 152, 153, 97, 174, 96, 192, 122, 129, 14, 167, 75, 70]. For example, TND studies in Canada report average VE of 33.5% (95% CI: 21.2%-44.0%) against H3N2, compared to 73.0% (95% CI: 61.9%-80.4%) against H1N1 and 57.6% (95% CI: 49.5%-64.3%) against B (equations 4.13 - 4.18) [150, 151, 158, 156, 157, 152, 153, 97]. In summary, the older RCT studies imply that vaccination should increase the prevalence of H3N2 relative to H1N1, but not necessarily relative to B (Table 4.1). It is unclear whether these conclusions should apply to recent seasons. More recent evidence from studies based on TND suggest that vaccination should increase the prevalence of H3N2 relative to both H1N1 and B (Table 4.1). Vaccines might also distinguish between strains defined on other phylogenetic scales, such as influenza B lineages (Table 4.2) and clades of H3N2 (Tables 4.3) [36, 71, 151].

Higher vaccine coverage should strengthen vaccine-driven selection. Seasonal vaccine coverage has differed consistently between countries over time (Fig. 4.8). For example, in the United States, seasonal vaccine coverage averaged 43.4% and ranged from 32.6% to 46.1% from the 2008-2009 to the 2014-2015 seasons [33]. In contrast, seasonal vaccine coverage in European countries averaged 13.5% (ranging from 10.1% to 18.1% over time) during the same time period [63]. Reported vaccine coverage for any individual European country has not exceeded 30%. Moreover, most European countries do not recommend vaccinating

children, in contrast to the United States (Fig. 4.9) [63, 35]. Thus, we expect signatures of vaccine-driven selection to be more apparent in the United States compared to Europe. In these temperate populations, annual epidemics are seeded from an external source and go extinct at the end of the season [6, 10, 142]. Therefore, vaccine-driven selection most likely occurs within individual seasons and on a local scale.

Here, we define expectations for vaccine-driven selection in seasonal influenza based on immunological and epidemiological evidence and test whether these expectations can be detected in available surveillance data. On average, we expect that compared to less vaccinated populations (e.g., European countries [63]), more vaccinated populations (e.g., the United States [33]) will have a lower prevalence of the strains that are better targeted by the vaccine. Since seasonality and incomplete mixing lead to regional variation in which strains dominate in each season [68], we compare type and subtype frequencies cumulatively over multiple seasons. We examine selection on three phylogenetic scales: among types or subtypes (H3N2, H1N1, and B), influenza B lineages (B/Victoria and B/Yamagata), and H3N2 clades.

## 4.2 Results

### *4.2.1 Spatial differences in influenza subtype and type frequencies are not always consistent with vaccine-driven selection caused by differential vaccine effectiveness.*

We test for vaccine-driven selection among influenza types and subtypes (hereafter referred to generally as subtypes) by comparing subtype frequencies from confirmed influenza cases between the United States and Europe from 2009-2010 to 2016-2017. We examine this range of seasons because other seasons lack the surveillance data required for the analysis. TND studies in Canada [150, 151, 158, 156, 157, 152, 153, 97] and the United States [96, 192, 122, 129, 14, 167, 75, 70] over this time period show significantly lower average effectiveness

against H3N2 compared to either H1N1 or B (Fig. 4.6, 4.7, equations 4.13 - 4.18). VE is also lowest against H3N2 in Europe [174] and Australia [67, 163, 107], although the local differences in VE by type and subtype are not always statistically significant. From 2008-2009 to 2014-2015, seasonal influenza vaccine coverage in European countries averaged 13.5% [63] compared to 43.4% in the United States [33]. Thus, if vaccines select for subtypes against which the vaccine is less effective, we expect the United States to have a greater proportion of H3N2 relative to H1N1 and relative to B in this period.

We computed influenza subtype frequencies using the number of influenza viruses detected by subtype in the WHO FluNet [187] database. The data are contributed by National Influenza Centers (NICs), which collect patients' respiratory samples and test for influenza positivity, type, and subtype. To account for temporal fluctuations in influenza's incidence (which is presently not directly measured by surveillance programs), we calculate a weighted average of seasonal subtype and type frequencies (equation 4.4). Frequencies are weighted using an estimated influenza intensity, which is the product of influenza-like illness (ILI) or acute respiratory illness (ARI) incidence and the fraction of influenza-positive respiratory samples (equation 4.1, Figs. 4.12, 4.13, 4.14) [81].

On average, from the 2009-2010 season to the 2016-2017 season, H3N2 was less abundant than B and more abundant than H1N1 in the United States compared to Europe (Fig. 4.1). Compared to influenza B, we estimate that H3N2 was 1.06 (95% CI: 1.06-1.07) times more abundant in the United States and 1.23 (95% CI: 1.22-1.25) times more abundant in Europe. This difference is in the opposite direction expected based on TND studies from the study period. Compared to influenza H1N1, H3N2 was 1.34 (95% CI: 1.33-1.35) times as abundant in the United States and 0.97 (95% CI: 0.95-0.98) times as abundant in Europe. This difference is consistent with expectations, since vaccines are more effective in general against H1N1 than H3N2 on average during the study period.

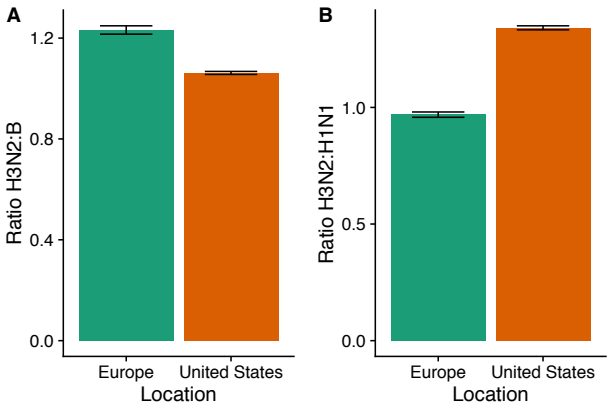


Figure 4.1: Comparing the ratios of (A) H3N2 to B and (B) H3N2 to H1N1 between the United States and Europe from the 2009-2010 season to the 2016-2017 season. Subtype frequencies from the WHO FluNet database are calculated seasonally. Ratios are calculated by first averaging seasonal subtype frequencies weighted according to influenza intensity in the same season (equation 4.4). Error bars show 95% confidence intervals estimated using multinomial distributions of seasonal subtype frequencies. Unweighted seasonal frequencies are shown in Figure 4.16 and seasonal influenza intensities are shown in Figure 4.17.

We also tested for selection at a national level by testing for a correlation between national vaccine coverage and subtype ratios. We expect the ratios of H3N2 to H1N1 to increase monotonically with vaccine coverage, since VE is lowest against H3N2 on average according to TND studies over the study period. We also expect the ratios of H3N2 to B to increase monotonically with vaccine coverage based on VE measured in TNDs (though to a lesser degree than H3N2 to H1N1). We found a significant correlation between average seasonal vaccine coverage and the ratio of H3N2 to H1N1 (Pearson's $r = 0.51, p = 0.03$) but no significant correlation between coverage and the ratio of H3N2 to B (Pearson's $r = 0.24, p = 0.34$) (Fig. 4.2). Results were similar when adjusting vaccine coverage for VE (Fig. 4.10), using Canadian VE [150, 151, 158, 156, 157, 152, 153, 97] for the Northern Hemisphere and Australian VE [67, 163, 107] for the Southern Hemisphere.

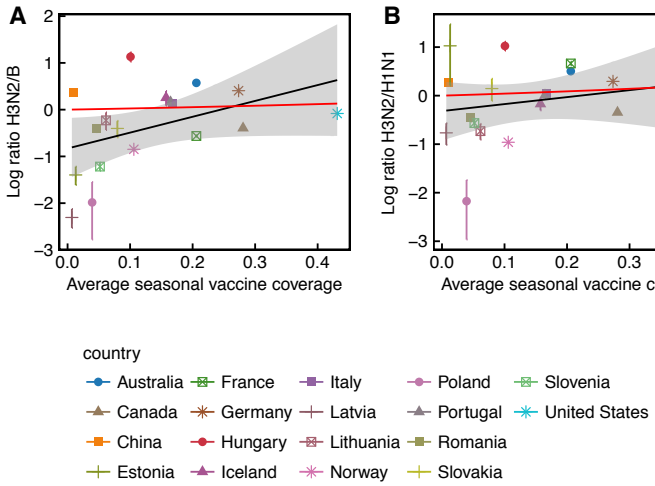


Figure 4.2: Differences in countries' subtype ratios are partially consistent with vaccine-driven selection. (A) The ratio of H3N2 to B among countries does not significantly correlate with the average seasonal vaccine coverage (Pearson's $r = 0.24, p = 0.33$). (B) The ratio of H3N2 to H1N1 among countries significantly correlates with the average seasonal vaccine coverage (Pearson's $r = 0.50, p = 0.03$). Subtype ratios are adjusted for seasonal influenza intensity (equation 4.4). Error bars show 95% confidence intervals estimated using multinomial distributions of seasonal subtype frequencies. Red lines show expectations based on equation 4.27, estimated using VE measured in Canada. Vaccine coverages are reported by national agencies (Fig. 4.8) [33, 63, 61, 27, 79]. Vaccine coverage in China is estimated using doses distributed as reported by the IFPMA IVS task force [133, 134]. The number of seasons contributing to each data point is shown in Figure 4.11.

### *4.2.2 Influenza B lineage frequencies do not differ significantly between more and less-vaccinated populations during seasons where only one lineage was included in the vaccine.*

Multiple lines of evidence offer conflicting expectations for how the trivalent inactivated vaccine should select for influenza B lineages (Tables 4.2). A quadrivalent vaccine containing viruses from both the B/Yamagata and the B/Victoria lineages was introduced in the 2013-2014 season and currently accounts for ∼80% of all influenza vaccinations in the

United States [34]. In clinical trials, the quadrivalent vaccine elicited significantly greater hemagglutination inhibition (HI) titers against both lineages than did the trivalent vaccine against the heterologous lineage [21, 20], suggesting that vaccine-induced immunity is partly lineage-specific. Mouse models and studies in children using the live attenuated vaccine (LAIV) suggest that vaccination with a Victoria strain (B/Brisbane/60/2008-like) induces antibodies responses against Victoria and Yamagata strains by HI, but vaccination with a Yamagata strain (B/Florida/4/2006-like) only elicits antibody responses against Yamagata [155, 154]. Despite these immunological differences measured by HI, the effectiveness of the trivalent vaccine against clinical infection has been comparable against both lineages in the three seasons for which dual estimates exist [129, 122, 96]. Moreover, trivalent vaccines are effective against influenza B even in seasons dominated by a lineage that mismatches the vaccine [97, 158]. Thus, based on TND studies, which measure vaccine-induced protection against clinical influenza infection (albeit with some bias [109]), we expect no difference in the ratios of vaccine-unmatched to matched influenza B lineages between the more vaccinated United States and less vaccinated Europe.

We computed influenza B lineage frequencies using sequence data from the GISAID database (Fig. 4.18) [148]. We use sequences instead of virological data from the FluNet database because B lineage typing was not performed on respiratory samples in most countries until after the quadrivalent vaccine was introduced. We examine data from the 2009-2010 to the 2012-2013 seasons (before the introduction of the quadrivalent vaccine), where there are sufficient sequences to detect a medium-sized difference in B lineage frequencies (Cohen's $h > 0.5$) with 80% power at 0.05 significance. As in the type- and subtype-level analysis, we attempted to minimize the effects of natural spatiotemporal variation in influenza's incidence by weighting each season by an estimated influenza intensity.

We detected no significant difference in the ratios of vaccine-unmatched (non-targeted) to vaccine-matched (targeted) influenza B lineages between the United States and Europe

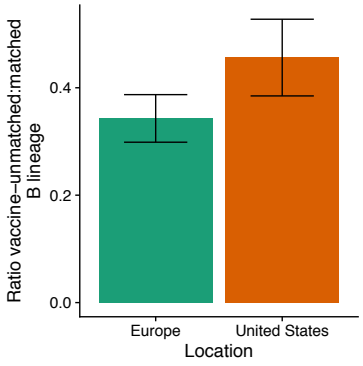


Figure 4.3: The ratios of vaccine-matched to unmatched B lineages do not differ between the United States and Europe on average from the 2009-2010 season to the 2012-2013 season. Lineage frequencies are calculated using sequences from GISAID [148]. Ratios are calculated by first averaging seasonal subtype frequencies weighted by influenza intensity in the same season (equation 4.4). Error bars indicate 95% binomial confidence intervals. Unweighted seasonal lineage frequencies are shown in Figure 4.18.

over this period (Fig. 4.3). We estimated that the vaccine-matched lineage was 0.47 (95% CI: 0.39-0.53) times as abundant as the vaccine-matched lineage in the United States and 0.34 (95% CI: 0.30-0.39) times as abundant in Europe relative to the vaccine-unmatched lineage.

### *4.2.3 In the 2014-2015 season, 3C.2a H3N2 clades are more prevalent in the United States than Europe, but antigenic distances from the vaccine strain are indistinguishable.*

We analyzed H3N2 strain frequencies from the 2014-15 season, where immunological and epidemiological evidence suggests large differences in VE among circulating clades (Table 4.3). During this season, circulating viruses belonging to the 3C.2a clade acquired a new glycosylation site and several other amino acid substitutions in the antigenic site B of HA [36]. The trivalent inactivated vaccine, which contained an A/Texas/50/2012-like H3N2 component, was ineffective against 3C.2a strains (VE: -13%, 95% CI, -51% to 15%), yet moderately effective against 3C.3b strains that were not glycosylated at this site (VE: 52%,

95% CI, -17% to 80%) [151, 71]. Based on these observations, we would expect a greater frequency of 3C.2a viruses in more vaccinated populations. In other seasons, differences in VE by clade are unknown.

Strains of the 3C.2a clade were significantly more frequent in North America compared to Europe ($p < 0.001$, Pearson $\chi^2$ test). Strains of the 3C.3b clade were also significantly less frequent in North America compared to Europe ($p < 0.001$, Pearson $\chi^2$ test) (Fig. 4.4). These patterns are consistent with expectations based on clade-specific VEs. The 3C.3a clade also acquired mutations that may have caused low VE (-48%; 95% CI: -169 to 19) [185, 71], although low incidence of this clade limited the power to measure VE. However, the frequencies of 3C.3a were not significantly different between North America and Europe.

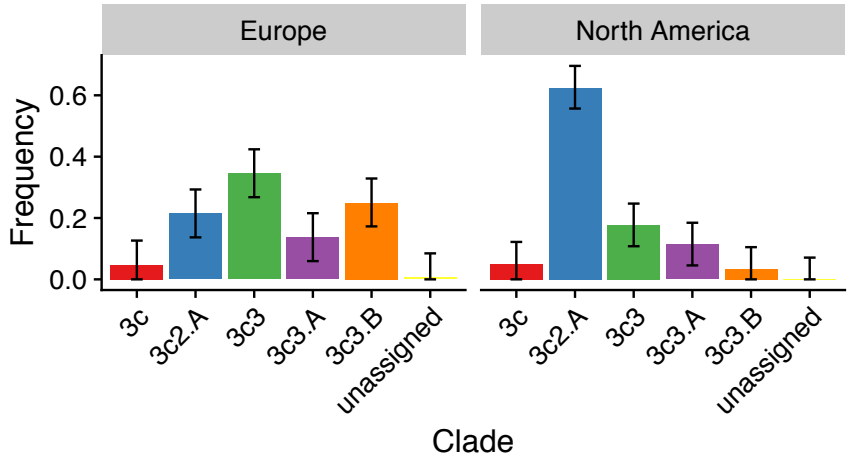


Figure 4.4: Frequencies of H3N2 clades circulating during the 2014-2015 season, stratified by region. Error bars indicate 95% multinomial confidence intervals. Notably, 3C.2a strains are significantly more frequent in North America compared to Europe, and 3C.3b strains are significantly less frequent in North America compared to Europe.

To test for vaccine-induced selection at a finer scale, we estimated the antigenic distances between the vaccine strain and H3N2 strains circulating in the 2014-2015 season. If vaccination selected for mutant H3N2 strains during the 2014-2015 season, then we would expect circulating strains in more vaccinated populations to be more antigenically distant from the vaccine strain compared to less vaccinated populations. Antigenic distances are often mea-

sured by HI assays using naive ferret antisera [127, 88] and can also be quantified by amino acid hamming distances among epitope sites [24, 110, 189]. While estimated antigenic distances are useful for studying general evolutionary patterns, both metrics have unmeasured error that probably varies between seasons and populations. For instance, epitope-based hamming distances could underestimate the immunological effects of glycosylation sites, which easily disrupt antibody binding [36]. For HI distances, antisera raised in naive ferrets can have different specificities compared to antisera from humans, because previous exposures affect the generation of new immune responses [47, 55, 56]. Nonetheless, measures of antigenic distance partly correlate with more direct measures of vaccine effectiveness [24, 110]. Strains from the 2014-2015 season carrying mutations in the antigenic site B of HA reduced the binding of antibodies elicited by vaccination with A/Texas/50/2012 in both ferrets and humans [36], suggesting agreement between ferret HI titers and VE in humans. Here, we test for selection at the level of individual genotypes using these two measures of antigenic distance as a proxy for antigenic phenotype.

We find that the antigenic distances between circulating H3N2 strains and the vaccine strain (A/Texas/50/2012) in North America and Europe are not consistent with vaccine-driven selection in the 2014-2015 seasons (Fig. 4.19). During the 2014-2015 season, H3N2 strains in North America were less antigenically distant from the vaccine strain by epitope hamming distance (9.2 units, 95% CI: 9.0, 9.4) compared to Europe (10.0 units, 95% CI: 9.7 10.3), opposite of expectations. Similarly, according to HI distance, North American H3N2 strains were significantly less distant from the vaccine strain (1.17 units, 95% CI: 1.12-1.21) compared to Europe (1.34 units, 95% CI: 1.26-1.42), also opposite of expectations. Thus, although clade frequencies are consistent with vaccine-driven selection among H3N2 strains in the 2014-2015 seasons, conventional measures of antigenic distances between circulating H3N2 strains are not consistent with expectations.

### *4.2.4 Power analysis*

Is the weak support for vaccine-induced evolution evidence of the vaccine's weak effects or a consequence of insufficient data? We conducted a power analysis using VE from TND studies conducted in Canada during the time period that we analyzed (2009-2010 to 2016-2017) [150, 151, 158, 156, 157, 152, 153, 97]. We first computed the expected difference in subtype and type proportions between two populations (equation 4.27), one vaccinated at 20% and the other at 40% (representing the Europe and the United States, respectively). We assume VEs of 34% against H3N2, 58% against B, and 73% against H1N1. The expected proportion of H3N2 out of H3N2 and B is 51.3% in Europe versus 53.0% in the United States (or H3N2:B ratios of 1.05 and 1.12). The expected proportion of H3N2 out of H3N2 and H1N1 is 52.2% in Europe versus 55.0% in the United States (or H3N2:H1N1 ratios of 1.09 and 1.22). For any given sample of influenza viruses from two populations, one vaccinated at 20% and one at 40% (representing the Europe and the United States, respectively), ∼10,000 samples per population are needed to detect the expected spatial difference in the relative abundance of H3N2 to H1N1, whereas ∼28,000 samples per population are needed to detect the expected difference for H3N2 to B at 0.90 power and 0.05 significance (Pearson's $\chi^2$ test, Fig. 4.21, 4.20). For a difference in VE comparable to those among H3N2 clades in 2014-2015 (about 50% against 3C2.a and 0% 3C3.b), ∼6,000 samples per population would be necessary to detect a difference in frequencies at 0.90 power and 0.05 significance.

