MODELING POPULATION BIOLOGY OF EMERGING INFECTIONS OF HOSTS WITH HETEROGENEOUS IMMUNE HISTORIES

by

DEVEN V. GOKHALE

(Under the Direction of Pejman Rohani)

ABSTRACT

Mathematical models, describing partially observed Markov processes, in combination with likelihood-based statistical inference, are applied to disease systems at distinct scales of biological organization. Initially, a simulation based case study, demonstrating key concepts and ideas of modeling infectious diseases, is presented. Subsequent chapters provide comprehensive analyses of epidemiological time-series generated at within-host, and between host perspective. For all these studies, hypothesis testing, parameter estimation, and uncertainty and sensitivity analysis is carried to infer probable mechanisms driving dynamics in the observed data. For each of the chapter, examples are provided to use the best fitting mathematical models to explore the unobserved states of the system of interest.

INDEX WORDS: SARS-CoV-2, influenza, mumps, viral infections, pathogen interaction, vaccine failure, within-host dynamics, innate immunity, dynamical systems, statistical inference

MODELING POPULATION BIOLOGY OF EMERGING INFECTIONS OF HOSTS WITH HETEROGENEOUS IMMUNE HISTORIES

by

DEVEN V. GOKHALE

B.Sc., Ramnarain Ruia College, University of Mumbai, 2013
M.Sc., Maharaja Sayajirao University, 2015
M.S., University of Vermont, 2017

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the
Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

MODELING POPULATION BIOLOGY OF EMERGING INFECTIONS OF HOSTS WITH HETEROGENEOUS IMMUNE HISTORIES

by

DEVEN V. GOKHALE

Major Professor:	Pejman Rohani
Committee:	John Drake
	Stephen M. Tompkins
	Christopher C. Whalen
	Aaron A. King

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education
and Dean of the Graduate School
The University of Georgia
August 2022

DEDICATION

I dedicate this work to misery that is this life, my parents who taught me how to suffer it, and my lovely wife Uma who taught me how to live it.

ACKNOWLEDGMENTS

This work has been funded with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272201400004C (NIAID Centers of Excellence for Influenza Research and Surveillance, CEIRS), Contract No.75N93021C00018 (NIAID Centers of Excellence for Influenza Research and Response, CEIRR) and from the National Science Foundation under Grant No. DGE-1545433 (IDEAS training program).

I would like to extend my deepest gratitude towards my lab mates, Dr. Tobias Brett, Dr. Christian Gunning, and Dr. Arash Seidpour for their insightful inputs in helping me in forming this dissertation. iwould also like to acknowledge Dr. Vanessa Ezenwa and Dr. Ford Ballantyne for their guidance during the early development of this work. Finally, I would also like to mention all my colleagues that constitute the wonderful Odum school of ecology community for their support throughout my journey. Thank you.

CONTENTS

Acknowledgments v

List of Figures vii

List of Tables xvii

1 Introduction 1
- 1.1 Anatomy of pathogen transmission . . . 2
- 1.2 Complexities in real world disease systems . . . 2
- 1.3 Model-data interface . . . 5

2 Key concepts for model-data interface: A simulation-based case study 8
- 2.1 Introduction . . . 8
- 2.2 Methods . . . 11
- 2.3 Results . . . 16
- 2.4 Discussion . . . 22

3 Informing antiviral effectiveness for influenza A and SARS-CoV-2 by quantifying within-host immune interactions 27
- 3.1 Introduction . . . 27
- 3.2 Methods . . . 30
- 3.3 Results . . . 45
- 3.4 Discussion . . . 51
- 3.5 Supplementary Results . . . 53

4 Unraveling mechanisms driving anomalous influenza B seasonality in Massachussetts 59
- 4.1 Introduction . . . 59
- 4.2 Methods . . . 61
- 4.3 Results . . . 66

4.4 Discussion . . . 73
4.5 Supplementary Results . . . 75

5 Disentangling the causes of mumps reemergence in the US 79
5.1 Introduction . . . 79
5.2 Methods . . . 81
5.3 Results . . . 88
5.4 Discussion . . . 99
5.5 Supplementary Results . . . 103

6 Conclusion 106

Bibliography 110

Appendices 132

A Supplementary Information Ch.2 132

B Supplementary Information Ch.4 136
B.1 Estimating annual migration rates . . . 136
B.2 Transforming the POLYMOD contact matrix . . . 137
B.3 Reconstructing missing vaccine coverage . . . 139
B.4 Full model of vaccine traits and mumps transmission . . . 142
B.5 Full model of Erlang distributed waning immunity . . . 150
B.6 Derivation of R_0 and R_p . . . 152
B.7 Sensitivity analysis . . . 157
B.8 Differential evolutionary algorithm . . . 159

LIST OF FIGURES

1.1 **Schematic representation of processes across scales.** (A) Demonstration of the hierarchical connections among various scales of biological organization in establishment, and maintenance of infectious pathogens. (B) Depiction of systemic heterogeneities intrinsic to scales represented in (A). 5

2.1 Schematic diagram representing partially Observed Markov Process [104] 10

2.2 **Model schematic** showing flows among constituent compartments 12

2.3 **Simulated System and Data**. Line plots represent a stochastic realization of the population dynamics of individuals in the system who are (A) susceptible, (B) vaccinated, (C) exposed, (D) Infectious, and (E) Recovered. (F) Represents the true number of cases per 10^5 individual. (G) shows the dynamics in the observed number of cases in the system. At strategically chosen set of parameter values, the system was simulated and allowed to reach an equilibrium. The first 100 y were discarded as transients and the remainder 80 y time-series was assumed to be data. 18

2.4 **Model-data agreement**. (A) Qualitative median model fits to the simulated data (0-40 years, purple) using the estimated MLE. Ribbon around the model trajectory represent the 95% prediction intervals. (B) Quantitative, log-linear model-data agreement at the MLE (Purple) assessed for the median trajectory. Error bars around the points represent the observation error distribution generated by simulating 1000 synthetic time series. (C) Log probability density, and (D) conditional log-likelihood evaluated for the best fitting model. 20

2.5 **Out of fit model-data agreement**. (A) Qualitative median model fits to the out of fit simulated data (41-80 years, orange) using the estimated MLE. Ribbons around the median model trajectory represent the 95% prediction intervals. (B) Quantitative, log-linear model-data agreement at the MLE (Orange) assessed for the median trajectory. Error bars around the points represent the observation error distribution generated by simulating 1000 synthetic time series. 22