The present sample sizes from 2006-2007 to 2016-2017 are more than large enough to detect expected differences in the relative abundances of H3N2 to H1N1 and B between the United States and Europe. However, for the B lineage analysis, the number of sequences available from 2009-2010 to 2012-2013 are insufficient to detect even the maximum expected difference in proportions (i.e., 100% effectiveness against one lineage and 0% against the other, implying vaccine-unmatched lineage prevalences of 0.56 in the less vaccinated population and 0.63 in the more vaccinated population) at 0.90 power and 0.05 significance.

Given the number of available sequences in the United States and Europe, the power to detect the maximum difference in B lineage proportions is ∼0.60 at 0.05 significance. For the H3N2 analysis, the power to detect the expected difference in clade frequencies (assuming 50% effectiveness against one clade and 0% against the other) at 0.05 significance is ∼0.74, although the actual difference in H3N2 clade proportions exceeds what is predicted by our model (equation 4.27). Statistical power may be larger in future seasons, assuming surveillance continues. For example, ∼6000 H3N2 sequences are available in GISAID from the 2016-2017 season, which would have been sufficient to detect the expected difference in H3N2 clade proportions based on VEs from the 2014-2015 season (at 0.90 power and 0.05 significance).

## 4.3 Discussion

We detected partial evidence of vaccine-driven selection on seasonal influenza. At the type and subtype level, TND studies from the 2009-2010 to the 2016-2017 seasons suggest that the vaccine has been less effective against H3N2 than against B or H1N1. Thus, we expect more vaccinated populations to have a greater proportion of H3N2 compared to less vaccinated populations during these seasons. Contrary to expectations, we find that H3N2 is relatively less common than B in the more vaccinated United States compared to Europe during this time period. However, consistent with expectations, we find that H3N2 is significantly more frequent relative to H1N1 in the United States compared to Europe during this period, and there was also a consistent trend of higher H3N2 to H1N1 ratios in more vaccinated countries. When we examined influenza B, we found no significant differences in the ratios of vaccine-matched and unmatched lineages between the United States and Europe, though small sample sizes limit statistical power. It is furthermore unclear if we should expect differences given the apparently high cross-protection after vaccination. Lastly, during the 2014-2015 influenza season, the vaccine was ineffective against the H3N2 3C.2a clade, which

carried several antigenic mutations, but moderately effective against the ancestral 3C.3b clade. We found that strains belonging to the 3C.2a clade were significantly more frequent in North America compared to Europe, suggesting vaccine-driven selection during this season. However, alternative measures of antigenic distance between strains in these regions were not consistent with vaccine-driven selection. Collectively, these results indicate that vaccine-driven selection could be influencing the frequencies of influenza A subtypes, and the distribution of H3N2 clades in one season is also consistent with vaccine-driven selection. However, there is little evidence of vaccine-induced selection on or within influenza B.

The analysis suggests that VE measured in 2009-2010 to 2014-2015 does not explain the relative frequencies of influenza A/H3N2 and B viruses over the same time period. The weak correlation between vaccine coverage and the ratio of H3N2 to B at the country level might be explained by their having similar VEs. Although VEs measured by TND during the period of study (2009-2010 to 2016-2017) are lower to H3N2 than B [67, 163, 107, 150, 151, 158, 156, 157, 97, 174, 96, 192, 122, 129, 14, 167, 75, 70], estimates based on TND studies and RCTs from earlier seasons (2005-2006 to 2008-2009) show comparable effectiveness [8, 17, 119, 125, 168, 153, 152] (Appendix 4.7). Taken together, these studies suggest that VE against H3N2 and B may be similar, although it is unclear to what extent these measurements are generalizable across time.

The higher H3N2 to H1N1 ratio in more vaccinated populations compared to less vaccinated populations suggests a greater difference in VE between H3N2 and H1N1 compared to H3N2 and B. Unlike VE measurements against influenza B, VE measurements against H1N1 (consisting of RCTs from early seasons and TND studies from early and recent seasons) are consistently higher than against H3N2 (Appendix 4.7) [67, 163, 107, 150, 151, 158, 156, 157, 97, 174, 96, 192, 122, 129, 14, 167, 75, 70, 8, 17, 119, 125, 168, 153, 152], suggesting that differences in VE between influenza A subtypes persist through time. Spatial differences in subtype frequencies might only be detectable when differences in subtype-specific VE are

consistently large.

Unmeasured bias in strain frequency data adds uncertainty to our analysis. In general, uncertainty in subtype and type frequencies is small due to large sample sizes. However, strains associated with more severe disease (e.g., H3N2 [104]) may be reported more frequently, since testing for subtype and type draws from symptomatic and medically attended influenza cases. Accordingly, H3N2 may be overrepresented in countries that have larger at-risk demographic groups compared to countries that have smaller at-risk groups. H3N2 may also be overrepresented in countries that use ARI or SARI case definitions to screen for influenza cases (e.g., France before 2014-2015 and Germany [62]), since these cases are more severe than ILI. Thus, compared to what we measured, the ratio of H3N2 to B in Europe may be more similar to that in the United States, and the ratio of H3N2 to H1N1 in Europe may be even lower than that in the United States.

Although our analysis attributes all error to variation in strain frequencies, there is also error *and potentially bias* in VE measurements. Conventional VE measures effectiveness against clinical influenza infection and may fail to capture effectiveness against typical influenza infections due to case ascertainment bias [109]. True VE is thus potentially lower than reported against viruses causing less severe disease, which would make VE between H3N2, B, and H1N1 more comparable. If the VEs are more similar, then we would expect that the type and subtype ratios would be less affected by vaccine coverage, which is partly consistent with what we observe for H3N2 to B. If vaccines are less effective at preventing infections, then they may also be less effective at preventing transmission. Prospective randomized case-control studies that estimate the rate of paucisymptomatic and asymptomatic infections could improve the accuracy of VE measurements.

Future analyses of vaccine-driven selection would benefit from major improvements in two areas influenza surveillance, accurate measurement of VE and standardized surveillance among study populations. RCTs with frequent testing for influenza infection would

help accurately measure effectiveness in preventing transmission. Alternatively, direct comparisons between VE measured by RCTs and TND studies in the same population could inform whether TND-based estimates are reliable in general [108]. VE studies should also include sufficiently large sample sizes to measure age- and type/subtype-specific VE. Standardized surveillance protocols would minimize systematic biases in strain frequencies. Well-documented surveillance protocols (e.g., [99]) and annotated metadata, including patient age and vaccination history, would also help models correct for biases. For these reasons, another place to test for vaccine-driven selection may be between regions of the United States, where surveillance is more consistent and mixing is not rapid enough to homogenize strain compositions [176].

## 4.4 Materials and Methods

### *4.4.1 Data collection*

We calculated type and subtype frequencies using the number of influenza viruses detected by subtype, as reported in the in the WHO FluNet [187] database. We also collected influenza-like illness data from the WHO FluID database [186]. We collected surveillance data from the United States, Australia, Canada, China, and all European countries with surveillance data available in the WHO FluNet [187] and FluID databases [186] (Austria, Belgium, Croatia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom). European frequencies are calculated using a population size-weighted sum of country-level frequencies, using census estimates from the United Nations World Population Prospects (Figs. 4.17, 4.16). We excluded European countries where sampling was clearly biased towards particular age groups (the Netherlands) and countries where an ILI denominator was not reported (Malta and Luxembourg) [62]. For the influenza

B lineage analysis, we calculated lineage frequencies using the number of sequences identified by lineage as reported in GISAID [148]. For the H3N2 analysis, we collected sequences from GISAID and inferred clade membership and frequencies using Nextstrain [88].

### *4.4.2 Estimating influenza intensity*

In calculating cumulative ratios of influenza type, subtype, or lineage incidences (referred to generally as types hereafter), we first calculate seasonal frequencies of each type. We then calculate an average frequency over the observation period by taking a weighted sum of seasonal frequencies weighted by influenza intensity. Influenza intensity is derived from influenza-like illness (ILI) incidence and the fraction of laboratory-confirmed influenza positive respiratory samples.

For a given weekly incidence of influenza-like illness (ILI) (Fig. 4.12) and a weekly fraction of laboratory-confirmed influenza positive respiratory samples (Fig. 4.13), the weekly influenza incidence intensity (Fig. 4.14, 4.17) [81], $F_{\text{week}}$, is

$$F_{\text{week}} = \text{ILI incidence} \times \text{fraction of influenza positive samples.} \tag{4.1}$$

The seasonal influenza intensity $F_{x,t}$ is the average weekly influenza intensity over each season for each country $x$ and each season $t$. We define a season in the traditional way, starting on week 40 of the year and ending on week 39 of the following year.

$$F_{x,t} \equiv \frac{1}{\text{weeks}} \sum_{\text{weeks}} F_{\text{week}}. \tag{4.2}$$

When calculating influenza intensity in Europe, we calculate a sum of European country-level influenza intensities, weighted by population size. The seasonal incidence proxy $I_{x,t,s}$ of type $s$ in season $t$ for country $x$ is given by the fraction of type $s$ during season $t$ (given

by $q_{x,t,s}$, Figs. 4.15, 4.16) multiplied by the seasonal influenza intensity,

$$I_{x,t,s} = q_{x,t,s} F_{x,t}. \tag{4.3}$$

Since epidemics are not always synchronized across populations [176], we calculate a cumulative incidence ratio for types $s_1$ and $s_2$, $I_{s_1,x}/I_{s_2,x}$ as the ratio of the influenza intensity-weighted sums of within-season type frequencies $\sum_t^T q_{x,t,s} \frac{F_{x,t}}{\sum_t F_{x,t}}$ (over all seasons $T$ where surveillance data are available).

$$\frac{I_{s_1,x}}{I_{s_2,x}} = \frac{\sum_t^T q_{x,t,s_1} \frac{F_{x,t}}{\sum_t F_{x,t}}}{\sum_t^T q_{x,t,s_2} \frac{F_{x,t}}{\sum_t F_{x,t}}} \tag{4.4}$$

These equations apply to data from all countries except for China, where influenza-like illness data are not reported. We use the fraction of influenza-positive laboratory samples to calculate the influenza intensity for China. For Germany, we use ARI instead of ILI since ILI is not reported. In France, ARI is reported before the 2014-2015 season and ILI is reported after. We interpolate ILI before the 2014-2015 season by multiplying weekly ARI by the ratio of mean ILI (from 2014-2015 onwards) to mean ARI (from 2009-2010 to 2013-2014).

### *4.4.3 Power analysis*

We approximate sample sizes required to achieve 0.90 power at 0.05 significance using Pearson's $\chi^2$ test by first assuming that respiratory sample sizes from each season are the same (Fig. 4.21, 4.20). We refine these approximations using bootstrapped estimates of power and significance (assuming that pairs of type/subtype abundances are binomially distributed) based on the temporal distributions of sample sizes in the United States. In general, variation in temporal sampling increases the requisite sample size for a given effect size. In the text, we report the sample sizes required accounting for historical temporal variation in sampling.

### *4.4.4 Estimating antigenic distances between H3N2 strains and the vaccine strain*

We inferred the H3N2 phylogenetic tree [88] using a dataset enriched for strains from North America and Europe. We then inferred HI distances to the 2014-2015 vaccine strain (A/Texas/50/2012) for all strains sampled during the 2014-2015 season [88, 127]. Epitope distances were calculated as hamming distances among epitope sites [189].

## 4.5 Acknowledgements

We thank Marcos Vieira for helpful discussion and Rohan Dandavati for assistance with data collection.

## 4.6 Appendix

### *4.6.1 Approximate effects of vaccination on prevalence*

We derive the approximate impact of vaccination on prevalence using an SIR model. $S$, $I$, and $R$ represent the fraction of susceptible, infected, and recovered individuals. The birth rate $\mu$ and the death rate are equal, so the population size is constant. All individuals are born into the susceptible class. Transmission occurs at rate $\beta$, and recovery occurs at rate $\gamma$. We vaccinate some fraction $p$ of newborns. Vaccinated individuals move into the recovered class.

$$\frac{dS}{dt} = \mu(1-p) - \beta SI - \mu S \tag{4.5}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I \tag{4.6}$$

$$\frac{dR}{dt} = \gamma I - \mu R + \mu p \tag{4.7}$$

The endemic equilibrium of $S_{\text{eq}}$, $I_{\text{eq}}$, and $R_{\text{eq}}$ is

$$S_{\text{eq}} = \frac{\gamma + \mu}{\beta} \equiv \frac{1}{R_0} \tag{4.8}$$

$$I_{\text{eq}} = \frac{\mu}{\beta}(R_0(1-p) - 1) \tag{4.9}$$

$$R_{\text{eq}} = 1 - \frac{1}{R_0} - \frac{\mu}{\beta}(R_0(1-p) - 1) \tag{4.10}$$

where $R_0 = \frac{\beta}{\gamma+\nu}$.

We assume the following parameters for influenza: $\beta = 0.36$, $\gamma = 0.2$ (i.e., a 5 day duration of infection), and $\mu = 1/30\text{years}^{-1}$ (implying $R_0 = 1.8$). The prevalence without vaccination is $2.03 \times 10^{-4}$. At 17% vaccine coverage, the prevalence is $1.25 \times 10^{-4}$, equivalent to a 38% reduction in prevalence.

## 4.7 Estimating average vaccine effectiveness by subtype

Vaccine effectiveness ($E$), reported in test-negative design studies is expressed as 1−Odds ratio:

$$E = 1 - \frac{P(\text{infected — vaccinated})/P(\text{not infected — vaccinated})}{P(\text{infected — not vaccinated})/P(\text{not infected — not vaccinated})} \tag{4.11}$$

Vaccine efficacy ($F$), reported in randomized control trials is expressed as $1 -$ Risk ratio:

$$F = 1 - \frac{P(\text{infected — vaccinated})}{P(\text{infected — not vaccinated})} \tag{4.12}$$

In this section, we refer to both metrics as vaccine effectiveness ($E$) for simplicity.

To calculate average vaccine effectiveness, we assume that the log odds ratios $\log(1-E)$ are approximately normally distributed. Then, the mean effectiveness over $T$ seasons $t$ is as

follows.

$$\log(1-\bar{E}) = \frac{1}{T}\sum_{t}^{T}\log(1-E_t) \tag{4.13}$$

$$\bar{E} = 1-\exp[\frac{1}{T}\sum_{t}^{T}\log(1-E_t)] \tag{4.14}$$

Given 95% confidence intervals for seasonal effectivenesses $(E_{t,l}, E_{t,u})$, we calculate a 95% confidence interval for the average effectiveness $(\bar{E}_{t,l}, \bar{E}_{t,u})$ by first calculating the 95% confidence interval for the average log odds ratio $\big(\log(1-\bar{E}_{t,l}), \log(1-\bar{E}_{t,u})\big)$. Again, we assume a normal approximation. For the lower bound,

$$\log(1-\bar{E}_l) = log(1-\bar{E}) + \frac{1}{T}\left[\sum_{t}^{T}(\log(1-E_{t,l}) - log(1-E_t))^2\right]^{\frac{1}{2}} \tag{4.15}$$

$$\bar{E}_l = 1-\exp\left(\log(1-\bar{E}) + \frac{1}{T}\left[\sum_{t}^{T}(\log(1-E_{t,l}) - log(1-E_t))^2\right]^{\frac{1}{2}}\right). \tag{4.16}$$

Similarly, for the upper bound,

$$\log(1-\bar{E}_u) = log(1-\bar{E}) - \frac{1}{T}\left[\sum_{t}^{T}(\log(1-E_{t,u}) - log(1-E_t))^2\right]^{\frac{1}{2}} \tag{4.17}$$

$$\bar{E_u} = 1-\exp\left(\log(1-\bar{E}) - \frac{1}{T}\left[\sum_{t}^{T}(\log(1-E_{t,u}) - log(1-E_t))^2\right]^{\frac{1}{2}}\right). \tag{4.18}$$

In Australia [67, 163, 107], Canada [150, 151, 158, 156, 157, 152, 153, 97], Europe [174], and the United States [96, 192, 122, 129, 14, 167, 75, 70], test-negative design (TND) studies from 2009-2010 to 2016-2017 consistently show lower VE averaged over time against H3N2 compared to H1N1 or B (Fig. 4.6, 4.7). In Canada and the United States, these differences are statistically significant. In Australia, VE is still lower against H3N2 than B, although

not significantly so. In Europe, the differences by subtype are consistent with the general patterns, but are not statistically significant. Fewer seasons VE data in Europe (3-4 seasons on average compared to 4-7 in other locations) reduce the power to detect significant differences in VE by subtype.

VE measurement protocols differ by location. VE measures effectiveness against ARI caused by influenza in the United States, but measures effectiveness against ILI caused by influenza elsewhere. Additionally, TND studies in Europe measure VE in people older than 9 years, while TND studies elsewhere measure VE in people older than 6 months. For our VE-adjusted analysis (Fig. 4.10), we use Canadian VEs for Northern hemisphere countries and Australian VEs for Southern hemisphere countries, since studies from these countries offer the best combination of the number of seasons available and consistency in study protocol.

Randomized control trials (RCTs) are only available from earlier seasons (2005-2006 to 2008-2009) [8, 17, 119, 125, 168] where ILI surveillance data are not available and information about incidence by subtype are less abundant. While we cannot analyze data from these seasons, we summarize the available evidence for differential vaccine effectiveness for completeness. RCTs seldom measure vaccine efficacy against specific subtypes. Thus, to calculate average vaccine efficacies by subtype, we first substitute subtype-specific vaccine efficacy with overall vaccine efficacy (or efficacy against type A where applicable) when one type or subtype clearly dominates the cases in the study population. We calculate the average efficacy for each season, and then calculate overall average efficacies by subtype. On average from 2005-2006 to 2008-2009, RCTs also show lower efficacy against H3N2 (54.7%, 95% CI: 41.8-64.3%) compared to H1N1 (73.6%, 95% CI: 60.3, 77.5%), but instead show comparable (if not slightly higher) efficacy against H3N2 compared to B (53.5%, 95% CI: 27.0, 69.9%). These differences are not statistically significant. RCTs in children are even less common, take place before 2002, and do not measure efficacy against specific types/subtypes [45, 83, 90, 98].

It is unclear whether differences in VE by subtype from the few RCTs from 2005-2006 to 2008-2009 are relevant to our analysis of more recent surveillance data (from 2009-2010 to 2016-2017). Canadian TND studies from a similar time period as RCTs show a similar average trends [153, 152]: lower VE against H3N2 (57%, 95% CI: 42.6-67.7%) compared to H1N1 (80.1%, 95% CI: 64.9%, 88.6%), but higher VE against H3N2 compared to B (45.1%, 95% CI: 21.6-61.5%), though the differences are again not statistically significant. For our analysis of the recent surveillance data, we base our expectations on the more recent TND studies.

### *4.7.1 Derivation of theoretical subtype ratios*

The expected number of infections in an unvaccinated and non-immune population of size N with prevalence $\alpha = \frac{I}{I+S}$ is

$$I = N\alpha \tag{4.19}$$

We include vaccination and specify in terms of vaccine efficacy $E$ (1 - risk ratio) as opposed to vaccine effectiveness (1 - odds ratio). When incidence is low, the two estimates approach each other. Vaccination status is indicated with a subscript $v$ for vaccinated or a subscript $u$ for unvaccinated.

$$E = 1 - \frac{I_v/(I_v + S_v)}{I_u/(I_u + I_u)} \tag{4.20}$$

When some fraction $p$ of the population is vaccinated, the expected number of infections is

$$I = N(1-p)\alpha + Np\left[I_v/(I_v + S_v)\right] \tag{4.21}$$

Since the unvaccinated and vaccinated hosts belong to the same population, we can substitute $\alpha$ in the expression for $E$ (equation 4.20). Here, we also assume that there are no indirect effects of vaccination.

$$E = 1 - \frac{I_v/(I_v + S_v)}{\alpha} \tag{4.22}$$

$$I_v/(I_v + S_v) = \alpha(1 - E) \tag{4.23}$$

Substitute equation 4.23 into equation 4.21.

$$I = N(1-p)\alpha + Np\alpha(1 - E) \tag{4.24}$$

For any two subtypes $s_1$ and $s_2$, the expected ratio of infections is

$$\frac{I_{s_1}}{I_{s_2}} = \frac{N(1-p)\alpha_{s_1} + Np\alpha_{s_1}(1 - E_{s_1})}{N(1-p)\alpha_{s_2} + Np\alpha_{s_2}(1 - E_{s_2})} \tag{4.25}$$

$$= \frac{\alpha_{s_1}}{\alpha_{s_2}} \frac{1 - p + p(1 - E_{s_1})}{1 - p + p(1 - E_{s_2})} \tag{4.26}$$

$$= \frac{\alpha_{s_1}}{\alpha_{s_2}} \frac{1 - pE_{s_1}}{1 - pE_{s_2}} \tag{4.27}$$

Equation 4.27 shows that the expected ratio of subtype incidences is proportional to the ratio of the effective fraction of the population that is unvaccinated against each subtype. In the following analysis, we model the ratios of subtype incidences using equation 4.27. The ratio of subtype incidence is expected to scale linearly with the ratio of the effective unvaccinated fractions. The expected change in the ratios of cumulative subtype incidence with vaccine coverage based on VE measured in Canadian test negative design studies [150, 151, 158, 156, 157, 152, 153, 97] is shown in Figure 4.5.

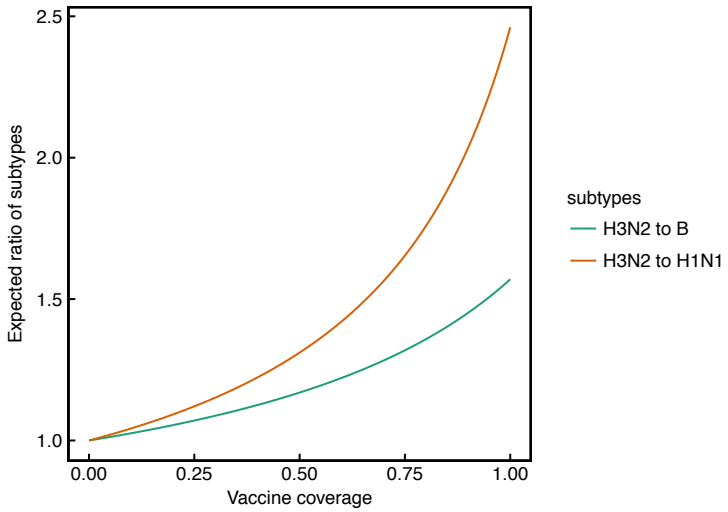


Figure 4.5: Expected change in the ratios of subtypes for increasing vaccine coverage. Lines are colored according to subtypes. Here, we assume that subtypes occur at equal frequencies without vaccination, and that the vaccine effectiveness over multiple seasons is the mean of vaccine effectiveness measured in each season. Vaccine effectiveness estimates are based on studies conducted in Canada [150, 151, 158, 156, 157, 152, 153, 97].

## 4.8 Supplementary tables and figures

Table 4.1: Evidence for potential vaccine-driven selection among influenza types and subtypes.

| Seasons | Observation | References | Interpretation |
|---|---|---|---|
| 2005-2006 to 2008-2009 | Lower vaccine efficacy against H3N2 (54.7%, 95% CI: 41.8-64.3%) compared to H1N1 (73.6%, 95% CI: 60.3, 77.5%), though not a statistically significant difference. Similar efficacy against H3N2 relative to B (53.5%, 95% CI: 27.0, 69.9%) reported in randomized control trials (RCTs) in adults (Table 4.5). | [8, 17, 119, 125, 168] | Possible selection for H3N2 relative to H1N1 but not necessarily relative to B. More recent RCTs have not yet been reported. |
| 2009-2010 to 2016-2017 | Lower vaccine effectiveness against H3N2, compared to either H1N1 or B based on test-negative design studies averaged over time. Differences are significant in Canada [150, 151, 158, 156, 157, 152, 153, 97] and the United States [96, 192, 122, 129, 14, 167, 75, 70], but not in Australia [67, 163, 107] for H3N2 vs. B, and not in Europe [174] for both H3N2 vs. B and H3N2 vs. H1N1 (Fig. 4.6, 4.7). | [67, 163, 107, 150, 151, 158, 156, 157, 152, 153, 97, 174, 96, 192, 122, 129, 14, 167, 75, 70] | Expect selection for H3N2 relative to either H1N1 or B. |

Table 4.2: Evidence for potential vaccine-driven selection among influenza B lineages.

| Observation | References | Interpretation |
|---|---|---|
| Vaccination with a Victoria strain (B/Brisbane/60/2008-like) induces antibodies responses against Victoria and Yamagata strains by HI, but vaccination with a Yamagata strain (B/Florida/4/2006-like) only elicits antibody responses against Yamagata (in mice and in children using the LAIV). | [155, 154] | Possible selection for Victoria during Yamagata vaccine seasons. No vaccine-driven selection for either lineage during Victoria vaccine seasons. |
| The quadrivalent vaccine elicited significantly greater HI titers against both lineages than did the trivalent vaccine against the heterologous lineage. | [21, 20] | Possible selection for the unmatched lineage. |
| TIV effectiveness against clinical influenza infection was comparable against both lineages in the 2011-2012, 2012-2013, and 2015-2016 seasons. Moreover, trivalent vaccines are effective against influenza B even in seasons dominated by a lineage that mismatches the vaccine. | [129, 122, 96, 97, 158] | Vaccine-driven selection for either lineage is not expected. |

Table 4.3: Evidence for potential vaccine-driven selection among H3N2 strains.

| Season | Observation | References | Interpretation |
|---|---|---|---|
| 2016-17 | Egg-adapted vaccine lacks glycosylation site found in circulating 3C.2a viruses. | [194] | The vaccine is poorly immunogenic against circulating strains in general. Little vaccine-driven selection is expected. |
| 2014-15 | Circulating 3C.2a viruses have a new glycosylation site and several other site B residues different from the vaccine strain. Clade-specific VEs differ: 3C.2a -13% (95% CI, -51% to 15%); 3C.3b 52% (95% CI, -17% to 80%) | [36, 151, 71] | 3C.2a strains differ immunologically from the vaccine strain. Vaccine-driven selection for 3C.2a strains is expected. |
| 2012-13 | In adults, the vaccine fails to induce responses to novel mutations on the vaccine strain's HA. | [47] | The vaccine is poorly immunogenic against circulating strains in general. Little vaccine-driven selection is expected. |

Table 4.4: Vaccine efficacy in adults measured in randomized control trials.

| | | Efficacy (CI) | | | | | | | Inclusion in averages[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ref. | Season | Overall | A | H3N2 | H1N1 | B | Type/subtype dominance | Location | H3N2 | H1N1 | B |
| [119] | 2008 | 41.0 (21.1, 55.9) | – | – | – | – | ∼50% B | Australia | (D) | (D) | (D) |
| [119] | 2009 | 42.7 (26.3, 55.4) | – | – | – | – | H1N1 dominated | Australia | (N) | **I** | (N) |
| [130] | 2004-05 | 74 (37, 89) | – | – | – | – | 50% H3N2 50% B | USA | (D) | (D) | (D) |
| [18] | 2005-06 | 22.3 (-49.1, 58.5) | 25.1 (-260.9, 82.2) | – | – | 21.5 (-65.9, 61.6) | Very few cases | Czechia | (F) | (F) | (F) |
| [95] | 2005-06 | 49.4 (12.7, 70.7) | – | – | – | – | Majority H3N2 | USA | **I** | (D) | (D) |
| [130] | 2005-06 | 23 (-153, 73) | – | – | – | – | Majority H3N2 | USA | (F) | (N) | (N) |
| [95] | 2006-07 | 49.2 (-0.04, 75.3) | – | – | – | – | Majority H3N2 | USA | **I** | (D) | (D) |
| [17] | 2006-07 | 61.6 (46, 72.8) | – | – | – | – | >99% H3N2 | Czechia, Finland | **I** | (N) | (N) |
| [74] | 2007-08 | 63.0 (46.7) | – | 49.3 (-9) | 81.5 (60.9) | 53.2 (22.2) | – | USA, Finland, Poland | **I** | **I** | **I** |
| [125] | 2007-08 | – | 72 (49, 84) | – | – | 40 (-189, 86) | 90% H3N2 | USA | **I** | (F) | (F) |
| [168] | 2007-08 | – | 49.0 (24.7, 65.9) | – | – | 37.2 (-8.9, 64.5) | 70% H3N2 among A | USA | **I** | (N) | **I** |
| [8] | 2008-09 | 71.5 (54.7, 82.1) | – | 50.0 (-173, 90.8) | 75.2 (55.4, 86.2) | 60.1 (9.5, 82.4) | – | USA | (F) | **I** | **I** |

[a] **I** Include in average (Table 4.5), (F) Too few cases, (D) Ambiguous type/subtype dominance, (N) Not reported and not estimable

Table 4.5: Estimated seasonal vaccine efficacy in adults.

| | Average efficacy (studies used to compute average)[a] | | |
|---|---|---|---|
| Season | H3N2 | H1N1 | B |
| 2005-2006 | 49.4 (12.7,70.7) [95] | – | – |
| 2006-2007 | 55.9 (35.5,70.4) [95, 17] | – | – |
| 2007-2008 | 58.3 (40.9,66.9) [74, 125, 168] | 81.5 (60.9) [74] | 45.8 (21.2, 59.2) [74, 168] |
| 2008-2009 | – | 62.3 (48.1, 72.6) [119, 8] | 60.1 (9.5, 82.4) [8] |
| Average | 54.7 (41.8, 64.3) | 73.6 (60.3, 77.5) | 53.5 (27.0, 69.9) |

[a]Seasonal averages are calculated using arithmetic means of log RRs (equation 4.13) reported by studies listed in Table 4.4. When subtype or type-specific efficacy is not explicitly measured, the efficacy against the dominant type/subtype in the study population is estimated as equal to the overall efficacy.

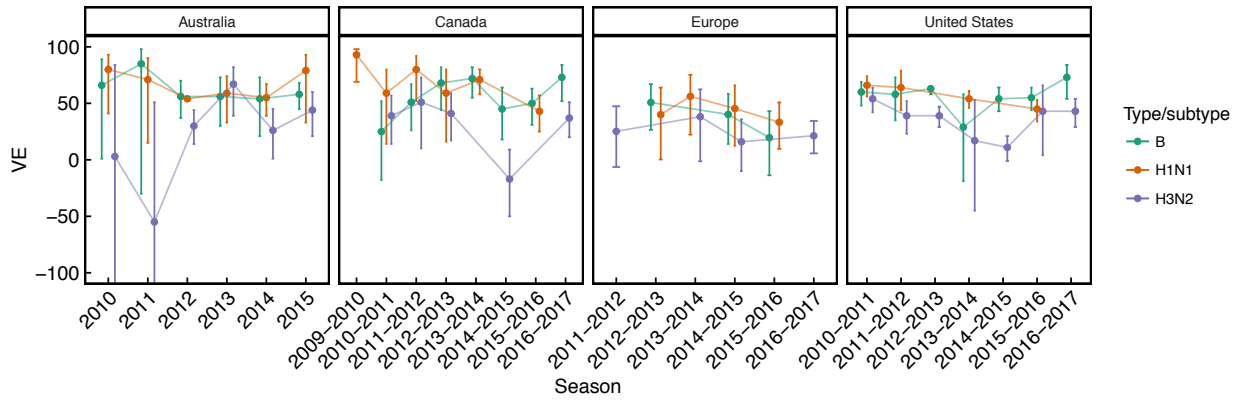


Figure 4.6: Seasonal vaccine effectiveness by type and subtype measured by test-negative design studies in Australia [67, 163, 107], Canada [150, 151, 158, 156, 157, 152, 153, 97], Europe [174], and the United States [96, 192, 122, 129, 14, 167, 75, 70] from the 2009-2010 season to the 2016-2017 season. European VEs include data from study sites in Germany, Spain, France, Croatia, Hungary, Ireland, Italy, The Netherlands, Poland, Portugal, Romania, and Sweden. Error bars show 95% CIs. VE in the United States measures effectiveness against medically attended ARI caused by influenza. Elsewhere, VE measures effectiveness against medically attended ILI caused by influenza. European studies enroll individuals older than 9 years, while other studies enroll individuals older than 6 months.

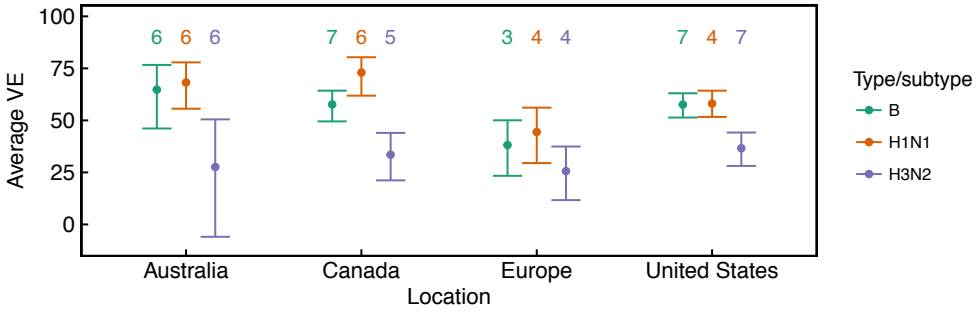


Figure 4.7: Seasonal vaccine effectiveness by type and subtype averaged over time in Australia [67, 163, 107], Canada [150, 151, 158, 156, 157, 152, 153, 97], Europe [174], and the United States [96, 192, 122, 129, 14, 167, 75, 70] from the 2009-2010 season to the 2016-2017 season. European VEs include data from study sites in Germany, Spain, France, Croatia, Hungary, Ireland, Italy, The Netherlands, Poland, Portugal, Romania, and Sweden. Numbers indicate the number of seasons used to calculate each mean. The specific seasons used to compute means are shown in Figure 4.6. Error bars show 95% CIs. Means and 95% CIs are calculated using arithmetic means of log ORs (equations 4.13-4.18)

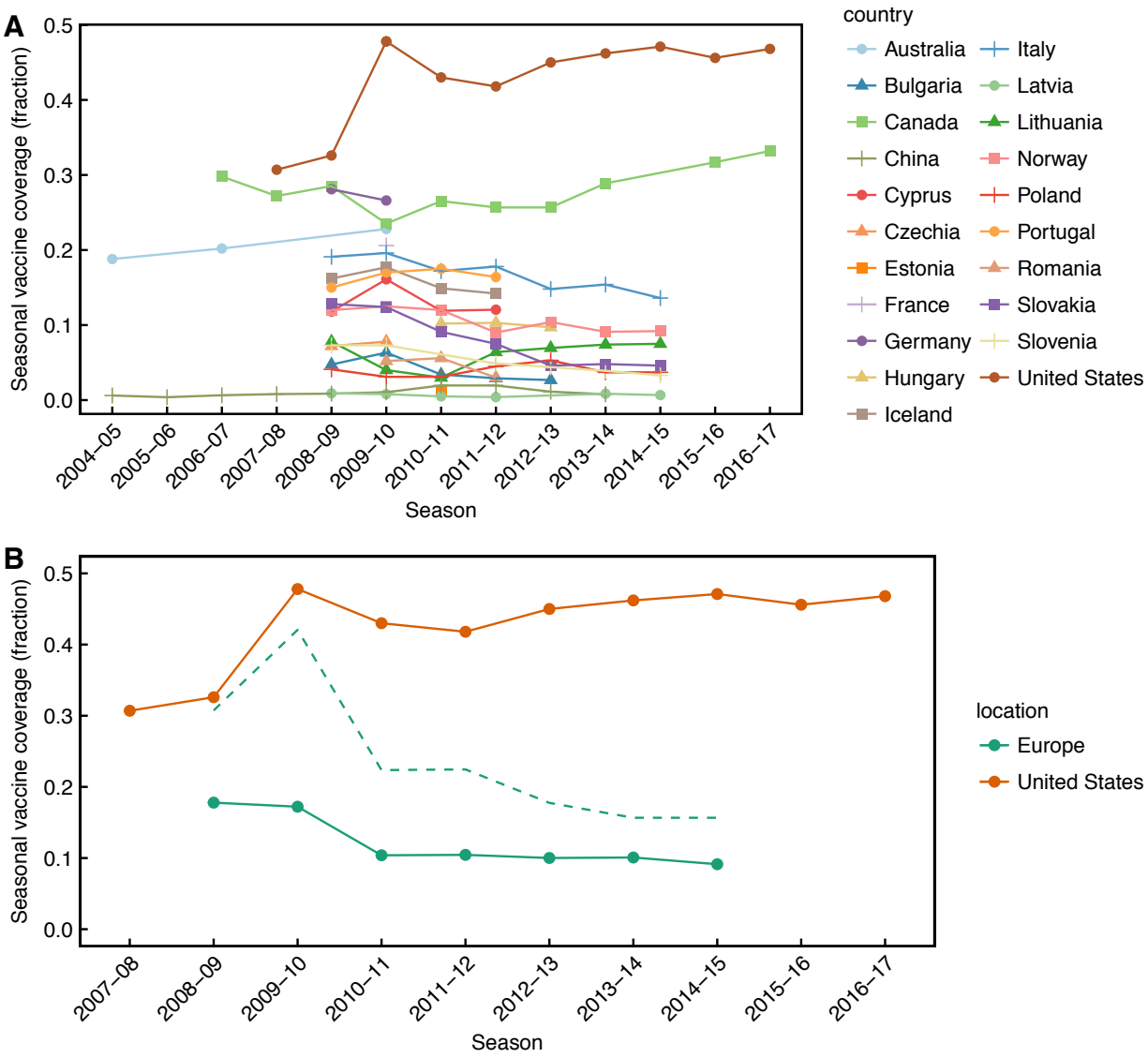


Figure 4.8: Seasonal vaccine coverage reported by national agencies [33, 63, 61, 27, 79], and vaccine coverage in China estimated from doses distributed as reported by the IFPMA IVS task force [133, 134]. (A) Vaccine coverage is reported by country. (B) European vaccine coverage is compared against United States vaccine coverage. The dashed line shows the total population size of countries reporting vaccine coverage divided by the total population of Europe in each season.


Figure 4.9: Seasonal influenza vaccination recommendations by age group in Europe [63] and the United States [35] during the 2007-2008 season and the 2014-2015 season.