2.6 **Bifurcation and spectral density analysis**. (A) 2 dimensional bifurcation diagram for the pre-vaccine era representing the dominant period of the model dynamics as a function of amplitude of seasonality (β_1) and basic reproductive number (R_0). (B) Temporal dynamics in the synthetic data (left axis) and normalized spectral density for period of oscillation (right axis). Temporal dynamics in the simulated model (left axis) with 95% predictions intervals (ribbons), and normalized spectral density for the period of oscillation for (C) $1\sigma = 13$d and (D) $1\sigma = 25$d. The first 1000y of model simulations were discarded as transients. The subsequent 40y were considered for the calculation of the dominant period of the median simulated model realization. 23

3.1 **Schematic of the focal hypotheses considered in this study.** All sub-populations involved in the host-virus interactions are colour coded. Arrows indicate promotion and flat-headed arrows indicate inhibition reactions describing cellular interactions. 32

3.2 **Experimental data from ferret infection experiments.** Facets represent within-host viral dynamics of H3N2 influenza (top) and SARS-CoV-2 (bottom) collected from nasal washes of 4 individual ferrets (colors). Viral RNA data is represented using points and solid lines; Viral titer data is represented using triangles and dotdashed lines. Different detection threshold (dt) against virus-data type combinations are represented using solid and dotdashed lines for viral RNA and titer respectively. X-axis represents time in days since ferrets were inoculated. 46

3.3 **Relative fits of within-host models to virus dynamics of H3N2 and SARS-CoV-2.** Models trajectories were fit to time series of virus titer (circles-solid lines) and RNA (triangles-dashed lines). Lines represent simulated median of the log-normal probability model and ribbons represent the 95% prediction intervals for H3N2 (golden) and SARS-CoV-2 (red) respectively. (***) denotes the best fitting model denotes the best fitting model to the observed dynamics ($\Delta AICc = 0$) 47

3.4 **Relationship between effectiveness and initiation time for antiviral enhancing cellular toxicity on the viral dynamics of H3N2 and SARS-CoV-2.** (A) Effectiveness of the antiviral therapy expressed as the reduction in of the area under the viral curve relative to the control region. (B) Simulated viral and anti-viral dynamics for regions (I-V) in (A). Control trajectories (region I, $\eta = 0$) of virus dynamics of H3N2 (golden) and SARS-CoV-2 (red) were generated using the parameter MLEs of best fitting models. For treatment (regions II-V), we assume a daily pulse of unit antiviral drug ($Amax = 1$) with moderate antiviral binding rate ($\beta_A = 100\, n_A^{-1} d^{-1}$) and duration ($\frac{1}{\delta} = 1$ day). 50

3.5 **Experimental data from ferret infection experiments stratified by the sex of the animal.** Facets represent within-host viral dynamics of H3N2 influenza (top) and SARS-CoV-2 (bottom). Viral RNA data is represented using points and solid lines; Viral titer data is represented using triangles and dotdashed lines. Different detection threshold (dt) against virus-data type combinations are represented using solid and dotdashed lines for viral RNA and titer respectively. X-axis represents time in days since ferrets were inoculated. 54

3.6 **Sensitivity analysis with hypothesis specific model parameters.** Partial rank correlation coefficients (PCC, color) were calculated to assess sensitivity of parameters in generating model outcomes. 5000 parameter configurations were generated using a Sobol sampling scheme. Bootstrapped confidence intervals were obtained by uniformly sampling with replacement and generating 150 samples. Statistical significance was ascertained at 5%, Bonferroni corrected, level of sigificance. 55

3.7 **Goodness of fit for virus-data-model combination.** Scatter plots show model-data agreement for viral RNA (empty circles) and viral titers (empty triangles) on a log_{10} scale. Error-bars represent individual level variation among replicates. Fill gradient represents time (in days) to demarcate the observed infection interval. Dotdashed line is the 1:1 reference line. For every model, coeficient of variation (R^2, reported inset) was calulated for each data-type at the medium of the replicate ditribution to quantify the proportion of variation explained by the models. (***) represents the best fitting model based on the AICc comparison. . . . 56

3.8 **Simulation study reflecting the relative action of antiviral effectiveness, initiation time and type on the viral dynamics of H3N2 and SARS-CoV-2 with a pulse dose administrative strategy.** (A, C) Effectiveness of the antiviral therapy expressed as the reduction in of the area under the viral curve relative to the control region. (B, D) Viral and anti-viral dynamics simulated for regions (1-5). Control trajectories, region (1), for the two viruses were generated using the parameter MLEs of two best fitting models and simulating dynamics for 12 days. For the remainder of regions, a moderate antiviral binding rate ($\beta_A = 100\ n_A^{-1}d^{-1}$) and duration ($\frac{1}{\delta} = 1$ day) were assumed during this study. 58

4.1 **Relative fits between epidemiological models for seasons 2016-2018.** Plots show matched type specific simulated trajectories - neutral (dashed) and cross-protection (dotted) models fitted to weekly case report data (solid) for MA. Ribbons represent Poisson observation error around model predictions. 68

4.2 **Model fit for 2019 Influenza season.** Plots show matched type specific simulated trajectories for the neutral model (dashed) fitted to weekly case report data (solid) for MA. Ribbons represent Poisson observation error around model predictions. 69

4.3 **Simulation study reflecting dynamic effects of cross-protection ($\chi_{AB} = \chi_{BA} = \chi$) and relative changes in type specific R_0s. A)** Depicts the difference in the epidemic phases of Types A and B (top) and the peaks ratios (bottom) of the two influenza types with relative changes of R_0^B with respect to R_0^A (x-axis). **B)** displays epidemic dynamics (cases per 100,000) for the two types resulting from parameter values selected at points (a), (b), (c) and (d) in **A**. R_0^A is fixed at 2, duration of cross protection is fixed at 1 month. Type-specific immunity is assumed to last 4 years for both types. See *SI Appendix*, Fig. 4.6 for results assuming that Type B-specific immunity lasts longer than Type A-specific immunity. 71