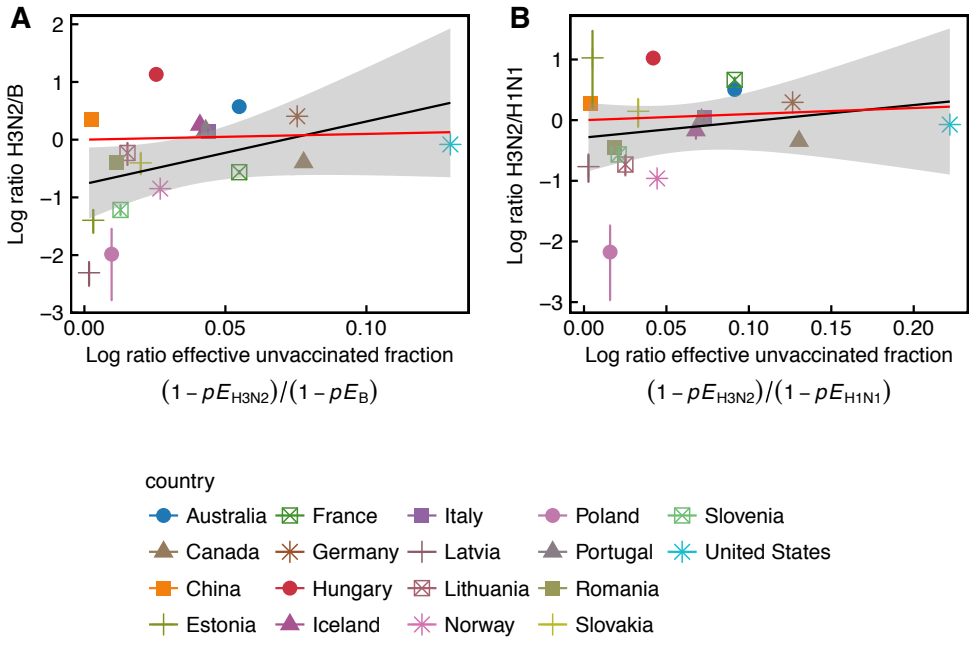


Figure 4.10: Country-level differences in subtype ratios of countries are not consistent with vaccine-driven selection, accounting for vaccine effectiveness. (A) The ratio of H3N2 to B among countries does not significantly correlate with the theoretical ratio (Pearson's $r = 0.21, p = 0.40$) (B) Similarly, the ratio of H3N2 to H1N1 among countries correlates significantly with the theoretical ratio (Pearson's $r = 0.50, p = 0.04$). Theoretical ratios are calculated using Canadian VEs [150, 151, 158, 156, 157, 152, 153, 97] for the Northern Hemisphere and Australian VEs for the Southern Hemisphere [67, 163, 107]. Error bars show 95% confidence intervals estimated using multinomial distributions of seasonal subtype frequencies. Red line shows expectation based on equation 4.27. The number of seasons contributing to each data point is shown in Figure 4.11.

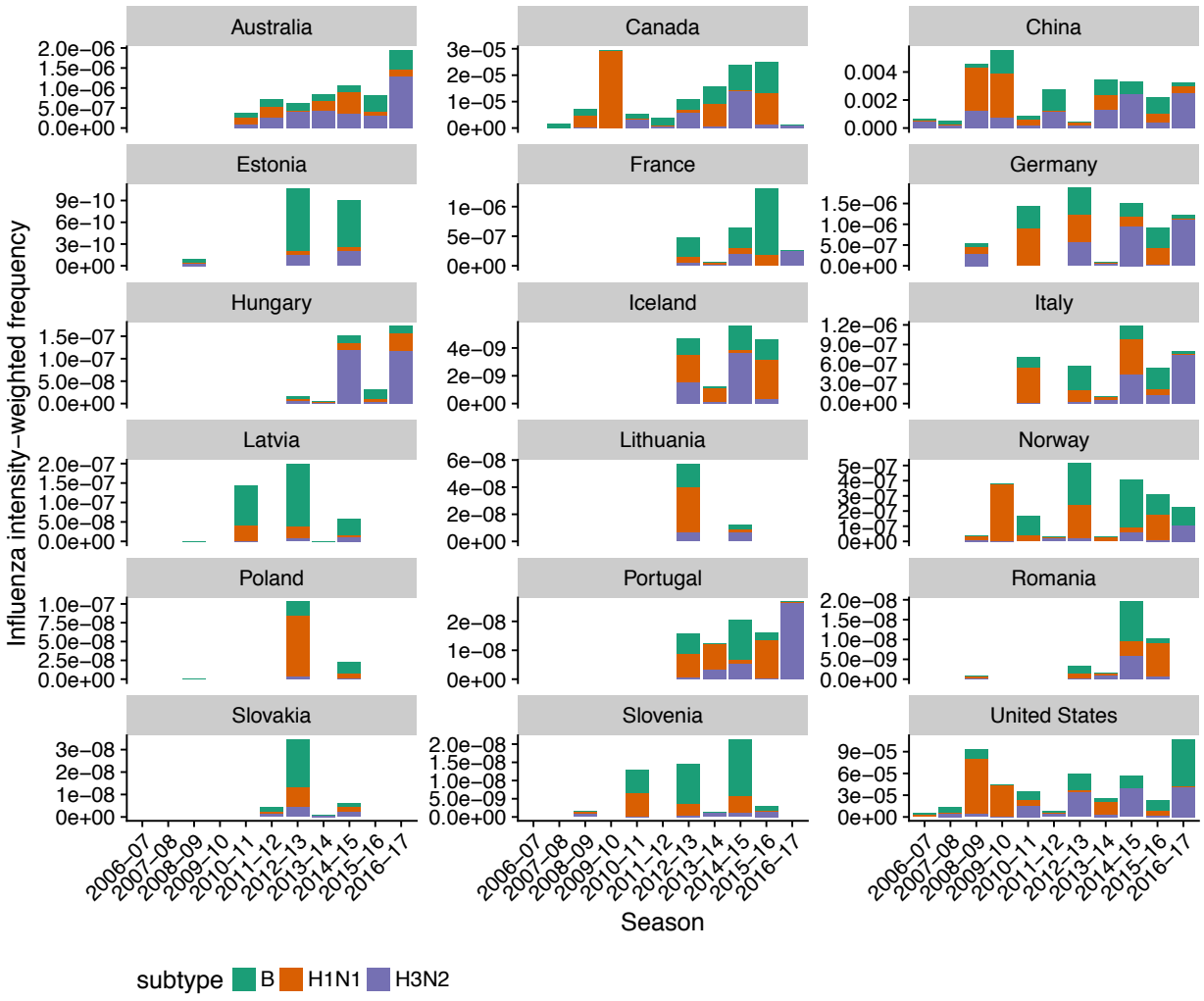


Figure 4.11: Influenza intensity-weighted seasonal subtype and type frequencies prior to summation for use in country-level analysis (Figs. 4.10, 4.2).

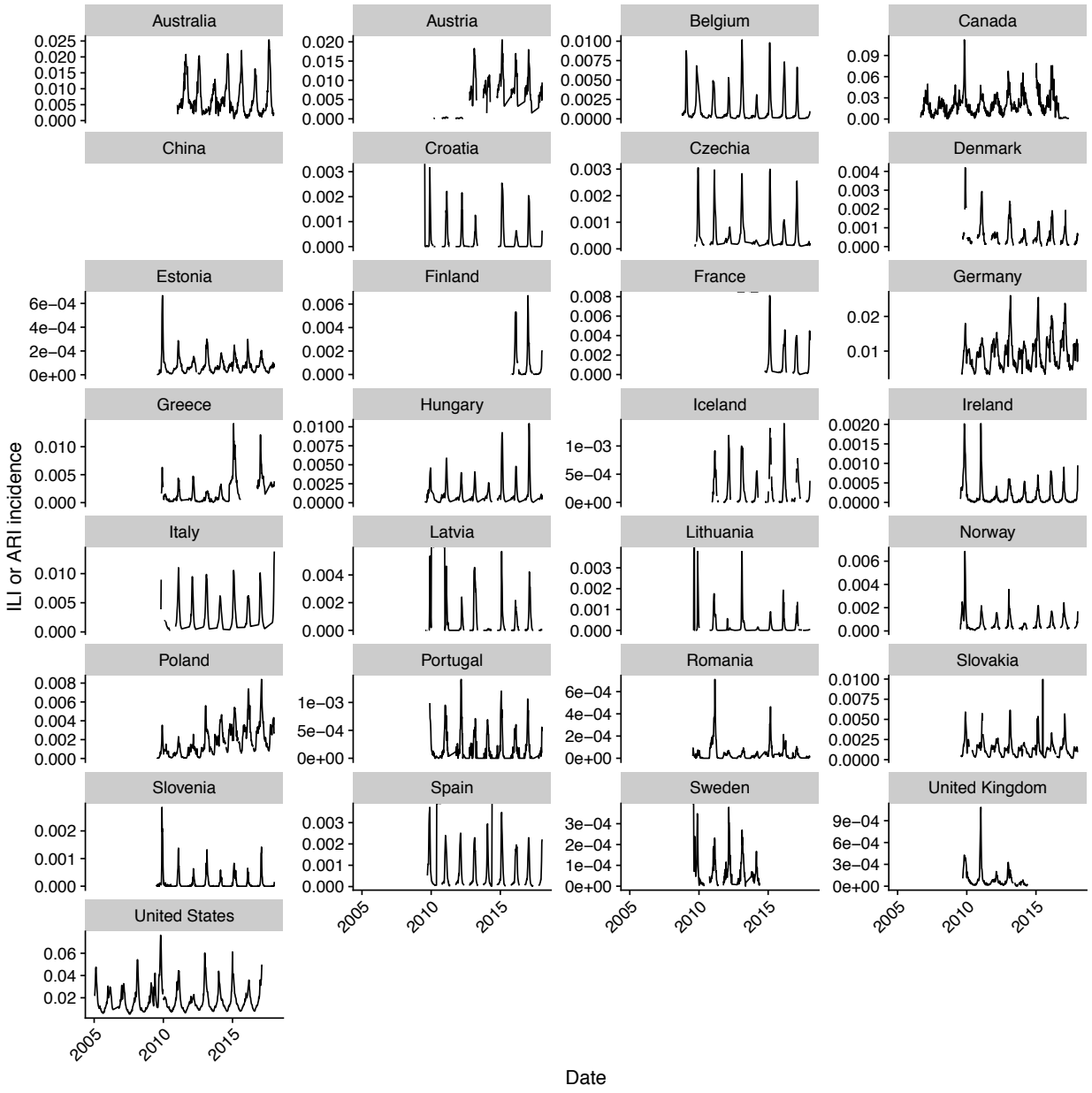


Figure 4.12: Country-level seasonal influenza-like illness (ILI) or acute respiratory illness (ARI) incidence are shown, as reported in the WHO FluID database [186]. ILI incidence as reported is shown for all countries except for France and Germany. Before the 2014-2015 season, ILI incidence in France is estimated from ARI incidence. ARI incidence is shown for Germany. Incidences are reported at weekly resolution.

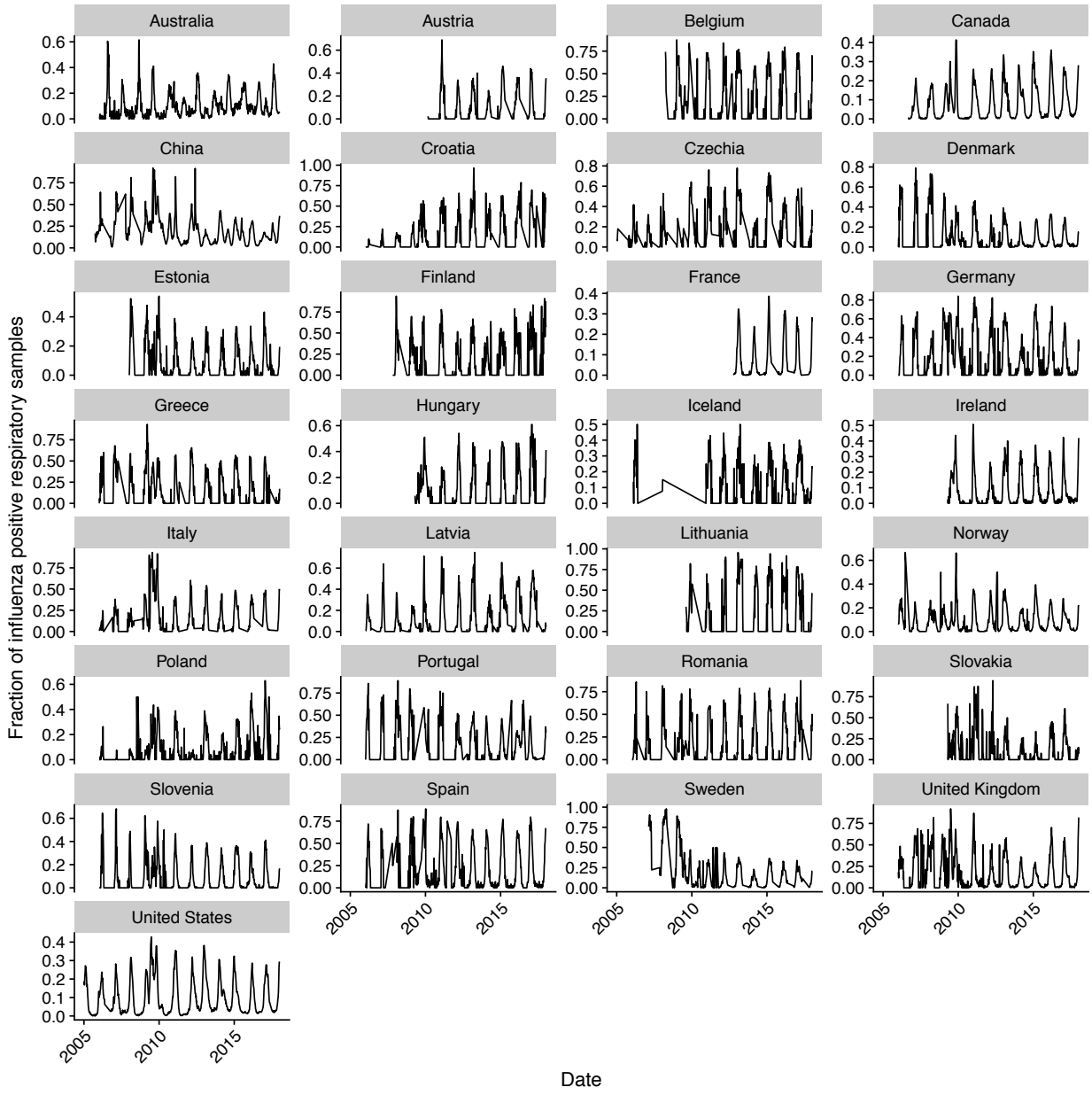


Figure 4.13: The fraction of laboratory tested influenza positive respiratory samples from National Influenza Centers is shown at weekly resolution, as reported in the WHO FluNet database [187].

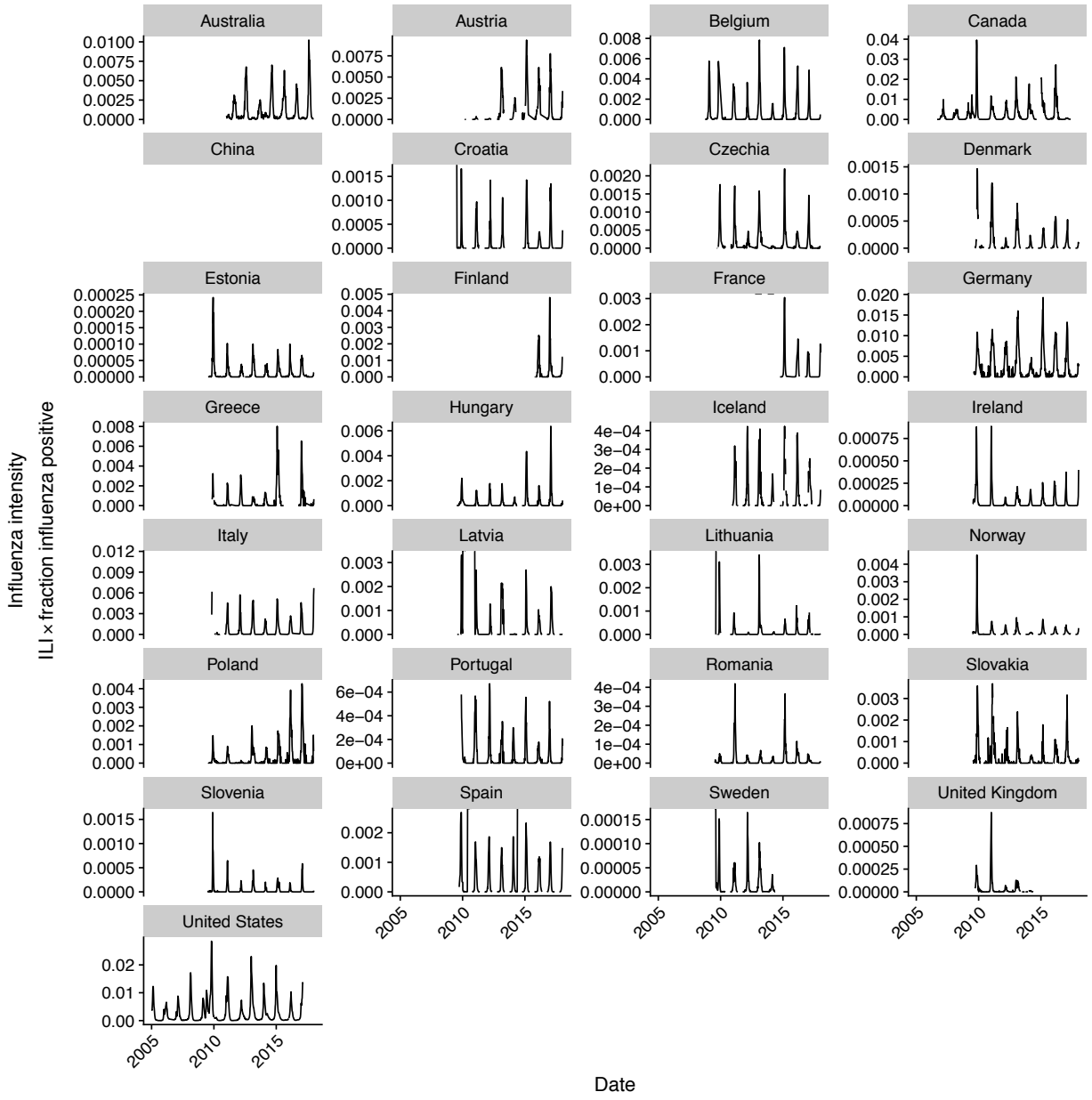


Figure 4.14: Influenza intensity (ILI × fraction of influenza positive respiratory samples [81]) is shown by country at weekly resolution. For China, the influenza intensity is simply the fraction of influenza positive respiratory samples, since ILI (or any other measure of respiratory illness incidence) is not reported.