4.4 **Simulation-based illustration of the susceptible backlog hypothesis. A)** Simulation experiments demonstrating susceptible dynamics (dotted lines) and the corresponding relative timing and amplitude of influenza A (solid red line) and influenza B (solid blue lines). For influenza B, we depict three distinct scenarios: R_0^B is either low throughout (= 1.56), high throughout (= 2.41), or R_0^B starts low (= 1.56), but increases (= 2.41) at the start of the 2019 season (highlighted in orange). The associated effective reproductive numbers ($R_{\text{eff}} = R_0 \times \frac{S}{N}$) are presented in panel **C)**. Panels **B)** & **D)** present similar information to **A)** & **C)** but perform an alternative experiment, testing whether the absence of an influenza B outbreak in the 2018/2019 season highlighted in gray and resulting accumulation of susceptible individuals alone would explain the anomalous dynamics in influenza season 2019. Parameter values are presented in Table 4.2 and Table 4.5). 72

4.5 **Two-type influenza transmission model schematic.** Definition of rate parameters and state variables are provided in Tables ?? S1 respectively. 75

4.6 **Simulation study demonstrating dynamic effects of cross-protection and changes in type-specific R_0s when the duration of natural immunity differs between influenza types.** Type-specific duration of natural-immunity for Type A and Type B was set to be 4 years and 10 years respectively. **A)** Shows changes in the peak case ratio (top) and phase-difference in the peak weeks (bottom) of the two influenza types with relative changes of R_0^B with respect to R_0^A (x-axis) and cross-protection (χ). **B)** displays epidemic-dynamics (cases per 100,000) of the two influenza types using parameter values at points (a), (b), (c), and (d) in **A**. R_0^A is fixed at 2 and the duration of cross-protection ($\frac{1}{\phi}$) is set to 1 month. χ is assumed to be symmetric across the two influenza types. 77

5.1 **Mumps distribution in the United States.** (A) Total reported mumps cases per 10^5. (B) Age-stratified case reports across 5 age cohorts (Purple gradient). (C) Age-distribution of mumps case reports (Purple gradient);(D–E) Spatial distribution of average mumps case reports per 10^5 during the two reemergence eras, 1985-1989 (panel D) and 2006-2012 (panel E). 89

5.2 **Model-data agreement.** (A) Age-specific qualitative model fits to within sample time-series (1977-2012, seagreen) and the corresponding 5 year out of sample prediction (2013-2018, orange) for the waning hypothesis. Ribbons represent 80% prediction intervals for the two prediction epochs. (B) Age-specific log-linear model agreement for within sample fit (seagreen lines and hollow circles) and out of sample predictions (orange lines and hollow circles). Statistical performance of the model was assessed using coefficient of variation (R^2, reported inset) for the two prediction epochs. Year of record is represented using a continuous grey gradient. 92

5.3 **Quantitative comparison of mumps incidence age distribution.** (A) Qualitative comparsion between expected, assuming the model of waning immunity, and the observed incidence per 10^5 age-distribution (purple gradient, left, y-axis). Simulated incidence is broken down into the two prediction epochs - within sample (seagreen gradient) and out of sample (orange gradient) as in figure 5.2. Lines represent comparison of temporal shifts in the mean age of infection (right, y-axis) under the waning model (dotted) and observed incidence (dashed). (B) Quantitative agreement between observed *v/s.* expected age-distribution of incidence using the Kullback-Leibler Divergence ($\overline{D}^{KL}$, left y-axis) through time. Boxplots represent bootsrapped distribution of $\overline{D}^{KL}$ calculated by comparing observed age-distribution to 1000 synthetically generated time-series under the estimated observation noise. The area plot represents an estimate of age-aggregated mumps incidence per 10^5 (right, y-axis). 94

5.4 **Reconstructed population immune profile and impact on mumps resurgence.** (A) Proportion of vaccinated individuals retaining immunity over time for the exponential waning. The average duration of immunity was fixed to $1/\delta$ =111.5 years (i.e. the maximum likelihood estimate). The green dashed line indicates the approximate critical vaccination level for the exponential waning model. Here the critical vaccination threshold was approximated using $1-1/R_0$. (B) Percent of the population with vaccine immune protection by age cohorts over time. (C) Dynamics in the effective reproductive number highlighting epidemic transitions between supercritical (red) and subcritical (green) phases (left, y-axis). Annual vaccination rates of neonatal (light green) and booster (dark blue) doses (right, y-axis). (D) Time series of age-specific susceptibility profiles.(E) Number of individuals who lose immunity each week per 10^5. (F) Weekly mumps prevalence per 10^5 individuals in each age cohort ($10^5 * I_a/N_a$) 97

5.5 **Age-specific immunity post vaccination** (A) Relative prevalence ratio $\left(I_i^w(t)/I_i^s(t)\right)$ of infection after vaccination across the five age-cohorts (facet columns), as a function of varying probability of immunity loss by the age of 18 years (colour gradient). Estimated MLE of immune duration was converted to probability of immune loss (indigo); Time underwent a reset to 0 years after administration of neonatal dose (grey background) and booster dose (black background). (B) Population level vaccine impact (solid lines) and reproductive number (dashed lines) calculated at 91.7% and 91.6% neonatal and booster vaccine coverage respectively (fixed at an average for 2008-2018) as a response, and (C) Stable-state prevalence distribution across five age-cohorts as a function of varying probability of immune loss by age 18. Relationship between duration of immunity and Probability of immune loss by age 18 are represented on the secondary axis (dotdashed lines). Covariate values were fixed at the last known value in the year 2018. Dynamics were simulated for 300 y and final values in the infectious compartments values we taken to be prevalence per 10^5 100

5.6 **Relative Model Fit.** Qualitative agreement of age-specific observed cases of mumps (facet rows, black) to 4 models of vaccine imperfection (facet columns) - No Loss (violet), Waning (Exponential, seagreen), Waning (Erlang, x = 3, darkblue), and Leaky (pink). Colored solid lines represent negative log probability density for the four models at the MLE. Ribbons represent 80% prediction intervals from the model simulation. Relative quantitative model-data agreement has been conveyed using Akaike information criterion (ΔAIC, reported inset) among various models. . . 103

5.7 **Relative distribution of average mumps cases.** Barplots represent the mean age-distribution of cases during 2000-2018, calculated using the median trajectories for the 4 models of vaccine imperfection – No Loss (violet), Waning (Exponential, seagreen), Waning (Erlang, x = 3, darkblue), and Leaky (pink). Solid lines represent observed case distribution over the same time period. 104

5.8 **Age-specific parameter estimates of Process and Observation model.** Faceted barplots represent maximum likelihood estimates of probability of infection (q), reporting probability (ρ) and dispersion parameter (ψ) corresponding to the four hypotheses of mumps reemergence (columns). Age-class has been reported using a blue gradient. 105

B.1 **Matrix of daily contact rates.** Daily diary records of who (y-axis) comes in contact with whom (x-axis) for the UK in 2005. Contact rates were caliberated for the US using age-specific national population sizes and then corrected for reciprocity. 138

B.2 **Reconstructed neonatal vaccine rate.** (A) Time series for the available mumps annual vaccine coverage rates for the neonatal dose. The blue region highlights the period of missing vaccine data. (B) Weekly mumps incidence during the period where vaccine data is missing. (C) Imputed, weekly vaccine uptake (upper panel) by recursively filtering from 1000 stochastically generated trajectories conditioned on the initial decline of mumps in (B). Filtering distribution is assumed to be an over-dispersed normal ($\psi = 0.8$, $\rho = 0.04$). Averaged annual coverage (lower panel) is used in the testing of hypothesis. 141
B.3 **Model covariates.** (A-F) Demographic and epidemiological variables used in introducing realizm to the simulated dynamics; (A) Probability that for a given case age was reported; (B) Total normalized annual births; Annual mumps vaccination rates for the neonatal dose (C) and booster dose (D), Missing early uptake values were interpolated (blue); (E) Age-specific population size; And (F) migration rates broken down by age cohorts. 143
B.4 **Transmission model schematic diagram** (A) Basic skeleton of model with vaccine imperfection. (B) Erlang distributed waning distribution. (C) Demographic model to account for vaccination, and aging. 144
B.5 Pseudo code for the DE implementation 160

LIST OF TABLES

2.1 Model parameters, ranges, and definitions. 11
2.2 Model parameter comparison 19

3.1 **Model compartments for the model of infection and immunity.** 33
3.2 Model parameters with definition and units for the full model of infection and immunity. Parameters are grouped by the hypotheses and arranged in an order of increasing biological complexity. 34
3.3 Model parameters with definition for the joint observation model for viral titers and RNA. 40
3.4 **Model compartments for the model of antiviral action.** . 43
3.5 Model parameters with definition and units for the full model of infection and immunity. Parameters are grouped by the hypotheses tested in this article and arranged in the increasing order of biological complexity. Individual model formulations are specified in Eq. 3.42-3.49 44
3.6 **Maximum likelihood estimates (MLE)**. Model specific parameter MLE's were obtained by maximizing the likelihood function. 49
3.7 **Maximum likelihood estimates (MLE) and 95% confidence intervals**. Model specific parameter MLE's were obtained by maximizing the likelihood function. Uncertainty in the MLE was ascertained using parametric bootstrap. 57

4.1 Parameters for Strategic Model 63
4.2 Parameter estimates and goodness of fit. 67
4.3 Model State Variables 75

4.4 Parameter estimates and goodness of fit (2016-2018 seasons). Cross-protection and neutral models were formally contested to assess relative goodness of fit using ΔAIC values. Models were integrated until equilibrium (burn-in period of 100 years) and subsequent 3 years were used in the fitting procedure. Simulated trajecteories were fit to weekly, type-specific, incidence data for influenza for the state of Massachusetts. 76
4.5 Parameter estimates and goodness of fit (2019 season). The best fitting model (neutral) from the previous era, Table 4.4, was used to estimate type specific epidemiological parameters. The model was integrated to reach equilibrium (burn-in period of 103 years) using parameter MLE from the previous era. Eventual sizes of model compartments from this simulation were fixed as initial conditions for 2019 season. Model parameters not estimated above were fixed at the MLE from the previous era. Simulated trajectories were fit to weekly, type-specific, incidence data for Influenza for the state of Massachusetts. 76
4.6 Parameter estimates and goodness of fit. The neutral model was fit to seasons 2016-2019 to assess relative goodness of fit against the best fitting model (neutral with season dependent parameter values, Tables 4.4 and 4.5) using ΔAIC. Simulated trajectories were fit to weekly, type-specific, incidence data for influenza for the state of Massachusetts. . 78

5.1 Table of parameter maximum likelihood estimates and derived quantities 91
5.2 95% Bootstraped Confidence Intervals 104

B.1 Model parameters with ranges were estimated by fitting model to a time series of Annual Mumps Incidence data for the US (1977-2018) 145
B.2 Model Compartments for the i^{th} age-cohort 146
B.3 Model Compartments for the i^{th} age-cohort 157

CHAPTER 1

INTRODUCTION

A model is a caricature or analogy of the perceived reality that contains the minimum, necessary elements to reproduce the essential attributes of this reality *e.g.*, a miniature toy plane is a model consists of elements about a real plane's appearance, mannequins that stand gracefully poised at the local mall are meant to be a representation of our appearance when adorning the $200 jeans that the model advertises, or a thumbnail of a well-lit apartment posted on a travel website is meant to inform us of the daily abundance of sunlight our brief sojourn in Bruges! Similarly, a mathematical model is a theoretical analogy, formally defined using a collection of postulates, usually derived from first principles, that are believed to govern an observed system, physical, chemical, biological or otherwise. Mathematical models serve to enhance our understanding of such a system of interest without the need of perfect reconstruction of the observed reality. However, as per the popular aphorism stated by George Box, *"All models are wrong, but some are useful"*, the utility of a mathematical model needs to be justified. Heuristically, this is carried out through a statistical confrontation of models with observed data, collected as a representational subset of the system of interest. This dissertation offers a compendium of applications of interfacing mathematical models with observations in exploring the complex biology surrounding the spread of infectious pathogens spreading.

1.1 Anatomy of pathogen transmission

What are the first principles that govern disease transmission? Interestingly, when it comes to successful proliferation of a pathogen, only three basic prerequisites exist: (1) the existence of a biologically naïve constituent of the population, also called the susceptibles (denoted by S), (2) its transmission capable counterpart, or the infectious (I), and (3) establishment of infection given contact between the two (at rate β) [97]. These components can be mathematically represented as

$$\frac{dI}{dt} = \beta SI. \tag{1.1}$$

Here Eq.1.1 describes the growth rate of the infectious population, also the measure of the spread of a pathogen, for a defined population. This representation, offers the most parsimonious description of disease transmission. Here it is assumed that constituents of a population, distinguished merely by their discrete infection status (here S and I), encounter each other at random, like gas molecules enclosed in a container. In many cases, depending upon the organization of the population, the assumption about homogeneous mixing can hold, making this a very useful model to describe the mechanism of disease spread. However, there are conditions when this simple model may fall short to account for the structural complexities. I outline some of these below.