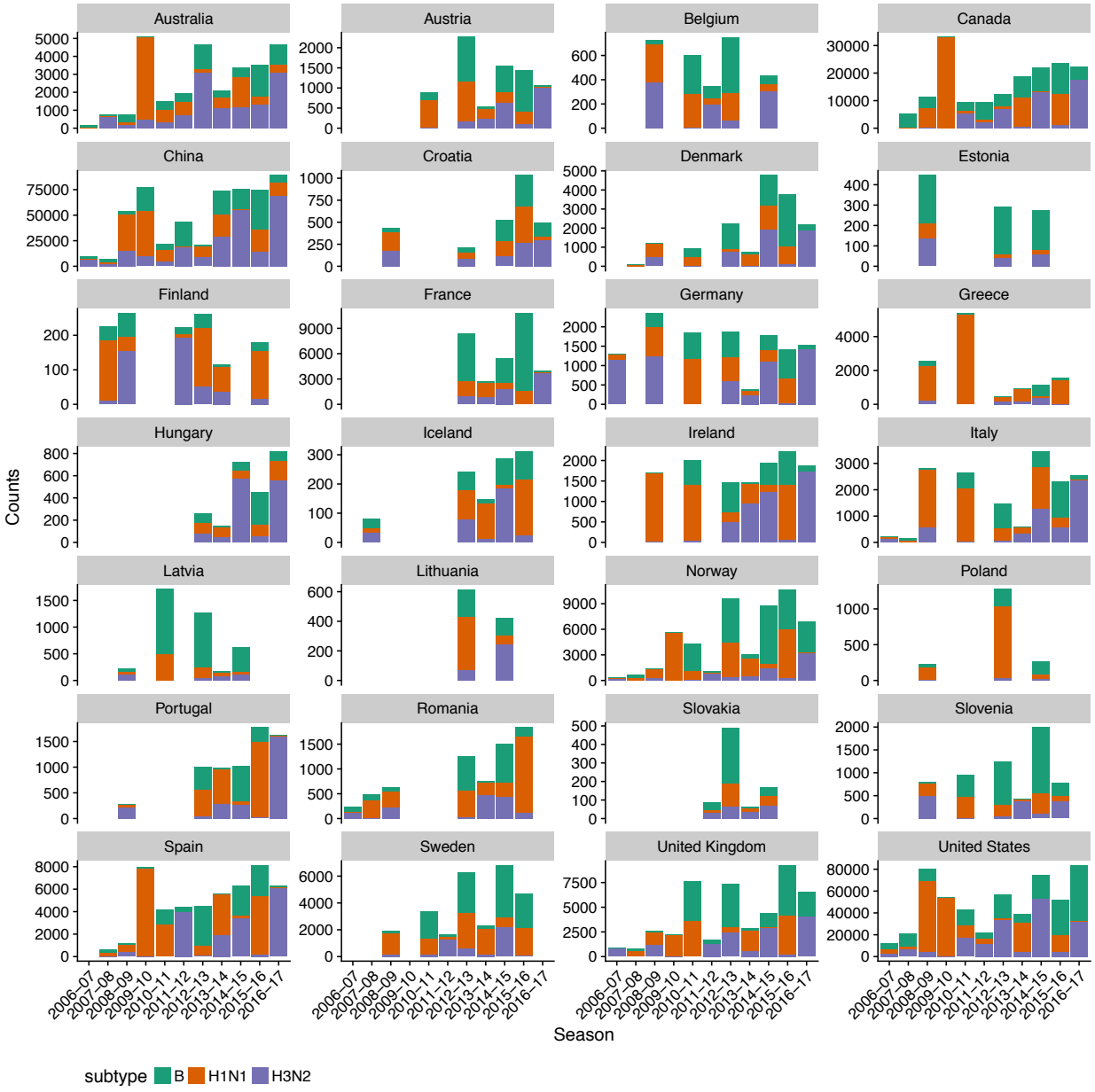


Figure 4.15: Seasonal counts of laboratory-tested respiratory samples identified by type and subtype, as reported in the WHO FluNet database are shown [187]. These data are used in the type and subtype-level analysis of vaccine-driven selection.

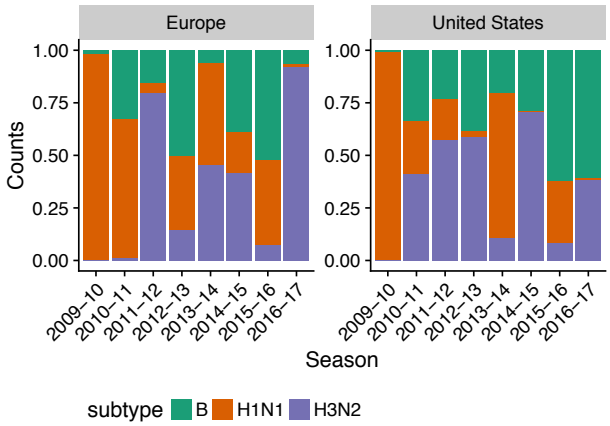


Figure 4.16: Seasonal subtype frequencies in the United States and Europe. European frequencies are calculated as a weighted average of country-level frequencies, weighted by population size.

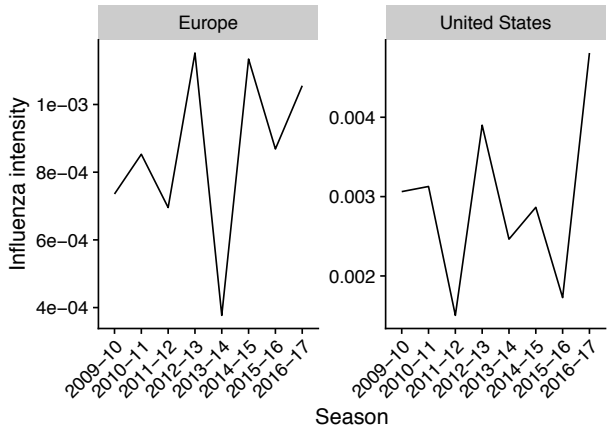


Figure 4.17: Seasonal influenza intensity in the United States and Europe. European influenza intensity is calculated as a weighted average of country-level influenza intensity, weighted by population size.

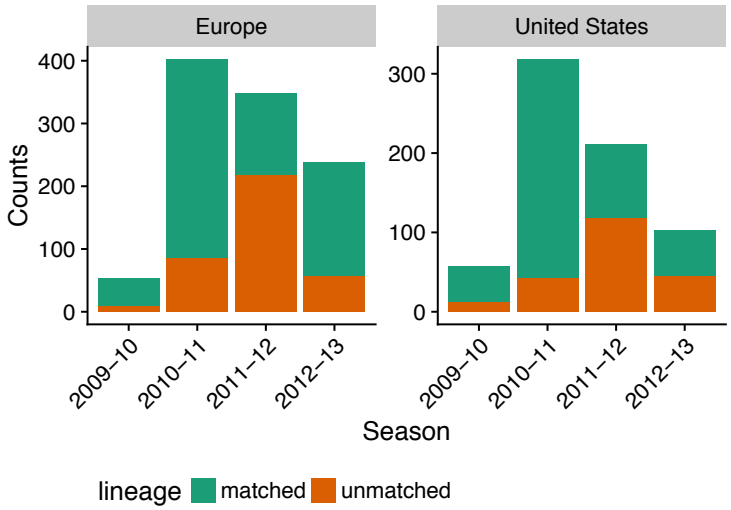


Figure 4.18: Seasonal counts of influenza B sequences contained in the GISAID database. Colors represent lineages that are matched or unmatched to the vaccine strain in the trivalent inactivated vaccine. These data are used in the influenza B lineage-level analysis of vaccine-driven selection.

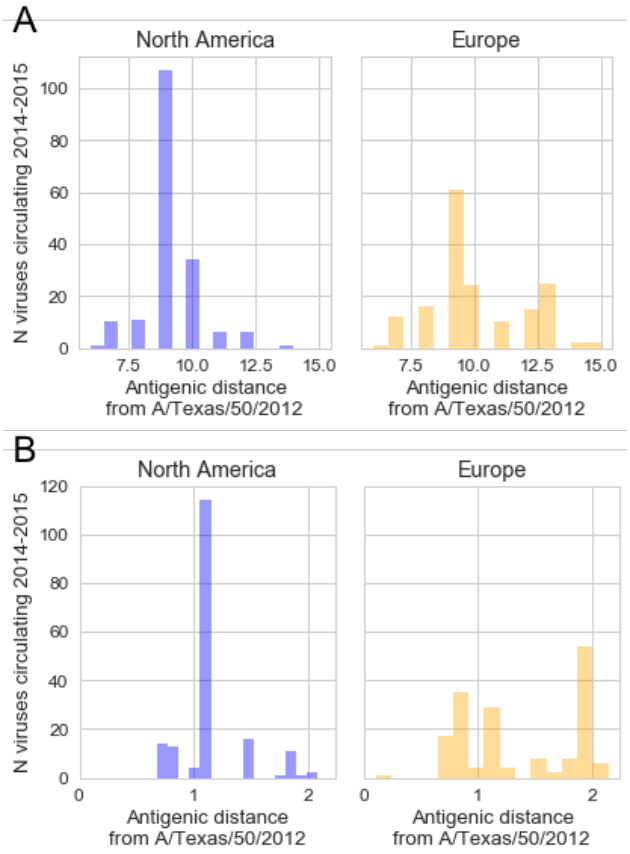


Figure 4.19: Distributions of circulating H3N2 antigenic distances from the 2014-2015 vaccine strain (A/Texas/50/2012), stratified by region. Antigenic distances in (A) are calculated as hamming distances between epitope sites, as defined in [189]. H3N2 strains in North America were more antigenically distant from the vaccine strain by epitope hamming distance (9.2 units, 95% CI: 9.0, 9.4) compared to Europe (10.0 units, 95% CI: 9.7 10.3). Antigenic distances in (B) are calculated using HI titers [127, 88]. North American H3N2 strains are significantly less distant from the vaccine strain (1.17 units, 95% CI: 1.12-1.21) compared to Europe (1.34 units, 95% CI: 1.26-1.42).

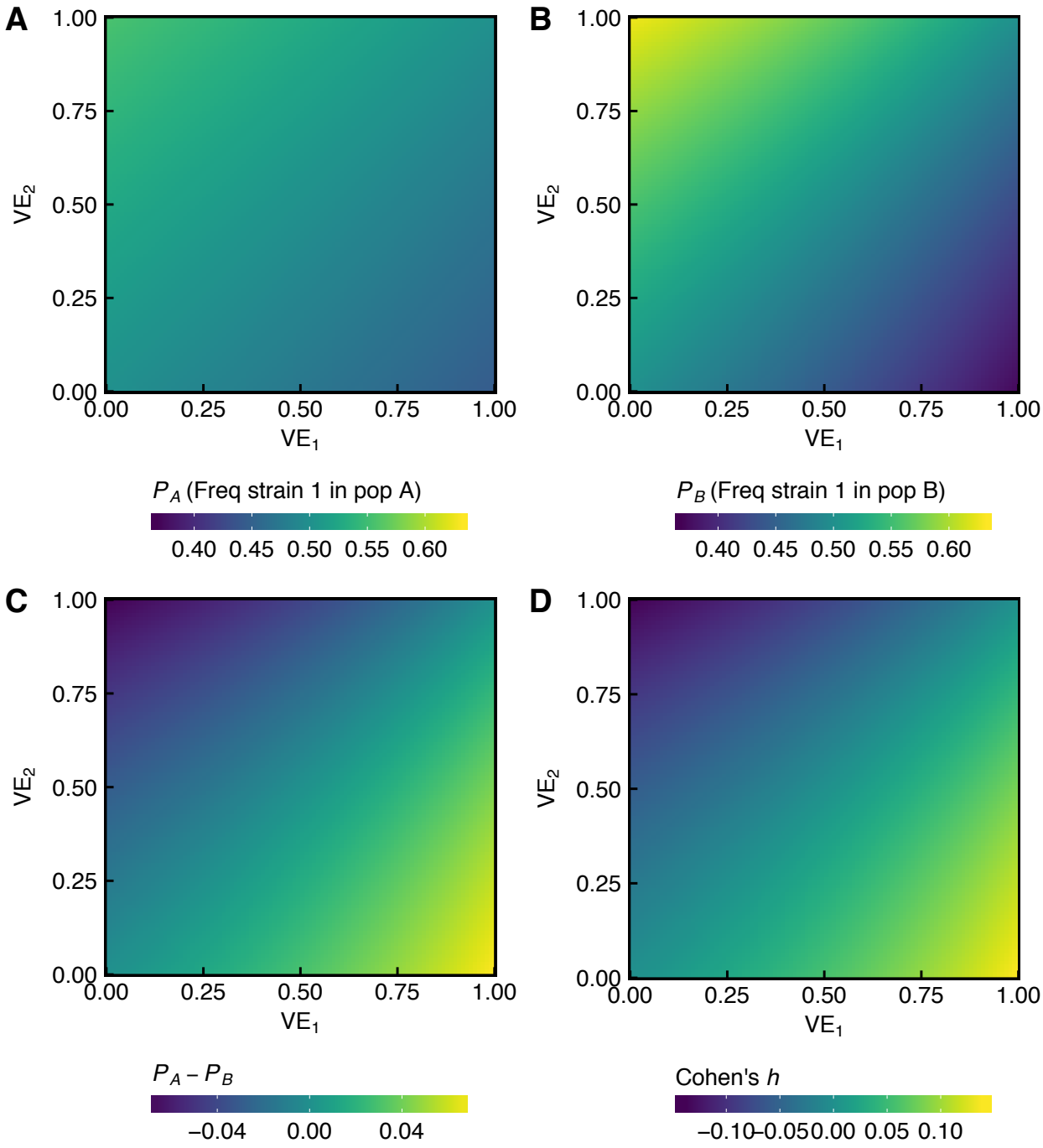


Figure 4.20: Simulated frequencies of two strains $s_1$ and $s_2$, circulating in populations $A$ (with 20% vaccine coverage) and $B$ with (40% vaccine coverage), calculated according to equation 4.27. Given the vaccine effectiveness against both strains, we calculated the fraction of strain 1 (out of strains 1 and 2) in (A) population $A$ and (B) population $B$. The absolute difference in frequencies between regions is shown in (C), and the Cohen's $h$ effect size is shown in (D). These quantities are used to calculate statistical power in Figure 4.21.

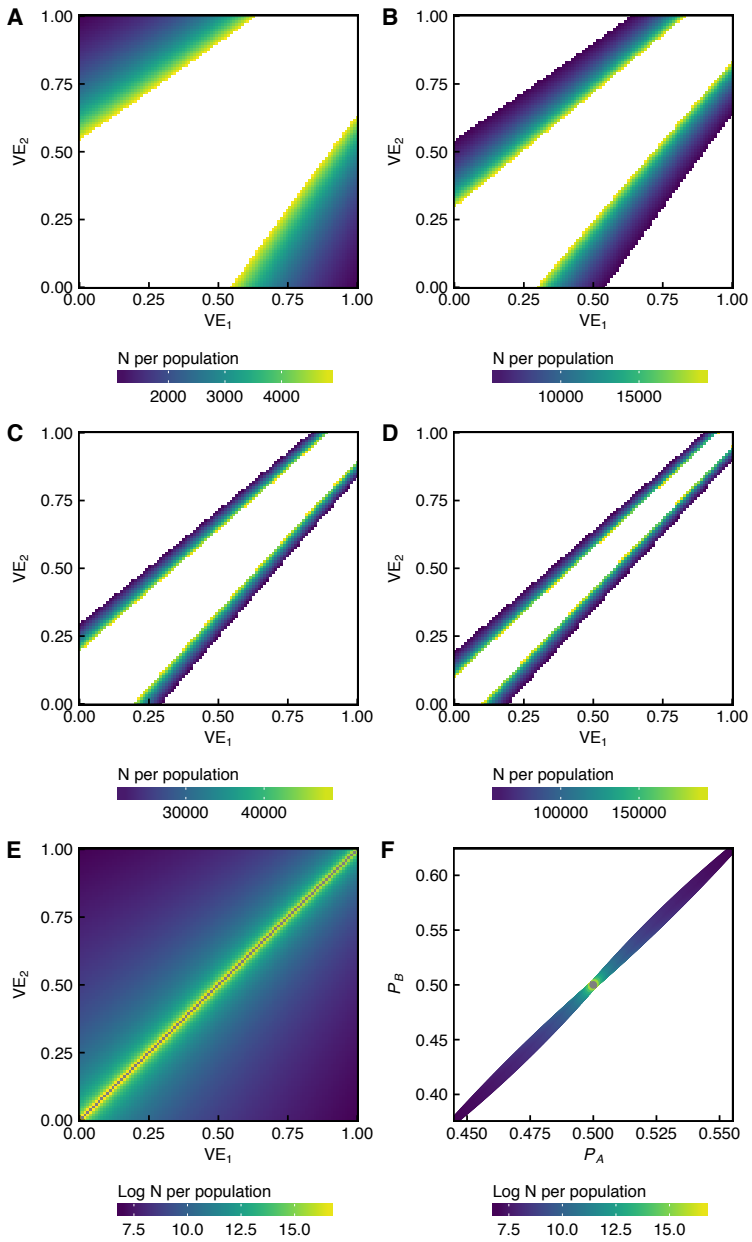


Figure 4.21: Given differences in vaccine effectiveness against strain 1 and strain 2 (Fig 4.20), we estimate the sample sizes per population required to achieve 0.90 statistical power to detect the corresponding difference in strain frequencies between populations A (20% vaccine coverage) and B (40% vaccine coverage) at 0.05 significance, assuming equal distribution of sample sizes over time. Required sample sizes are divided into several ranges for visual clarity in (A-D), and are shown on a log scale in (E). The required sample size to detect a difference in the proportion of strain 1 in population A versus population B (which is determined by the difference in VE against strain 1 and 2) is shown on a log scale in (F). Also in (F), the white space is outside the range of equation 4.27 when VE is between 0 and 1.

# CHAPTER 5

# CONCLUSIONS

This dissertation examines contributions to seasonal influenza's fitness from two approaches. Among host populations, the fitness of strains emerges as a property of host metapopulation ecology and transmission dynamics. For instance, populations with high transmission rates drive faster evolution, and strains that emerge here are better able to compete against globally circulating strains. Strain fitness is also determined by direct interactions with host immunity, which in turn can be modulated to some extent using vaccines. Here, the differential effectiveness of vaccines makes some strains able to infect more hosts than others. These analyses have important implications for the management of seasonal influenza.

Chapter 2 examined the ecological factors that cause influenza strains from certain host populations to be more evolutionarily successful. We tested the effects of demography, seasonality, population size, initial conditions, and transmission rates on a population's ability to produce the most evolutionarily successful strains. We find that higher transmission rates increase the rate of phenotypic evolution; thus, populations with higher transmission rates are more likely to produce evolutionarily successful strains.

Chapter 3 examined the theoretical long-term effects of vaccination on the evolution of seasonal influenza. We found that low rates of vaccination could easily cause eradication. We detected adaptation to the vaccine in extremely rare scenarios when vaccines induced immunity to a much narrower set a strains relative to natural immunity. These results also imply that long-term estimates of vaccines' indirect benefits are potentially greater than presently appreciated, due to additional evolutionary effects of extinction.

Chapter 4 examined potential empirical evidence for vaccine-driven selection of seasonal influenza in temperate populations. We tested for selection based on the premise that differential vaccine effectiveness should create selective gradients that allow less-protected strains to increase in frequency relative to more-protected strains. We found no consistent sup-

port for vaccine-driven strain replacement among influenza types and subtypes and between influenza B lineages. However, during the 2014-2015 season, we detected an increased frequency of H3N2 strains carrying mutations in epitope sites in more vaccinated populations compared to less vaccinated populations, supporting the possibility of vaccine-driven strain replacement within subtypes. We did not find evidence of vaccine-driven strain replacement in other seasons or for H1N1. These results suggest that the seasonal influenza vaccine in general does not produce sufficiently large fitness differences among influenza strains to cause detectable evolutionary changes.