1.2 Complexities in real world disease systems

The principle of parsimony, also referred to as the Occam's razor, succinctly provides an abstract threshold for simplicity. It states that *"Everything should be made as simple as possible, but never simpler"*. This means there are conditions where the parsimonious assumptions stated in Eq.1.1 may fall short to account for the structural complexities that might be crucial to to the description of the system.

To provide an example of such a violation, consider the assumption of "random contacts" described in Eq.1.1. In the real world, contacts are nearly never random. Most of the time, individuals mixing within a population will do so assortatively commonly reflecting a certain social behavior *e.g.*infants come in contact with their parents at a rate higher than they do with other infants [135]; Individuals within polygamous groups have disproportionately more sexual contacts on average than individuals engaged in monogamy [11], some professions like human health care typically require far more human engagement than other occupations, which may conversely require extended periods of isolation indoors [41]. When considering pathogen transmission occurring among multiple geographically or socially organized populations, *e.g.* cities, countries, or continents, some populations might be better connected to one another than others. This may be by the virtue of geographic proximity or socio-economic and political interactions like tourism, trade or war [40, 67, 172]. Even from a within-host perspective, some chronic infections can persist for a long duration of time because of the refugia generated by the adjacent organs during the various stages of the disease progression [130]. In short, contact structures among individual entities fueling the transmission process can be quite complex, depending upon the disease system and the scale of biological organization.

Mixing patterns are not the only source of heterogeneity within a system of interest. Eq.1.1 treats population constituents to be stratified into discrete, binary states depending upon their infection status. In reality, the state of a given individual lies along a complex continuum between susceptibility and infectiousness. Such a heterogeneous distribution of susceptibility and infectivity, can therefore introduce differentials in shedding and acquisition rate of the pathogenic particles circulating among the individual constituents of the system. In case of the viral infections like influenza, infected epithelial cells can exhibit extreme heterogeneity in their potential as viral producers within human host [175]. At the scale of diseases spreading between individuals, many disease systems pertaining to proximity driven, vector-borne, or venereal transmission routes show extreme aggregation in the number of secondary cases produced by individuals in the population [119, 209]. Even in case of a meta-population connected over

a social network, defined by age, occupation, or behavior, risk differentials to exposure can drive some populations to become superior transmission sources while others might behave as effective sinks [96].

Yet another complexity in ascertaining the transmission dynamics is that many of the factors influencing the spread of the pathogen are inhomogenous over time. This could be a consequence of temporal variation driven by fluctuating environmental conditions [3] or demographic characteristics of hosts like births [83], deaths, geographical movement [17, 127] or physiological and genetic characteristics that modulate susceptibility or resistance of the host to invading pathogens thereby generating seasonal variation in the population level immunity [49].

Moreover, the spread of an infectious agent through a population is inherently a multi-scale phenomenon [200]. Specifically, the process of infection of an obligate pathogen like a virus starts at the cellular level. Viruses are capable of hijacking the cellular machinery, converting a cell into a virus producing factory [176]. Upon successful within-host establishment and proliferation of the infecting pathogen, the infected hosts become viable to transmit among individuals within a population. With continued and persistent transmission, the proliferating pathogen can gradually spread between multiple populations spanning socio-geographical scales of cities, countries and, in many cases, even continents as is apparent with the ongoing pandemic of the severe acute respiratory syndrome corona virus 2 (Fig. 1.1A).

Under these circumstances, depending upon the problem at hand, one must incorporate the right level of complexity within a model in order to accurately capture the salient characteristics of the system of interest. In practice, not considering the intrinsic structural heterogeneities while formulating a transmission model can lead to misleading results, in many cases underestimating the true transmission potential of the proliferating pathogen [96]. Deciding what is the appropriate degree of complexity can therefore be very complicated, making inference and ecological modeling of infectious diseases challenging [19].

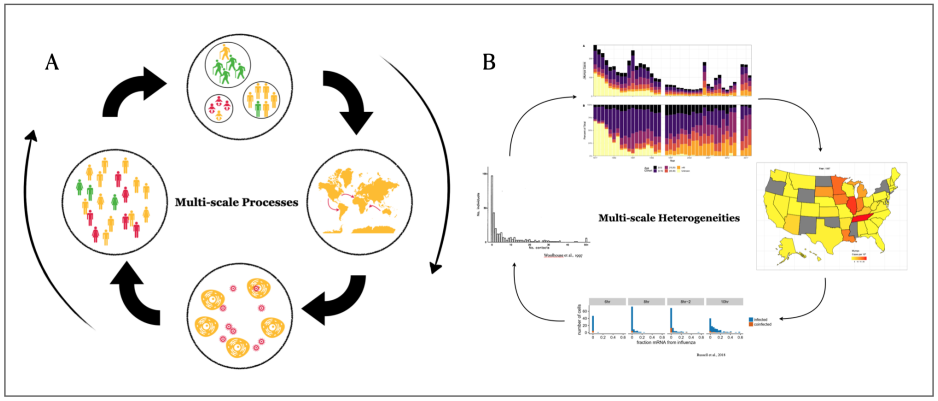

Figure 1.1: **Schematic representation of processes across scales.** (A) Demonstration of the hierarchical connections among various scales of biological organization in establishment, and maintenance of infectious pathogens. (B) Depiction of systemic heterogeneities intrinsic to scales represented in (A).

1.3 Model-data interface

Mathematical, and statistical models are greatly flexible, and widely applied tools in analysing complex processes underpinning the spread of infectious agents. This ubiquitous employment of transmission models in the field of disease ecology stems from two crucial factors. (I) An arbitrary degree of complexity can be incorporated within a dynamical process model [7, 96]. This ensures the capacity of incorporating the appropriate degree of realism and complexity necessary to portray the biology that is under investigation. (II) In nearly all cases, there is contention about the underlying mechanism that leads to the observed dynamics, often this pertains to a time series of case incidence of a certain pathogen circulating through a population. Under such circumstances, not only is the parametric configuration of a model not known, there may be uncertainty about the structure of the underlying model. Formulating multiple plausible models to test which can reproduce the observed transmission patterns can very easily be formulated under this paradigm [167].

Mathematical models in combination with likelihood-based statistical analysis can be employed to conduct hypothesis testing and model selection to

identify, from a collection of hypothesized models, the best fitting explanations to the observed data. [83, 186]. Further, dynamical models have shown to be useful to conduct counterfactual scenario analysis. This allows us to understand how a system will evolve under several conflicting scenarios [59, 204]. Mathematical models can thus be instructive in design and evaluation of various public health strategies.

This dissertation broadly emphasizes the role of mathematical and statistical modeling in uncovering mechanisms that drive dynamics in epidemiological and virological time-series data. Specifically, I demonstrate the application of likelihood-based frequentist methods to arrive at a singular mechanism, among multiple competing ecological hypotheses, that most likely drives the dynamics in the observed data [53]. This method can further be used in the estimation of the unknown parameters of the system of interest, assessing the sensitivity of the system to the changes in the values key model parameters, quantifying uncertainty in parameter estimates and assessing validity of our best fitting model.

In the first chapter, I provide a demonstration of the key workflow involved in model formulation, fitting and estimation of parameters using a simulated noisy time series of a hypothetical disease system. This chapter presents some of the key steps that constitute time-series analysis of epidemiological data using mechanistic models of infectious disease transmission. Subsequent chapters, demonstrate the applications of theoretical techniques to conduct inference and statistical analysis of infectious disease time-series observed at distinct scales of biological organization.

In the second chapter, using the infection data collected using challenge experiments on ferrets, we attempt to quantify the modulatory nature of the host's innate immune system in controlling cellular level dynamics of influenza A H3N2 (H3N2) and Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) viruses. Since their introduction in 1968 and 2019, H3N2 and, SARS-CoV-2 respectively, have caused unprecedented disruption of the public health systems globally. In addition to the substantial burden of morbidity and mortality around the world, both viruses exhibit rapid evolution leading to antigenic escape from the prevailing interventions [26, 74, 158, 181]. In this bi-phasic study, we develop

a suite of within-host models encompassing a number of hypotheses regarding virus-specific functional responses of the host's innate immune system and their impacts on the proliferation of H3N2 and SARS-CoV-2 viruses. Antiviral therapies are among the most effective pharmaceutical interventions in treatment of a variety of viral pathogens. Having identified the biological mechanism driving the infections of H3N2 and SARS-CoV-2 respectively, we analyze the effectiveness of a hypothetical antiviral in rapidly controlling the two infections.

In the third chapter, we assess the role of facilitative interaction among pathogens, at an individual level, in explaining the atypical in 2019/2020 influenza season in the United States. This recent influenza season began earlier than any season since the 2009 H1N1 pandemic [44], with an increase in influenza-like illnesses observed as early as August [148]. Also noteworthy was the numerical domination of influenza B cases early in this influenza season, in contrast to their typically later peak in the past [24]. In this chapter, I provide a case study epidemiological transmission models of multiple pathogens fitted to weekly type-specific influenza time-series from Massachusetts. This study investigates whether changes in the effective reproduction number or short-term cross-immunity between these viruses can explain the novel dynamics of influenza A and B seasonality.

In the fourth chapter, we present an analysis of the reemergent epidemics of mumps at the using an age-structured transmission model. Over the past two decades, multiple countries with high vaccine coverage have experienced resurgent outbreaks of mumps [14, 75, 92, 125, 151, 157]. Worryingly, in these countries a high proportion of cases have been among those who have completed the recommended vaccination schedule, raising alarm about the effectiveness of existing vaccines. Two putative mechanisms of vaccine failure have been proposed as driving observed trends: (1) gradual waning of vaccine-derived immunity [115, 208] (necessitating additional booster doses), and (2) the introduction of novel viral genotypes capable of evading vaccinal immunity [45]. Focusing on the United States, we conduct statistical likelihood-based hypothesis testing using a mechanistic transmission model on age-structured epidemiological, demographic and vaccine uptake time-series data.

CHAPTER 2

KEY CONCEPTS FOR MODEL-DATA INTERFACE: A SIMULATION-BASED CASE STUDY

2.1 Introduction

Depending upon the choice of a model, a computational scientist is often confronted with a trade-off between interpretability of results and representation of complexity contained within the model. A simpler model may offer a clear mapping between the values of the control parameters, and a corresponding effect on the model dynamics. Such a model, however, may fail to capture some of the complex mechanistic details of the phenomenon of interest. Thus a parsimonious model, may be a gross misrepresentation of the reality that a researcher might be interested in uncovering [96]. On the other hand, a model encompassing a higher degree of structural complexity can greatly enhance the realism being captured thus improving the generalizability, and practicability of the model. However, complex models which are a very close representation of the biology of interest can often contain parameter trade-offs, such that multiple parametric combinations may lead to the production of qualitatively similar dynamics, there by

obfuscating the effect of individuals model parameters. This can greatly limit our ability to identify unknown parameters about the system [171], and in many cases, may make interpretability of results very difficult [182].

Throughout this chapter, and this dissertation as whole, the use of the word model refers to a specific type of mathematical model often used in describing partially observed or latent state Markov processes. The two defining characteristics that this suite of models possess are – (1) Only a small subset of the structural complexity generating the biological phenomenon is in fact recorded or observed, where as the majority of the system remains hidden [91], and in most instances cannot be observed, (2) There is interplay among deterministic, and stochastic forcing generating time-series that resemble a noisy clockwork [19].

Dynamics generated by such models are often represented using a system of ordinary or stochastic differential equations. This approach facilitates a researcher in formally incorporating multiple mechanistic explanations, often derived through first principles. Often these hypothesized mechanisms are believed to be generative of the observed data. Figure 2.1 provides a conceptual link between the theoretical model and the observed data. Here a process, consisting of partially hidden states, often characterizing the crucial components of the biological system of interest, is mapped to a recorded time-series through the lens of an observation model. Often the observation model is a probability distribution which and the records are as a random variable following this distribution.

Another special property of these type of a models is that all states (X_t) depend only on 2 components - (1) the control parameters of the model (θ) and the immediate past configuration of the constituent states (X_{t-1}). This dependency of the state variable(s) on their immediate past values is called the Markovian or the memory-less property of theses models [64].