These analyses are largely based on a simplistic view of influenza's interactions with the host immune system. Emerging biological and epidemiological insights highlight important future directions.

## 5.1 Future directions

Individuals with different immune history respond differently to subsequent exposures, creating heterogeneity in the specificity of immunity. At the type and subtype level, individuals tend to respond best to the first dominant antigens [57, 55, 56] they are exposed to, and subsequent exposures often fail to elicit specific responses to novel antigenic sites. Instead, subsequent exposures appear to boost antibody responses against conserved antigens [4, 178, 190]. Consequently, immunity against circulating influenza will vary in time and between individuals.

The effects of immune history have implications for the expected selective effects of vaccines. Among hosts with previous exposure history, vaccines appear to boost any existing responses against antigenic sites that are conserved between the vaccine components and the host's previous exposures [4, 178, 190]. For hosts with recent first exposures, vaccines potentially cause the greatest difference in the strength of immunity directed against an ancestral variant that is conserved in the hosts' exposure history and a circulating variant

mutated at the same site. For example, in the 2014-2015 H3N2-dominated season, the ancestral K166 viruses co-circulated with mutant Q166 viruses [194]. Hosts previously exposed to K166 would have boosted immunity against K166 when vaccinated with the vaccine carrying K166, driving strong selection for Q166. However, if the vaccine strain carried Q166, then vaccinated hosts with previous exposure history would not necessarily mount a Q166-specific response, but might instead mount a response against shared antigens. Thus, the selective versus neutral effects of vaccination depends on a combination of immune history, the composition of circulating viruses, and the vaccine components. Similar trends could be expected for the 2013-2014 and 2015-2016 H1N1-dominated seasons [112, 137, 150]. In contrast to experienced hosts, children without exposure history would be expected to mount a strong primary immune response with antibodies that target the specific vaccine sites [57, 55, 56]. Thus, vaccination of children could be selective regardless of the vaccine strain, assuming that some immunological diversity exists in the viral population.

Identifying correlates of protection will be essential to managing influenza. Much attention has been given to ferret HI-based measurements of antigenicity, which do not always correlate with protection [48]. A simple measure of epitope distance, based on a normalized hamming distance between epitope amino acid sequences, appear better correlated with VE than HI titers [24]. An important distinction between these two metrics is that HI titer closely depends to host immunity, while epitope distance is more a property of the virus itself (although of course, epitope sites are identified because they are targetable by antibodies). While HI titers will vary between individuals and are not always generalizable for predictive purposes, epitope distance can be used to develop baseline expectations that can be subsequently informed using individuals' unique immunity.

Contact patterns may have important implications for influenza's ecology and evolution, especially given potential age-associated heterogeneity in immune specificity [102, 103]. Assuming age-associated immune heterogeneity described above, age-assortative mixing pat-

terns [126] could concentrate immune selection in children and adolescents. In contrast, long-range transmission facilitated mostly by adults may be mostly neutral. Models that incorporate both contact network structure and heterogeneity in the acquisition of immune specificity will elucidate how viral evolution in subpopulations contribute to the global evolutionary patterns.

# REFERENCES


[1] H J Adam, S E Richardson, F B Jamieson, P Rawte, D E Low, and D N Fisman. Changing epidemiology of invasive Haemophilus influenzae in Ontario, Canada: Evidence for herd effects and strain replacement due to Hib vaccination. *Vaccine*, 28(24):4073–4078, May 2010.

[2] Ben Adams and Alice Carolyn McHardy. The impact of seasonal and year-round transmission regimes on the evolution of influenza A virus. *Proceedings. Biological sciences / The Royal Society*, 278(1716):2249–2256, August 2011.

[3] F R Adler. The effects of averaging on the basic reproduction ratio. *Mathematical Biosciences*, 111(1):89–98, September 1992.

[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):316ra192, dec 2015.

[5] Nimalan Arinaminpathy, Oliver Ratmann, Katia Koelle, Suzanne L Epstein, Graeme E Price, Cecile Viboud, Mark A Miller, and Bryan T Grenfell. Impact of cross-protective vaccines on epidemiological and evolutionary dynamics of influenza. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8):3173–3177, February 2012.

[6] Justin Bahl, Martha I Nelson, Kwok H Chan, Rubing Chen, Dhanasekaran Vijaykrishna, Rebecca A Halpin, Timothy B Stockwell, Xudong Lin, David E Wentworth, Elodie Ghedin, Yi Guan, J S Malik Peiris, Steven Riley, Andrew Rambaut, Edward C Holmes, and Gavin J D Smith. Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 108(48):19359–19364, November 2011.

[7] Shweta Bansal, Babak Pourbohloul, and Lauren Ancel Meyers. A comparative analysis of influenza vaccination programs. *PLoS Medicine*, 3(10):e387–10, October 2006.

[8] P Noel Barrett, Gregory Berezuk, Sandor Fritsch, Gerald Aichinger, Mary Kate Hart, Wael El-Amin, Otfried Kistner, and Hartmut J Ehrlich. Efficacy, safety, and immunogenicity of a Vero-cell-culture-derived trivalent influenza vaccine: a multicentre, double-blind, randomised, placebo-controlled trial. *The Lancet*, 377(9767):751–759, feb 2011.

[9] B. Bean, B. M. Moore, B. Sterner, L. R. Peterson, D. N. Gerding, and H. H. Balfour. Survival of Influenza Viruses on Environmental Surfaces. *Journal of Infectious Diseases*, 146(1):47–51, jul 1982.

[10] Trevor Bedford, Sarah Cobey, Peter Beerli, and Mercedes Pascual. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathogens*, 6(5):e1000918, May 2010.

[11] Trevor Bedford, Andrew Rambaut, and Mercedes Pascual. Canalization of the evolutionary trajectory of the human influenza virus. *BMC Biology*, 10(1):38, 2012.

[12] Trevor Bedford, Steven Riley, Ian G Barr, Shobha Broor, Mandeep Chadha, Nancy J Cox, Rodney S Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, Alexander Klimov, Nicola S Lewis, Xiyan Li, John W McCauley, Takato Odagiri, Varsha Potdar, Andrew Rambaut, Yuelong Shu, Eugene Skepner, Derek J Smith, Marc A Suchard, Masato Tashiro, Dayan Wang, Xiyan Xu, Philippe Lemey, and Colin A Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–220, July 2015.

[13] Trevor Bedford, Marc A Suchard, Philippe Lemey, Gytis Dudas, Victoria Gregory, Alan J Hay, John W McCauley, Colin A Russell, Derek J Smith, and Andrew Rambaut. Integrating influenza antigenic dynamics with molecular evolution. *eLife*, 3:e01914, 2014.

[14] Edward A. Belongia, Burney A. Kieke, James G. Donahue, Laura A. Coleman, Stephanie A. Irving, Jennifer K. Meece, Mary Vandermause, Stephen Lindstrom, Paul Gargiullo, and David K. Shay. Influenza vaccine effectiveness in Wisconsin during the 2007–08 season: Comparison of interim and final results. *Vaccine*, 29(38):6558–6563, sep 2011.

[15] Edward A Belongia, Burney A Kieke, James G Donahue, Robert T Greenlee, Amanda Balish, Angie Foust, Stephen Lindstrom, David K Shay, and Marshfield Influenza Study Group. Effectiveness of inactivated influenza vaccines varied substantially with antigenic match from the 2004-2005 season to the 2006-2007 season. *The Journal of Infectious Diseases*, 199(2):159–167, January 2009.

[16] Edward A Belongia, Melissa D Simpson, Jennifer P King, Maria E Sundaram, Nicholas S Kelley, Michael T Osterholm, and Huong Q McLean. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *The Lancet. Infectious diseases*, 16(8):942–951, August 2016.

[17] Jiri Beran, Timo Vesikari, Veronika Wertzova, Aino Karvonen, Karel Honegr, Niklas Lindblad, Pascale Van Belle, Mathieu Peeters, Bruce L Innis, and Jeanne-Marie Devaster. Efficacy of inactivated split-virus influenza vaccine against culture-confirmed influenza in healthy adults: a prospective, randomized, placebo-controlled trial. *The Journal of infectious diseases*, 200(12):1861–9, dec 2009.

[18] Jirí Beran, Veronika Wertzova, Karel Honegr, Eva Kaliskova, Martina Havlickova, Jirí Havlik, Helena Jirincova, Pascale Van Belle, Varsha Jain, Bruce Innis, and Jeanne-Marie Devaster. Challenge of conducting a placebo-controlled randomized efficacy

study for influenza vaccine in a season with low attack rate and a mismatched vaccine B strain: a concrete example. *BMC Infectious Diseases*, 9(1):2, January 2009.

[19] Matthew Biggerstaff, Simon Cauchemez, Carrie Reed, Manoj Gambhir, and Lyn Finelli. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infectious Diseases*, 14(1):1–20, September 2014.

[20] Stan L. Block, Judith Falloon, Jeffrey A. Hirschfield, Leonard R. Krilov, Filip Dubovsky, Tingting Yi, and Robert B. Belshe. Immunogenicity and Safety of a Quadrivalent Live Attenuated Influenza Vaccine in Children. *The Pediatric Infectious Disease Journal*, 31(7):745–751, jul 2012.

[21] Stan L. Block, Tingting Yi, Eric Sheldon, Filip Dubovsky, and Judith Falloon. A randomized, double-blind noninferiority study of quadrivalent live attenuated influenza vaccine in adults. *Vaccine*, 29(50):9391–9397, nov 2011.

[22] 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*, 18(3):469–476, March 2011.

[23] M F Boni, J R Gog, V Andreasen, and M W Feldman. Epidemic dynamics and antigenic evolution in a single season of influenza A. *Proceedings of the Royal Society B*, 273(1592):1307–1316, June 2006.

[24] Melia E Bonomo and Michael W Deem. Predicting Influenza H3N2 Vaccine Efficacy from Evolution of the Dominant Epitope. *Clinical Infectious Diseases*, (May), apr 2018.

[25] Noel T Brewer, Gretchen B Chapman, Frederick X Gibbons, Meg Gerrard, Kevin D McCaul, and Neil D Weinstein. Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology*, 26(2):136–145, 2007.

[26] C B Bridges, W W Thompson, M I Meltzer, G R Reeve, W J Talamonti, N J Cox, H A Lilac, H Hall, A Klimov, and K Fukuda. Effectiveness and cost-benefit of influenza vaccination of healthy working adults: A randomized controlled trial. *JAMA*, 284(13):1655–1663, October 2000.

[27] Sarah A Buchan and Jeffrey C Kwong. Trends in influenza vaccine coverage and vaccine hesitancy in Canada, 2006/07 to 2013/14: results from cross-sectional survey data. *CMAJ open*, 4(3):E455–E462, aug 2016.

[28] W F Carman, A R Zanetti, P Karayiannis, J Waters, G Manzillo, E Tanzi, A J Zuckerman, and H C Thomas. Vaccine-induced escape mutant of hepatitis B virus. *The Lancet*, 336(8711):325–329, August 1990.

[29] F Carrat and A Flahault. Influenza vaccine: The challenge of antigenic drift. *Vaccine*, 25(39-40):6852–6862, September 2007.

[30] 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, April 2008.

[31] Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boëlle, Antoine Flahault, and Neil M Ferguson. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 452(7188):750–754, April 2008.

[32] CDC. *FluVaxView*. Centers for Disease Control and Prevention, 2015.

[33] CDC. Seasonal influenza vaccine and total doses distributed. Technical report, Centers for Disease Control and Prevention, 2016.

[34] Centers for Disease Control and Prevention. *Quadrivalent Influenza Vaccine*, 2018.

[35] Centers for Disease Control and Prevention. *Seasonal Influenza Vaccination Resources for Health Professionals*, 2018.

[36] Benjamin S Chambers, Kaela Parkhouse, Ted M Ross, Kevin Alby, and Scott E Hensley. Identification of Hemagglutinin Residues Responsible for H3N2 Antigenic Drift during the 2014-2015 Influenza Season. *Cell reports*, 12(1):1–6, jul 2015.

[37] Joseph Chan, Antony Holmes, and Raul Rabadan. Network Analysis of Global Influenza Spread. *PLoS computational biology*, 6(11):e1001005–10, November 2010.

[38] Dennis L Chao, M Elizabeth Halloran, and Ira M Longini Jr. School Opening Dates Predict Pandemic Influenza A(H1N1) Outbreaks in the United States. *The Journal of Infectious Diseases*, 202(6):877–880, September 2010.

[39] G B Chapman and E J Coups. Predictors of influenza vaccine acceptance among healthy adults. *Preventive medicine*, 29(4):249–262, October 1999.

[40] Rubing Chen and Edward C Holmes. The evolutionary dynamics of human influenza B virus. *Journal of Molecular Evolution*, 66(6):655–663, June 2008.

[41] 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.e10, apr 2018.

[42] X Cheng, Y Tan, M He, T T Y Lam, X Lu, C Viboud, J He, S Zhang, J Lu, C Wu, S Fang, X Wang, X Xie, H Ma, M I Nelson, H f Kung, E C Holmes, and J Cheng. Epidemiological dynamics and phylogeography of influenza virus in southern China. *Journal of Infectious Diseases*, 207(1):106–114, January 2013.

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

[44] CIESIN. Center for International Earth Science Information Network - CIESIN - Columbia University, United Nations Food and Agriculture Programme - FAO, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid, Future Estimates. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)., 2018.

[45] R. D. Clover, S. Crawford, W. P. Glezen, L. H. Taber, C. C. Matson, and R. B. Couch. Comparison of Heterotypic Protection against Influenza A/Taiwan/86 (H1N1) by Attenuated and Inactivated Vaccines to A/Chile/83-like Viruses. *Journal of Infectious Diseases*, 163(2):300–3004, feb 1991.

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

[47] Sarah Cobey, Sigrid Gouma, Kaela Parkhouse, Benjamin S Chambers, Hildegund C Ertl, Kenneth E Schmader, Rebecca A Halpin, Xudong Lin, Timothy B Stockwell, Suman R Das, Emily Landon, Vera Tesic, Ilan Youngster, Benjamin A Pinsky, David E Wentworth, Scott E Hensley, and Yonatan H Grad. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine effectiveness in 2012-13. *Clinical Infectious Diseases*, 2018.

[48] Sarah Cobey, Sigrid Gouma, Kaela Parkhouse, Benjamin S Chambers, Hildegund C Ertl, Kenneth E Schmader, Rebecca A Halpin, Xudong Lin, Timothy B Stockwell, Suman R Das, Emily Landon, Vera Tesic, Ilan Youngster, Benjamin A Pinsky, David E Wentworth, Scott E Hensley, and Yonatan H Grad. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine effectiveness in 2012-13. *Clinical Infectious Diseases*, 2012.

[49] Sarah Cobey and Scott E Hensley. Immune history and influenza virus susceptibility. *Current opinion in virology*, 22:105–111, 2017.

[50] Davide Corti, Amorsolo L Suguitan, Debora Pinna, Chiara Silacci, Blanca M Fernandez-Rodriguez, Fabrizia Vanzetta, Celia Santos, Catherine J Luke, Fernando J Torres-Velez, Nigel J Temperton, Robin A Weiss, Federica Sallusto, Kanta Subbarao, and Antonio Lanzavecchia. Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *The Journal of clinical investigation*, 120(5):1663–73, may 2010.

[51] Benjamin J Cowling, Eric H Y Lau, Conrad L H Lam, Calvin K Y Cheng, Jana Kovar, Kwok Hung Chan, J S Malik Peiris, and Gabriel M Leung. Effects of school closures, 2008 winter influenza season, Hong Kong. *Emerging infectious diseases*, 14(10):1660–1662, October 2008.

[52] N J Cox and C A Bender. The molecular epidemiology of influenza viruses. *Seminars in Virology*, 1995.

[53] N J Cox, T L Brammer, and H L Regnery. Influenza: global surveillance for epidemic and pandemic variants. *European journal of epidemiology*, 10(4):467–470, August 1994.

[54] N J Cox and K Subbarao. Global epidemiology of influenza: past and present. *Annual review of medicine*, 51:407–421, 2000.

[55] F M DAVENPORT and A V HENNESSY. A serologic recapitulation of past experiences with influenza A; antibody response to monovalent vaccine. *The Journal of experimental medicine*, 104(1):85–97, jul 1956.

[56] F M DAVENPORT and A V HENNESSY. Predetermination by infection and by vaccination of antibody response to influenza virus vaccines. *The Journal of experimental medicine*, 106(6):835–50, dec 1957.

[57] F M Davenport, A V HENNESSY, and T Francis. Epidemiologic and immunologic significance of age distribution of antibody to antigenic variants of influenza virus. *The Journal of experimental medicine*, 98(6):641–656, December 1953.

[58] Vittorio Demicheli, Tom Jefferson, Eliana Ferroni, Alessandro Rivetti, and Carlo Di Pietrantonj. Vaccines for preventing influenza in healthy adults. *Cochrane Database of Systematic Reviews*, 2:CD001269, feb 2018.

[59] Andrew Dobson. Population dynamics of pathogens with multiple host species. *The American naturalist*, 164 Suppl 5(S5):S64–78, nov 2004.

[60] Jonathan Dushoff and Simon Levin. The effects of population heterogeneity on disease invasion. *Mathematical Biosciences*, 128(1-2):25–40, jul 1995.

[61] Amalie Dyda, Surendra Karki, Andrew Hayen, C. Raina MacIntyre, Robert Menzies, Emily Banks, John M Kaldor, and Bette Liu. Influenza and pneumococcal vaccination in Australian adults: A systematic review of coverage and factors associated with uptake. *BMC Infectious Diseases*, 16(1), 2016.

[62] ECDC. Influenza surveillance country, territory, and area profiles 2017. Technical report, World Health Organization, European Centre for Disease Prevention and Control, 2017.

[63] ECDC. Seasonal influenza vaccination in europe. vaccination recommendations and coverage rates in the eu member states for eight influenza seasons: 2007–2008 to 2014–2015. Technical report, European Centre for Disease Prevention and Control, 2017.

[64] Damian C Ekiert, Gira Bhabha, Marc-André Elsliger, Robert H E Friesen, Mandy Jongeneelen, Mark Throsby, Jaap Goudsmit, and Ian A Wilson. Antibody recognition of a highly conserved influenza virus epitope. *Science (New York, N.Y.)*, 324(5924):246–51, apr 2009.

[65] A Elomaa, A Advani, D Donnelly, M Antila, J Mertsola, H Hallander, and Q He. Strain Variation among Bordetella pertussis Isolates in Finland, Where the Whole-Cell Pertussis Vaccine Has Been Used for 50 Years. *Journal of Clinical Microbiology*, 43(8):3681–3687, August 2005.

[66] Daniel R. Feikin, Eunice W. Kagucia, Jennifer D. Loo, Ruth Link-Gelles, Milo A. Puhan, Thomas Cherian, Orin S. Levine, Cynthia G. Whitney, Katherine L. O'Brien, Matthew R. Moore, and the Serotype Replacement Study Group. Serotype-Specific Changes in Invasive Pneumococcal Disease after Pneumococcal Conjugate Vaccine Introduction: A Pooled Analysis of Multiple Surveillance Sites. *PLoS Medicine*, 10(9):e1001517, sep 2013.

[67] James E Fielding, Kristina A Grant, Georgina Papadakis, and Heath A Kelly. Estimation of type- and subtype-specific influenza vaccine effectiveness in Victoria, Australia using a test negative case control method, 2007-2008. *BMC Infectious Diseases*, 11(1):170, dec 2011.

[68] Brian S. Finkelman, Cécile Viboud, Katia Koelle, Matthew J. Ferrari, Nita Bharti, and Bryan T. Grenfell. Global Patterns in Seasonal Activity of Influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral Coexistence and Latitudinal Gradients. *PLoS ONE*, 2(12):e1296, dec 2007.

[69] W M Fitch, J M Leiter, X Q Li, and P Palese. Positive Darwinian evolution in human influenza A viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 88(10):4270–4, may 1991.

[70] Brendan Flannery. Preliminary end-of-season estimates of 2016–17 seasonal influenza vaccine effectiveness against medically attended influenza from the us flu ve network. *National Adult and Influenza Immunization Summit*, 2017.

[71] Brendan Flannery, Richard K Zimmerman, Larisa V Gubareva, Rebecca J Garten, Jessie R Chung, Mary Patricia Nowalk, Michael L Jackson, Lisa A Jackson, Arnold S Monto, Suzanne E Ohmit, Edward A Belongia, Huong Q McLean, Manjusha Gaglani, Pedro A Piedra, Vasiliy P Mishin, Anton P Chesnokov, Sarah Spencer, Swathi N Thaker, John R Barnes, Angie Foust, Wendy Sessions, Xiyan Xu, Jacqueline Katz, and Alicia M Fry. Enhanced Genetic Characterization of Influenza A(H3N2) Viruses and Vaccine Effectiveness by Genetic Group, 2014-2015. *The Journal of infectious diseases*, 214(7):1010–9, oct 2016.

[72] Emuella M Flood, Matthew D Rousculp, Kellie J Ryan, Kathleen M Beusterien, Victoria M Divino, Seth L Toback, Medha Sasané, Stan L Block, Matthew C Hall, and

Parthiv J Mahadevia. Parents' decision-making regarding vaccinating their children against influenza: A web-based survey. *Clinical therapeutics*, 32(8):1448–1467, August 2010.

[73] Christian Flück, Sonja Schöpflin, Tom Smith, Blaise Genton, Michael P. Alpers, Hans-Peter Beck, and Ingrid Felger. Effect of the malaria vaccine Combination B on merozoite surface antigen 2 diversity. *Infection, Genetics and Evolution*, 7(1):44–51, jan 2007.

[74] Sharon Frey, Timo Vesikari, Agnieszka SzymczakiewiczMultanowska, Maria Lattanzi, Allen Izu, Nicola Groth, and Sandra Holmes. Clinical Efficacy of Cell Culture–Derived and EggDerived Inactivated Subunit Influenza Vaccines in Healthy Adults. *Clinical Infectious Diseases*, 51(9):997–1004, nov 2010.

[75] Manjusha Gaglani, Jessica Pruszynski, Kempapura Murthy, Lydia Clipper, Anne Robertson, Michael Reis, Jessie R Chung, Pedro A Piedra, Vasanthi Avadhanula, Mary Patricia Nowalk, Richard K Zimmerman, Michael L Jackson, Lisa A Jackson, Joshua G Petrie, Suzanne E Ohmit, Arnold S Monto, Huong Q McLean, Edward A Belongia, Alicia M Fry, and Brendan Flannery. Influenza Vaccine Effectiveness Against 2009 Pandemic Influenza A(H1N1) Virus Differed by Vaccine Type During 2013-2014 in the United States. *The Journal of infectious diseases*, 213(10):1546–56, may 2016.

[76] Alison P Galvani, Timothy C Reluga, and Gretchen B Chapman. Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. *Proceedings of the National Academy of Sciences*, 104(13):5692–5697, March 2007.

[77] G F Gause. Experimental Studies on the Struggle for Existence. *Journal of Experimental Biology*, 9(4), 1932.

[78] Blaise Genton, Inoni Betuela, Ingrid Felger, Fadwa Al-Yaman, Robin F. Anders, Allan Saul, Lawrence Rare, Moses Baisor, Kerry Lorry, Graham V. Brown, David Pye, David O. Irving, Thomas A. Smith, Hans-Peter Beck, and Michael P. Alpers. A Recombinant Blood-Stage Malaria Vaccine Reduces Plasmodium falciparum Density and Exerts Selective Pressure on Parasite Populations in a Phase 1–2b Trial in Papua New Guinea. *The Journal of Infectious Diseases*, 185(6):820–827, mar 2002.

[79] Linda Gionet. Health at a Glance: Flu vaccination rates in Canada. *National Advisory Committee on Immunization*, (82-624-X):7, 2015.

[80] Julia R Gog and Bryan T Grenfell. Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences*, 99(26):17209–17214, December 2002.

[81] Edward Goldstein, Sarah Cobey, Saki Takahashi, Joel C. Miller, and Marc Lipsitch. Predicting the Epidemic Sizes of Influenza A/H1N1, A/H3N2, and B: A Statistical Method. *PLoS Medicine*, 8(7):e1001051, jul 2011.

[82] 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*, 303(5656):327–332, January 2004.

[83] William C. Gruber, Larry H. Taber, W. Paul Glezen, Richard D. Clover, Troy D. Abell, Richard W. Demmler, and Robert B. Couch. Live Attenuated and Inactivated Influenza Vaccine in School-age Children. *Archives of Pediatrics & Adolescent Medicine*, 144(5):595, may 1990.

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

[85] S Gupta, K Trenholme, R M Anderson, and K P Day. Antigenic diversity and the transmission dynamics of Plasmodium falciparum. *Science (New York, N.Y.)*, 263(5149):961–3, feb 1994.

[86] 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, apr 1996.

[87] Vishal Gupta, David J Earl, and Michael W Deem. Quantifying influenza vaccine efficacy and antigenic distance. *Vaccine*, 24(18):3881–3888, May 2006.

[88] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *bioRxiv*, page 224048, nov 2017.

[89] Anthony Heymann, Gabriel Chodick, Brian Reichman, Ehud Kokia, and Joseph Laufer. Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. *The Pediatric infectious disease journal*, 23(7):675–677, July 2004.

[90] Alejandro Hoberman, David P. Greenberg, Jack L. Paradise, Howard E. Rockette, Judith R. Lave, Diana H. Kearney, D. Kathleen Colborn, Marcia Kurs-Lasky, Mary Ann Haralam, Carol J. Byers, Lisa M. Zoffel, Irene A. Fabian, Beverly S. Bernard, and Jill D. Kerr. Effectiveness of Inactivated Influenza Vaccine in Preventing Acute Otitis Media in Young Children. *JAMA*, 290(12):1608–1616, sep 2003.

[91] Peter Horby, Quang Thai Pham, Niel Hens, Thi Thu Yen Nguyen, Quynh Mai Le, Dinh Thoang Dang, Manh Linh Nguyen, Thu Huong Nguyen, Neal Alexander, W John Edmunds, Nhu Duong Tran, Annette Fox, and Tran Hien Nguyen. Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS ONE*, 6(2):e16965, 2011.

[92] E S Hurwitz, M Haber, A Chang, T Shope, S Teo, M Ginsberg, N Waecker, and N J Cox. Effectiveness of influenza vaccination of day care children in reducing influenza-related morbidity among household contacts. *JAMA*, 284(13):1677–1682, October 2000.

[93] ICAO. *Annual report of the ICAO council: 2014*. International Civil Aviation Organization, 2014.

[94] Charlotte Jackson, Emilia Vynnycky, and Punam Mangtani. Estimates of the transmissibility of the 1968 (Hong Kong) influenza pandemic: evidence of increased transmissibility between successive waves. *American Journal of Epidemiology*, 171(4):465–478, February 2010.

[95] Lisa A Jackson, Manjusha J Gaglani, Harry L Keyserling, John Balser, Nancy Bouveret, Louis Fries, and John J Treanor. Safety, efficacy, and immunogenicity of an inactivated influenza vaccine in healthy adults: a randomized, placebo-controlled trial over two influenza seasons. *BMC Infectious Diseases*, 10(1):71, March 2010.

[96] Michael L Jackson, Jessie R Chung, Lisa A Jackson, C Hallie Phillips, Joyce Benoit, Arnold S Monto, Emily T Martin, Edward A Belongia, Huong Q McLean, Manjusha Gaglani, Kempapura Murthy, Richard Zimmerman, Mary P Nowalk, Alicia M Fry, and Brendan Flannery. Influenza Vaccine Effectiveness in the United States during the 2015-2016 Season. *The New England journal of medicine*, 377(6):534–543, 2017.

[97] Naveed Z. Janjua, Danuta M. Skowronski, Gaston De Serres, Jim Dickinson, Natasha S. Crowcroft, Marsha Taylor, Anne-Luise Winter, Travis S. Hottes, Kevin Fonseca, Hugues Charest, Steven J. Drews, Suzana Sabaiduc, Nathalie Bastien, Yan Li, Jennifer L. Gardy, and Martin Petric. Estimates of Influenza Vaccine Effectiveness for 2007–2008 From Canada's Sentinel Surveillance System: Cross-Protection Against Major and Minor Variants. *The Journal of Infectious Diseases*, 205(12):1858–1868, jun 2012.

[98] Tom Jefferson, Alessandro Rivetti, Carlo Di Pietrantonj, and Vittorio Demicheli. Vaccines for preventing influenza in healthy children. *Cochrane Database of Systematic Reviews*, 2:CD004879, feb 2018.

[99] Barbara Jester, Joy Schwerzmann, Desiree Mustaquim, Tricia Aden, Lynnette Brammer, Rosemary Humes, Pete Shult, Shahram Shahangian, Larisa Gubareva, Xiyan Xu, Joseph Miller, and Daniel Jernigan. Mapping of the US Domestic Influenza Virologic Surveillance Landscape. *Emerging Infectious Diseases*, 24(7):1300–1306, jul 2018.

[100] David A Kennedy and Andrew F Read. Why does drug resistance readily evolve but vaccine resistance does not? *Proceedings. Biological sciences*, 284(1851):20162562, mar 2017.

[101] Adam Kucharski and Julia R Gog. Influenza emergence in the face of evolutionary constraints. *Proceedings. Biological sciences / The Royal Society*, 279(1729):645–652, February 2012.

[102] Adam J. Kucharski and Julia R. Gog. The Role of Social Contacts and Original Antigenic Sin in Shaping the Age Pattern of Immunity to Seasonal Influenza. *PLoS Computational Biology*, 8(10):e1002741, oct 2012.

[103] 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, jun 2014.

[104] Kin On Kwok, Steven Riley, Ranawaka A. P. M. Perera, Vivian W. I. Wei, Peng Wu, Lan Wei, Daniel K. W. Chu, Ian G. Barr, J. S. Malik Peiris, and Benjamin J. Cowling. Relative incidence and individual-level severity of seasonal influenza A H3N2 compared with 2009 pandemic H1N1. *BMC Infectious Diseases*, 17(1):337, dec 2017.

[105] C W Lee, D A Senne, and D L Suarez. Effect of Vaccine Use in the Evolution of Mexican Lineage H5N2 Avian Influenza Virus. *Journal of Virology*, 78(15):8372–8381, July 2004.

[106] Philippe Lemey, Andrew Rambaut, Trevor Bedford, Nuno Faria, Filip Bielejec, Guy Baele, Colin A Russell, Derek J Smith, Oliver G Pybus, Dirk Brockmann, and Marc A Suchard. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathogens*, 10(2):e1003932, February 2014.

[107] Avram Levy, Sheena G. Sullivan, Simone S. Tempone, Kerry L.M. Wong, Annette K. Regan, Gary K. Dowse, Paul V. Effler, and David W. Smith. Influenza vaccine effectiveness estimates for Western Australia during a period of vaccine and virus strain stability, 2010 to 2012. *Vaccine*, 32(47):6312–6318, oct 2014.

[108] Joseph Lewnard and Sarah Cobey. Immune History and Influenza Vaccine Effectiveness. *Vaccines*, 6(2):28, may 2018.

[109] Joseph A Lewnard, Christine Tedijanto, Benjamin J Cowling, and Marc Lipsitch. Quantifying biases in test-negative studies of vaccine effectiveness. *bioRxiv*, 2017.

[110] Xi Li and Michael W Deem. Influenza evolution and H3N2 vaccine effectiveness, with application to the 2014/2015 season. *Protein engineering, design & selection : PEDS*, 29(8):309–15, 2016.

[111] Juan Lin, Viggo Andreasen, Renato Casagrandi, and Simon A Levin. Traveling waves in a model of influenza A drift. *Journal of theoretical biology*, 222(4):437–445, June 2003.

[112] Susanne L Linderman, Benjamin S Chambers, Seth J Zost, Kaela Parkhouse, Yang Li, Christin Herrmann, Ali H Ellebedy, Donald M Carter, Sarah F Andrews, Nai-Ying Zheng, Min Huang, Yunping Huang, Donna Strauss, Beth H Shaz, Richard L Hodinka, Gustavo Reyes-Terán, Ted M Ross, Patrick C Wilson, Rafi Ahmed, Jesse D Bloom, and Scott E Hensley. Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013-2014 influenza season. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44):15798–15803, November 2014.

[113] Lisa C Lindesmith, Eric F Donaldson, Anna D Lobue, Jennifer L Cannon, Du-Ping Zheng, Jan Vinje, and Ralph S Baric. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS medicine*, 5(2):e31, February 2008.

[114] M Lipsitch. Vaccination against colonizing bacteria with multiple serotypes. *Proceedings of the National Academy of Sciences*, 94(12):6571–6576, June 1997.

[115] M Lipsitch. Bacterial vaccines and serotype replacement: lessons from Haemophilus influenzae and prospects for Streptococcus pneumoniae. *Emerging infectious diseases*, 1999.

[116] Mark Loeb, Margaret L Russell, Lorraine Moss, Kevin Fonseca, Julie Fox, David J D Earn, Fred Aoki, Gregory Horsman, Paul Van Caeseele, Khami Chokani, Mark Vooght, Lorne Babiuk, Richard Webby, and Stephen D Walter. Effect of influenza vaccination of children on infection rates in Hutterite communities: a randomized trial. *JAMA*, 303(10):943–950, March 2010.

[117] Maia Martcheva, Benjamin M Bolker, and Robert D Holt. Vaccine-induced pathogen strain replacement: what are the mechanisms? *Journal of the Royal Society, Interface*, 5(18):3–13, January 2008.

[118] Maia Martcheva, Benjamin M Bolker, and Robert D Holt. Vaccine-induced pathogen strain replacement: what are the mechanisms? *Journal of The Royal Society Interface*, 5(18):3–13, jan 2008.

[119] William J H Mcbride, Walter P Abhayaratna, Ian Barr, Robert Booy, Jonathan Carapetis, Simon Carson, Ferdinandus De Looze, Rod Ellis-Pegler, Leon Heron, Jeff Karrasch, Helen Marshall, Jodie Mcvernon, Terry Nolan, William Rawlinson, Jim Reid, Peter Richmond, Sepehr Shakib, Russell L Basser, Gunter F Hartel, Michael H Lai, Steven Rockman, and Michael E Greenberg. Efficacy of a trivalent influenza vaccine against seasonal strains and against 2009 pandemic H1N1: A randomized, placebo-controlled trial. *Vaccine*, 34(41):4991–4997, sep 2016.

[120] A R McLean. Vaccination, evolution and changes in the efficacy of vaccines: a theoretical framework. *Proceedings. Biological sciences / The Royal Society*, 261(1362):389–393, sep 1995.

[121] H Q McLean, M G Thompson, M E Sundaram, J K Meece, D L McClure, T C Friedrich, and E A Belongia. Impact of Repeated Vaccination on Vaccine Effectiveness Against Influenza A(H3N2) and B During 8 Seasons. *Clinical Infectious Diseases*, 59(10):1375–1385, October 2014.

[122] Huong Q McLean, Mark G Thompson, Maria E Sundaram, Burney A Kieke, Manjusha Gaglani, Kempapura Murthy, Pedro A Piedra, Richard K Zimmerman, Mary Patricia Nowalk, Jonathan M Raviotta, Michael L Jackson, Lisa Jackson, Suzanne E Ohmit, Joshua G Petrie, Arnold S Monto, Jennifer K Meece, Swathi N Thaker, Jessie R

Clippard, Sarah M Spencer, Alicia M Fry, and Edward A Belongia. Influenza vaccine effectiveness in the United States during 2012-2013: variable protection by age and virus type. *Journal of Infectious Diseases*, 211(10):1529–1540, May 2015.

[123] Jan Medlock and Alison P Galvani. Optimizing influenza vaccine distribution. *Science*, 325(5948):1705–1708, September 2009.

[124] J C Miller. Spread of infectious disease through clustered populations. *Journal of The Royal Society Interface*, 6(41):1121–1134, October 2009.

[125] Arnold S Monto, Suzanne E Ohmit, Joshua G Petrie, Emileigh Johnson, Rachel Truscon, Esther Teich, Judy Rotthoff, Matthew Boulton, and John C Victor. Comparative efficacy of inactivated and live attenuated influenza vaccines. *The New England journal of medicine*, 361(13):1260–7, sep 2009.

[126] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, March 2008.

[127] Richard A Neher, Trevor Bedford, Rodney S Daniels, Colin A Russell, and Boris I Shraiman. Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 113(12):E1701–9, March 2016.

[128] OECD. *Education at a Glance 2015, OECD Indicators*, 2015.

[129] S E Ohmit, M G Thompson, J G Petrie, S N Thaker, M L Jackson, E A Belongia, R K Zimmerman, M Gaglani, L Lamerato, S M Spencer, L Jackson, J K Meece, M P Nowalk, J Song, M Zervos, P Y Cheng, C R Rinaldo, L Clipper, D K Shay, P Piedra, and A S Monto. Influenza Vaccine Effectiveness in the 2011-2012 Season: Protection Against Each Circulating Virus and the Effect of Prior Vaccination on Estimates. *Clinical Infectious Diseases*, 58(3):319–327, January 2014.

[130] Suzanne E Ohmit, John C Victor, Judy R Rotthoff, Esther R Teich, Rachel K Truscon, Laura L Baum, Bhavya Rangarajan, Duane W Newton, Matthew L Boulton, and Arnold S Monto. Prevention of antigenically drifted influenza by inactivated and live attenuated vaccines. *The New England Journal of Medicine*, 355(24):2513–2522, December 2006.

[131] Michael T Osterholm, Nicholas S Kelley, Alfred Sommer, and Edward A Belongia. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *The Lancet. Infectious diseases*, 12(1):36–44, January 2012.

[132] Rekha Pai, Matthew R Moore, Tamara Pilishvili, Robert E Gertz, Cynthia G Whitney, Bernard Beall, and Active Bacterial Core Surveillance Team. Postvaccine genetic structure of Streptococcus pneumoniae serotype 19A from children in the United States. *The Journal of Infectious Diseases*, 192(11):1988–1995, December 2005.

[133] Abraham Palache, Valerie Oriol-Mathieu, Atika Abelin, and Tamara Music. Seasonal influenza vaccine dose distribution in 157 countries (2004–2011). *Vaccine*, 32(48):6369–6376, November 2014.

[134] Abraham Palache, Valerie Oriol-Mathieu, Mireli Fino, Margarita Xydia-Charmanta, and Influenza Vaccine Supply task force (IFPMA IVS). Seasonal influenza vaccine dose distribution in 195 countries (2004-2013): Little progress in estimated global vaccination coverage. *Vaccine*, 33(42):5598–5605, October 2015.

[135] Andrew W Park, Janet M Daly, Nicola S Lewis, Derek J Smith, James L N Wood, and Bryan T Grenfell. Quantifying the impact of immune escape on transmission dynamics of influenza. *Science*, 326(5953):726–728, October 2009.