A typical analysis of a partially observed Markov process, such as infectious disease transmission, often begins by considering multiple biological hypotheses. The subsequent step is to find, among these hypothesis, the most probable explanation driving the dynamics in the observed data. There are a plethora of available techniques however this dissertation em-

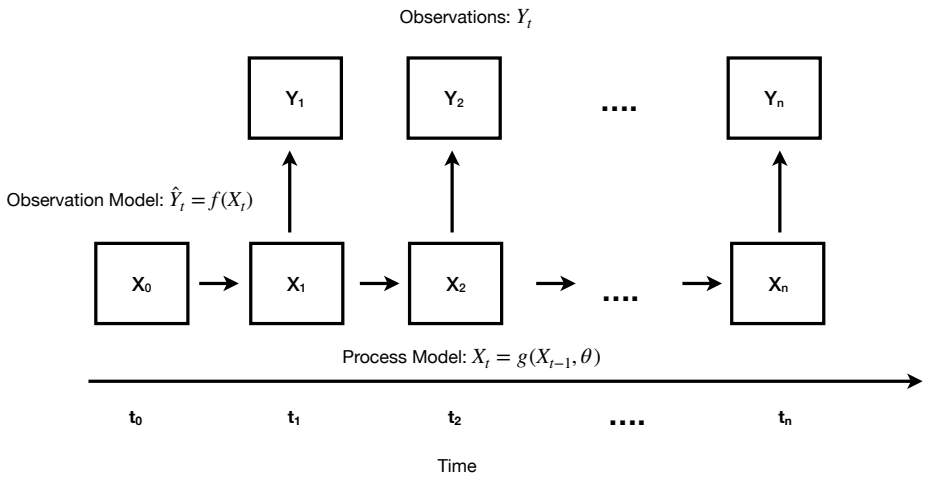

Figure 2.1: Schematic diagram representing partially Observed Markov Process [104]

ploys likelihood based statistical inference to perform model selection, and estimation of the unknown parameters [53, 155].

Following model selection, one ascertains model validity by calculating qualitative, and quantitative measures of goodness of fit pertaining to one or models of interests. Such diagnostic procedures allow quantification of the explanatory power possessed by the hypothesized model(s) of interest and provide useful insights about how these competing hypotheses qualitatively offer differential explanations about the observed reality. During this step, one is also interested in understanding how well were the parameters of interest, identified for the given dataset. Finally, the best fitting model can be employed to make predictions, and forecasts about the unobserved states.

In this study I simulate a synthetic time-series from a simple Vaccinated (V), Susceptible (S), Exposed (E), Infectious (I), Recovered (R) compartmental model with demographic stochasticity seasonality in transmission, and a time-inhomogenous rate of vaccination. The primary focus of this chapter is to lay-out some of the key steps involved in modeling infectious diseases. For this known biological scenario, I attempt to recover the parameter estimates that were used in the generation of data. Further I assess

the goodness of fit of our best fitting model. Finally, I perform a out of fit prediction using the best fitting model which is often considered as the gold standard for assessing the validity of the model of interest.

2.2 Methods

Data Simulation. The disease incidence time series data in this study was generated using a Euler-Maruyama representation of a system of stochastic differential equations (SDE) [30, 36, 68, 105]. For the dynamical system, a population of size N into Susceptible (S), Exposed (E), Infectious (I), Recovered (R), and Vaccinated (V) discrete states. Within this population, individuals received a neonatal dose of vaccination whence they enter into the V compartment at the rate $p(t)\omega$. The remainder of the births, entered into the S compartment at rate $\big(1-p(t)\big)\omega$ and remained susceptible to disease transmission. Upon infection, individuals move to the latent class (E). After a transient period of incubation lasting for on average $1/\sigma$ days, individuals leave the E compartment and become viably contagious as they enter the I class. The infection is cleared within $1/\gamma$ days on average. All model parameters and their corresponding values used in the simulation of time-series are described in Table 2.1.

Table 2.1: Model parameters, ranges, and definitions.

Parameter	True values	Definition
N	10^8	Population size
R_0	10	Basic reproductive number
$1/\gamma$	5 d	Duration of Infection
$1/\sigma$	13 d	Duration of Latency
β_1	0.1	Amplitude of Seasonality
ι	1 y^{-1}	Transient importation Rate
$1/\omega$	80 y	Average life expectancy
$1/\delta$	110 y	Vaccine derived immune duration
ρ	0.005	Probability of reporting
$1/\psi$	0.05	Dispersion Parameter

Force of infection is proportional to the number of individuals in the I compartment and the seasonality in the transmission,

$$\lambda(t) = R_0\gamma(I/N + \iota)\big(1 + \beta_1 sin(2\pi t/T)\big) \tag{2.1}$$

Here R_0 is the basic reproductive number which is defined as the number of secondary infections generated by a index case when introduced in a completely susceptible population. ι, β_1, and T are the average annual rate of transient importations, the amplitude, and the period (assumed to be 52 weeks) of seasonality respectively.

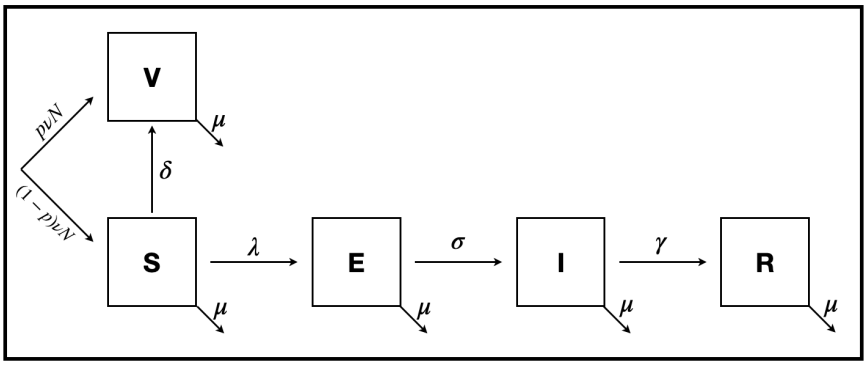

Figure 2.2: **Model schematic** showing flows among constituent compartments

Unlike a deterministic model, where individuals are assumed to flow from one compartment to the other at a fixed rate, a stochastic system is defined in terms of probable events that can occur with a fixed probability distribution for every compartment in the system. Thus a model represented by a flow diagram like Fig. 2.2, often used to describe deterministic dynamics, can be discretized such that every rate reaction is replaced by its corresponding probability counterpart. The changes in the population within individuals compartments are now written as $\Delta N_{..} = N_{..}(t+\Delta t) - N_{..}(t)$. This accounts for the change in the number individuals within a compartment as they experience a certain event within a small interval of time $[t, t+\Delta t)$. Specifically, for the system described by Fig. 2.2, individuals belonging to a constituent compartment can either experience birth ($\Delta N_{.,x}$), death ($\Delta N_{x,.}$), or they can transition from one compartment to the other ($\Delta N_{x,y}$) for all $x, y \in \{V, S, E, I, R\}$. The continuous time Markov-process for small increments in probability for this system is then given as