[136] R G Pebody, H K Green, and N Andrews. Uptake and impact of vaccinating school age children against influenza during a season with circulation of drifted influenza A and B strains, England, 2014/15. *Euro surveillance*, 20(39):1560–7917, 2015.

[137] Joshua G Petrie, Suzanne E Ohmit, Caroline K Cheng, Emily T Martin, Ryan E Malosh, Adam S Lauring, Lois E Lamerato, Katherine C Reyes, Brendan Flannery, Jill M Ferdinands, and Arnold S Monto. Influenza Vaccine Effectiveness Against Antigenically Drifted Influenza Higher Than Expected in Hospitalized Adults: 2014-2015. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 63(8):1017–1025, oct 2016.

[138] Tamara Pilishvili, Catherine Lexau, Monica M. Farley, James Hadler, Lee H. Harrison, Nancy M. Bennett, Arthur Reingold, Ann Thomas, William Schaffner, Allen S. Craig, Philip J. Smith, Bernard W. Beall, Cynthia G. Whitney, and Matthew R. Moore. Sustained Reductions in Invasive Pneumococcal Disease in the Era of Conjugate Vaccine. *The Journal of Infectious Diseases*, 201(1):32–41, jan 2010.

[139] Nicola Principi, Susanna Esposito, Paola Marchisio, Roberto Gasparini, and Piero Crovari. Socioeconomic impact of influenza on healthy children and their families. *The Pediatric Infectious Disease Journal*, 22(10 Suppl):S207–10, October 2003.

[140] Andrew Rambaut, Oliver G Pybus, Martha I Nelson, Cecile Viboud, Jeffery K Taubenberger, and Edward C Holmes. The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453(7195):615–619, May 2008.

[141] J M Read, J Lessler, S Riley, S Wang, L J Tan, K O Kwok, Y Guan, C Q Jiang, and D A T Cummings. Social mixing patterns in rural and urban areas of southern China. *Proceedings of the Royal Society B: Biological Sciences*, 281(1785):20140268–20140268, April 2014.

[142] C A Russell, T C Jones, I G Barr, N J Cox, R J Garten, V Gregory, I D Gust, A W Hampson, A J Hay, A C Hurt, J C de Jong, A Kelso, A I Klimov, T Kageyama, N Komadina, A S Lapedes, Y P Lin, A Mosterin, M Obuchi, T Odagiri, A D M E Osterhaus, G F Rimmelzwaan, M W Shaw, E Skepner, K Stohr, M Tashiro, R A M Fouchier, and D J Smith. The Global Circulation of Seasonal Influenza A (H3N2) Viruses. *Science (New York, N.Y.)*, 320(5874):340–346, April 2008.

[143] 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, February 2009.

[144] Jeffrey Shaman and Melvin Kohn. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3243–3248, March 2009.

[145] Jeffrey Shaman, Virginia E Pitzer, Cécile Viboud, Bryan T Grenfell, and Marc Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental united states. *PLoS Biol*, 8(2):e1000316, Feb 2010.

[146] Eunha Shim and Alison P Galvani. Distinguishing vaccine efficacy and effectiveness. *Vaccine*, 30(47):6700–6705, October 2012.

[147] K F Shortridge and C H Stuart-Harris. An influenza epicentre? *Lancet (London, England)*, 2(8302):812–813, October 1982.

[148] Yuelong Shu and John McCauley. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13):30494, mar 2017.

[149] Danuta M Skowronski, Catharine Chambers, Gaston De Serres, Suzana Sabaiduc, Anne-Luise Winter, James A Dickinson, Jonathan B Gubbay, Kevin Fonseca, Steven J Drews, Hugues Charest, Christine Martineau, Mel Krajden, Martin Petric, Nathalie Bastien, Yan Li, and Derek J Smith. Serial vaccination and the antigenic distance hypothesis: effects on influenza vaccine effectiveness during A(H3N2) epidemics in Canada, 2010-11 to 2014-15. *Journal of Infectious Diseases*, February 2017.

[150] Danuta M Skowronski, Catharine Chambers, Suzana Sabaiduc, Gaston De Serres, Anne-Luise Winter, James A Dickinson, Jonathan B Gubbay, Steven J Drews, Christine Martineau, Hugues Charest, Mel Krajden, Nathalie Bastien, and Yan Li. Beyond Antigenic Match: Possible Agent-Host and Immuno-epidemiological Influences on Influenza Vaccine Effectiveness During the 2015–2016 Season in Canada. *The Journal of Infectious Diseases*, 216(12):1487–1500, dec 2017.

[151] Danuta M. Skowronski, Catharine Chambers, Suzana Sabaiduc, Gaston De Serres, Anne-Luise Winter, James A. Dickinson, Mel Krajden, Jonathan B. Gubbay, Steven J. Drews, Christine Martineau, Alireza Eshaghi, Trijntje L. Kwindt, Nathalie Bastien, and Yan Li. A Perfect Storm: Impact of Genomic Variation and Serial Vaccination on

Low Influenza Vaccine Effectiveness During the 2014–2015 Season. *Clinical Infectious Diseases*, 63(1):21–32, jul 2016.

[152] Danuta M. Skowronski, Gaston De Serres, Natasha S. Crowcroft, Naveed Z. Janjua, Nicole Boulianne, Travis S. Hottes, Laura C. Rosella, James A. Dickinson, Rodica Gilca, Pam Sethi, Najwa Ouhoummane, Donald J. Willison, Isabelle Rouleau, Martin Petric, Kevin Fonseca, Steven J. Drews, Anuradha Rebbapragada, Hugues Charest, Marie-Ève Hamelin, Guy Boivin, Jennifer L. Gardy, Yan Li, Trijntje L. Kwindt, David M. Patrick, Robert C. Brunham, and for the Canadian SAVOIR Team. Association between the 2008–09 Seasonal Influenza Vaccine and Pandemic H1N1 Illness during Spring–Summer 2009: Four Observational Studies from Canada. *PLoS Medicine*, 7(4):e1000258, apr 2010.

[153] Danuta M. Skowronski, Gaston De Serres, Jim Dickinson, Martin Petric, Annie Mak, Kevin Fonseca, Trijntje L. Kwindt, Tracy Chan, Nathalie Bastien, Hugues Charest, Yan Li, and Yan Li. ComponentSpecific Effectiveness of Trivalent Influenza Vaccine as Monitored through a Sentinel Surveillance Network in Canada, 2006–2007. *The Journal of Infectious Diseases*, 199(2):168–179, jan 2009.

[154] Danuta M. Skowronski, Marie-Eve Hamelin, Naveed Z. Janjua, Gaston De Serres, Jennifer L. Gardy, Chantal Rhéaume, Xavier Bouhy, and Guy Boivin. Cross-Lineage Influenza B and Heterologous Influenza A Antibody Responses in Vaccinated Mice: Immunologic Interactions and B/Yamagata Dominance. *PLoS ONE*, 7(6):e38929, jun 2012.

[155] Danuta M. Skowronski, Travis S. Hottes, Gaston De Serres, Brian J. Ward, Naveed Z. Janjua, Suzana Sabaiduc, Tracy Chan, and Martin Petric. Influenza B/Victoria Antigen Induces Strong Recall of B/Yamagata But Lower B/Victoria Response in Children Primed With Two Doses of B/Yamagata. *The Pediatric Infectious Disease Journal*, 30(10):833–839, oct 2011.

[156] Danuta M. Skowronski, Naveed Z. Janjua, Gaston De Serres, Suzana Sabaiduc, Alireza Eshaghi, James A. Dickinson, Kevin Fonseca, Anne-Luise Winter, Jonathan B. Gubbay, Mel Krajden, Martin Petric, Hugues Charest, Nathalie Bastien, Trijntje L. Kwindt, Salaheddin M. Mahmud, Paul Van Caeseele, and Yan Li. Low 2012–13 Influenza Vaccine Effectiveness Associated with Mutation in the Egg-Adapted H3N2 Vaccine Strain Not Antigenic Drift in Circulating Viruses. *PLoS ONE*, 9(3):e92153, mar 2014.

[157] Danuta M. Skowronski, Naveed Z. Janjua, Gaston De Serres, Anne-Luise Winter, James A. Dickinson, Jennifer L. Gardy, Jonathan Gubbay, Kevin Fonseca, Hugues Charest, Natasha S. Crowcroft, Monique Douville Fradet, Nathalie Bastien, Yan Li, Mel Krajden, Suzana Sabaiduc, and Martin Petric. A Sentinel Platform to Evaluate Influenza Vaccine Effectiveness and New Variant Circulation, Canada 2010–2011 Season. *Clinical Infectious Diseases*, 55(3):332–342, aug 2012.

[158] Danuta M. Skowronski, Naveed Z. Janjua, Suzana Sabaiduc, Gaston De Serres, Anne-Luise Winter, Jonathan B. Gubbay, James A. Dickinson, Kevin Fonseca, Hugues Charest, Nathalie Bastien, Yan Li, Trijntje L. Kwindt, Salaheddin M. Mahmud, Paul Van Caeseele, Mel Krajden, and Martin Petric. Influenza A/Subtype and B/Lineage Effectiveness Estimates for the 2011–2012 Trivalent Vaccine: Cross-Season and Cross-Lineage Protection With Unchanged Vaccine. *The Journal of Infectious Diseases*, 210(1):126–137, jul 2014.

[159] D J Smith, S Forrest, D H Ackley, and A S Perelson. Variable efficacy of repeated annual influenza vaccination. *Proceedings of the National Academy of Sciences*, 96(24):14001–14006, November 1999.

[160] Derek J Smith, Alan S Lapedes, Jan C de Jong, Theo M Bestebroer, Guus F Rimmelzwaan, Albert D M E Osterhaus, and Ron A M Fouchier. Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305(5682):371–376, July 2004.

[161] 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, oct 2013.

[162] Rahul Subramanian, Andrea L Graham, Bryan T Grenfell, and Nimalan Arinaminpathy. Universal or Specific? A Modeling-Based Comparison of Broad-Spectrum Influenza Vaccines against Conventional, Strain-Matched Vaccines. *PLoS Computational Biology*, 12(12):e1005204–17, December 2016.

[163] S. G. Sullivan, K. S. Carville, M. Chilver, J. E. Fielding, K. A. Grant, H. Kelly, A. Levy, N. P. Stocks, S. S. Tempone, and A. K. Regan. Pooled influenza vaccine effectiveness estimates for Australia, 2012-2014. *Epidemiology and Infection*, 144(11):2317–2328, aug 2016.

[164] John S Tam, Maria Rosario Z Capeding, Lucy Chai See Lum, Tawee Chotpitayasunondh, Zaifang Jiang, Li-Min Huang, Bee Wah Lee, Yuan Qian, Rudiwilai Samakoses, Somsak Lolekha, K Pillai Rajamohanan, S Noel Narayanan, Chellam Kirubakaran, Ruth Rappaport, Ahmad Razmpour, William C Gruber, Bruce D Forrest, and Pan-Asian CAIV-T Pediatric Efficacy Trial Network. Efficacy and safety of a live attenuated, cold-adapted influenza vaccine, trivalent against culture-confirmed influenza in young children in Asia. *The Pediatric Infectious Disease Journal*, 26(7):619–628, July 2007.

[165] Kok Keng Tee, Tommy Tsan-Yuk Lam, Yoke Fun Chan, Jon M Bible, Adeeba Kamarulzaman, C Y William Tong, Yutaka Takebe, and Oliver G Pybus. Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. *Journal of Virology*, 84(7):3339–3350, April 2010.

[166] Pham Quang Thai, Marc Choisy, Tran Nhu Duong, Vu Dinh Thiem, Nguyen Thu Yen, Nguyen Tran Hien, Daniel J Weiss, Maciej F Boni, and Peter Horby. Seasonality of absolute humidity explains seasonality of influenza-like illness in Vietnam. *Epidemics*, 13:65–73, December 2015.

[167] J. J. Treanor, H. K. Talbot, S. E. Ohmit, L. A. Coleman, M. G. Thompson, P.-Y. Cheng, J. G. Petrie, G. Lofthus, J. K. Meece, J. V. Williams, L. Berman, C. Breese Hall, A. S. Monto, M. R. Griffin, E. Belongia, D. K. Shay, and US Flu-VE Network. Effectiveness of Seasonal Influenza Vaccines in the United States During a Season With Circulation of All Three Vaccine Strains. *Clinical Infectious Diseases*, 55(7):951–959, oct 2012.

[168] John J Treanor, Hana El Sahly, James King, Irene Graham, Ruvim Izikson, Robert Kohberger, Peter Patriarca, and Manon Cox. Protective efficacy of a trivalent recombinant hemagglutinin protein vaccine (FluBlok®) against influenza in healthy adults: a randomized, placebo-controlled trial. *Vaccine*, 29(44):7733–9, oct 2011.

[169] James Truscott, Christophe Fraser, Simon Cauchemez, Aronrag Meeyai, Wes Hinsley, Christl A Donnelly, Azra Ghani, and Neil Ferguson. Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *Journal of the Royal Society, Interface / the Royal Society*, 9(67):304–312, February 2012.

[170] Tim K Tsang, Lincoln L H Lau, Simon Cauchemez, and Benjamin J Cowling. Household Transmission of Influenza Virus. *Trends in Microbiology*, pages 1–11, November 2015.

[171] UN. *World Population Prospects: The 2012 Revision*. United Nations, Department of Economic and Social Affairs, Population Division, New York, 2013.

[172] UNSD. *UNSD Demographic Statistics*. United Nations, Department of Economic and Social Affairs, Statistics Division, New York, 2016.

[173] Lori Uscher-Pines, Andrew Mulcahy, Jurgen Maurer, and Katherine Harris. The relationship between influenza vaccination habits and location of vaccination. *PLoS ONE*, 9(12):e114863, 2014.

[174] Marta Valenciano, Esther Kissling, Amparo Larrauri, Baltazar Nunes, Daniela Pitigoi, Joan O'Donnell, Annicka Reuss, Judit Krisztina Horváth, Iwona Paradowska-Stankiewicz, Caterina Rizzo, Alessandra Falchi, Isabelle Daviaud, Mia Brytting, Adam Meijer, Bernard Kaic, Alin Gherasim, Ausenda Machado, Alina Ivanciuc, Lisa Domegan, Brunhilde Schweiger, Annamária Ferenczi, Monika Korczyńska, Antonino Bella, Ana Maria Vilcu, Anne Mosnier, Katherina Zakikhany, Marit de Lange, Sanja Kurečić Filipovićović, Kari Johansen, and Alain Moren. Exploring the effect of previous inactivated influenza vaccination on seasonal influenza vaccine effectiveness against medically attended influenza: Results of the European I-MOVE multicentre test-negative case-control study, 2011/2012-2016/2017. *Influenza and other Respiratory Viruses*, may 2018.

[175] Cecile Viboud, Wladimir J Alonso, and Lone Simonsen. Influenza in tropical regions. *PLoS medicine*, 3(4):e89, April 2006.

[176] Cecile Viboud, Ottar N Bjørnstad, David L Smith, Lone Simonsen, Mark A Miller, and Bryan T Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science (New York, N.Y.)*, 312(5772):447–451, April 2006.

[177] J Wallinga, P Teunis, and M Kretzschmar. Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *American journal of epidemiology*, 164(10):936–944, September 2006.

[178] Taia T. Wang, Gene S. Tan, Rong Hai, Natalie Pica, Erin Petersen, Thomas M. Moran, and Peter Palese. Broadly Protective Monoclonal Antibodies against H3 Influenza Viruses following Sequential Immunization with Different Hemagglutinins. *PLoS Pathogens*, 6(2):e1000796, feb 2010.

[179] R G Webster, W J Bean, O T Gorman, T M Chambers, and Y Kawaoka. Evolution and ecology of influenza A viruses. *Microbiological reviews*, 56(1):152–179, March 1992.

[180] Daniel M Weinberger, Richard Malley, and Marc Lipsitch. Serotype replacement in disease after pneumococcal vaccination Serotype replacement in disease following pneumococcal vaccination: A discussion of the evidence. *The Lancet*, 378(9807):1962–1973, 2011.

[181] Frank Wen, Trevor Bedford, and Sarah Cobey. Explaining the geographical origins of seasonal influenza A (H3N2). *Proceedings of the Royal Society B*, 283(1838):20161312–9, September 2016.

[182] Frank Tian Wen, Anup Malani, and Sarah Cobey. Vaccination and the evolution of seasonal influenza. *bioRxiv*, page 162545, jul 2017.

[183] Derek Weycker, John Edelsberg, M Elizabeth Halloran, Ira M Longini, Azhar Nizam, Vincent Ciuryla, and Gerry Oster. Population-wide benefits of routine vaccination of children against influenza. *Vaccine*, 23(10):1284–1293, January 2005.

[184] WHO. *Influenza Fact Sheet*. World Health Organization, 2014.

[185] WHO. Recommended composition of influenza virus vaccines for use in the 2015 southern hemisphere influenza season. Technical report, World Health Organization, 2014.

[186] WHO. Fluid - a global influenza epidemiological data sharing platform, 2018.

[187] WHO. Flunet, 2018.

[188] 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, feb 2012.

[189] Yuri I Wolf, Cecile Viboud, Edward C Holmes, Eugene V Koonin, and David J Lipman. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biology Direct*, 1(1):34, oct 2006.

[190] Jens Wrammert, Dimitrios Koutsonanos, Gui-Mei Li, Srilatha Edupuganti, Jianhua Sui, Michael Morrissey, Megan McCausland, Ioanna Skountzou, Mady Hornig, W Ian Lipkin, Aneesh Mehta, Behzad Razavi, Carlos Del Rio, Nai-Ying Zheng, Jane-Hwei Lee, Min Huang, Zahida Ali, Kaval Kaur, Sarah Andrews, Rama Rao Amara, Youliang Wang, Suman Ranjan Das, Christopher David O'Donnell, Jon W Yewdell, Kanta Subbarao, Wayne A Marasco, Mark J Mulligan, Richard Compans, Rafi Ahmed, and Patrick C Wilson. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *The Journal of experimental medicine*, 208(1):181–93, jan 2011.

[191] Veronika I. Zarnitsyna, Irina Bulusheva, Andreas Handel, Ira M. Longini, M. Elizabeth Halloran, and Rustom Antia. Intermediate levels of vaccination coverage may minimize seasonal influenza outbreaks. *PLOS ONE*, 13(6):e0199674, jun 2018.

[192] Richard K Zimmerman, Mary Patricia Nowalk, Jessie Chung, Michael L Jackson, Lisa A Jackson, Joshua G Petrie, Arnold S Monto, Huong Q McLean, Edward A Belongia, Manjusha Gaglani, Kempapura Murthy, Alicia M Fry, and Brendan Flannery. 2014–2015 Influenza Vaccine Effectiveness in the United States by Vaccine Type. *Clinical Infectious Diseases*, 63(12):1564–1573, December 2016.

[193] Daniel Zinder, Trevor Bedford, Edward B Baskerville, Robert J Woods, Manojit Roy, and Mercedes Pascual. Seasonality in the migration and establishment of H3N2 Influenza lineages with epidemic growth and decline. *BMC evolutionary biology*, 14(1):272, 2014.

[194] Seth J Zost, Kaela Parkhouse, Megan E Gumina, Kangchon Kim, Sebastian Diaz Perez, Patrick C Wilson, John J Treanor, Andrea J Sant, Sarah Cobey, and Scott E Hensley. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proceedings of the National Academy of Sciences*, 2017.