$$P\big(\Delta N_{.,S}(t)=1|S(t),I(t),R(t)\big)=\omega\big(1-p(t)\big)N\Delta t+o(\Delta t), \tag{2.2}$$
$$P\big(\Delta N_{S,.}(t)=1|S(t),I(t),R(t)\big)=\omega S(t)\Delta t+o(\Delta t), \tag{2.3}$$
$$P\big(\Delta N_{S,E}(t)=1|S(t),I(t),R(t)\big)=\lambda(t)S(t)\Delta t+o(\Delta t), \tag{2.4}$$

$$P\big(\Delta N_{.,V}(t)=1|S(t),I(t),R(t)\big)=\omega p(t)N\Delta t+o(\Delta t), \tag{2.5}$$
$$P\big(\Delta N_{V,.}(t)=1|S(t),I(t),R(t)\big)=\omega V(t)\Delta t+o(\Delta t), \tag{2.6}$$
$$P\big(\Delta N_{V,S}(t)=1|S(t),I(t),R(t)\big)=\delta(t)V(t)\Delta t+o(\Delta t), \tag{2.7}$$

$$P\big(\Delta N_{E,.}(t)=1|S(t),I(t),R(t)\big)=\omega E(t)\Delta t+o(\Delta t), \tag{2.8}$$
$$P\big(\Delta N_{E,I}(t)=1|S(t),I(t),R(t)\big)=\sigma E(t)\Delta t+o(\Delta t), \tag{2.9}$$

$$P\big(\Delta N_{I,.}(t)=1|S(t),I(t),R(t)\big)=\omega I(t)\Delta t+o(\Delta t), \tag{2.10}$$
$$P\big(\Delta N_{I,R}(t)=1|S(t),I(t),R(t)\big)=\gamma P(t)\Delta t+o(\Delta t), \tag{2.11}$$

$$P\big(\Delta N_{R,.}(t)=1|S(t),I(t),R(t)\big)=\omega R(t)\Delta t+o(\Delta t). \tag{2.12}$$

The corresponding changes in the state variables can then be given as

$$\Delta V(t)=\Delta N_{.,V}(t)-\Delta N_{V,S}(t)-\Delta N_{V,.}(t), \tag{2.13}$$
$$\Delta S(t)=\Delta N_{.,S}(t)-\Delta N_{S,I}(t)-\Delta N_{S,.}(t), \tag{2.14}$$
$$\Delta E(t)=\Delta N_{S,E}(t)-\Delta N_{E,I}(t)-\Delta N_{E,.}(t), \tag{2.15}$$
$$\Delta I(t)=\Delta N_{E,I}(t)-\Delta N_{I,R}(t)-\Delta N_{I,.}(t), \tag{2.16}$$
$$\Delta R(t)=\Delta N_{I,R}(t)-\Delta N_{R,.}(t). \tag{2.17}$$

Deterministic Transmission model. To investigate the simulated data and conduct statistical analysis, we formulated a deterministic model whose dynamics are represented using the following system of differential equations.

$$\frac{dV}{dt} = \omega p(t)N - \big(\delta + \omega\big)V, \tag{2.18}$$
$$\frac{dS}{dt} = \omega\big(1 - p(t)\big)N + \delta V - \big(\lambda(t) + \omega\big)S, \tag{2.19}$$
$$\frac{dE}{dt} = \lambda(t)S - (\sigma + \omega)E, \tag{2.20}$$
$$\frac{dI}{dt} = \sigma E - (\gamma + \omega)I, \tag{2.21}$$
$$\frac{dR}{dt} = \gamma I - \omega R. \tag{2.22}$$

For statistical inference, we defined a state variable that kept track of new cases generated by the model. This was achieved by calculating a piecewise continuous integral over time comparable to that the simulated data (assumed to be one week), and the expression for the new cases over the weekly interval [t-1, t) given by

$$C_t = \int_{t-1}^{t} \gamma I(u)du \tag{2.23}$$

Vaccine Reproductive Number (R_0^V). To quantify the effect of the vaccine in controlling diseases transmission, we derived a vaccine reproductive number using the next generation operator (NGO) [47]. This quantity allows to estimate the reduction in the transmissibility for a vaccine whose effectiveness wanes over time. A more general expression similar to ours has been derived by Magpanty *et al.* [123] which incorporates multiple modes of the vaccine imperfections. Initially consider a few terms, for vaccine coverage $p(t)$ at time t, the probability that vaccine derived immunity has waned in the vaccinated individuals is proportional to $\varepsilon^W = \delta/(\delta + \omega)$; $S^* = \Big(1 - (1 - \varepsilon^W)p(t)\Big)$ is the disease-free equilibrium for susceptible compartment. According to the method of next generation operator, I start by defining two rate matrices. One accounts for the rate at which of new infections can be generated. Here $\mathscr{F} = \begin{bmatrix} 0 & \beta S^* \\ 0 & 0 \end{bmatrix}$, and other provides the flux rates to and from the compartments which are involved in the infection generation process. In this case it is the E

compartment. Here $\mathscr{V} = \begin{bmatrix} (\omega+\sigma) & 0 \\ \sigma & (\omega+\gamma) \end{bmatrix}$. Then $\mathscr{K} = \mathscr{F}\mathscr{V}^{-1}$ is defined as the next generation operator. For this system, this is given as $\mathscr{K} = \begin{bmatrix} \frac{\beta\sigma S^*}{(\omega+\sigma)(\omega+\gamma)} & \frac{\beta(\omega+\sigma)}{(\omega+\sigma)(\omega+\gamma)} \\ 0 & 0 \end{bmatrix}$.

The highest eigen value of the next generation matrix estimates the basic vaccine reproductive number $\left(R_0^V(t)\right)$ at time t. In this case we get

$$R_0^V(t) = \frac{\beta\sigma\left(1-(1-\varepsilon^W)p(t)\right)}{(\omega+\gamma)(\omega+\sigma)}. \tag{2.24}$$

Parameter Estimation and Goodness of Fit. Trajectory matching was employed to estimate the parameters of the model that were assumed to unknown. The parameters that were estimated are represented in Table 2.2. The mean (μ) and variance (σ^2) of the observed cases was given by

$$\mu_t = \rho C_t, \tag{2.25}$$
$$\sigma_t^2 = \mu_t + \mu_t^2\psi. \tag{2.26}$$

Here, ρ, and ψ are the probability of reporting and the dispersion parameter respectively. The likelihood of the observed data conditioned on the model parameter vector Θ is given as

$$\mathscr{L}(\Theta) = \prod_t f(D_t;\mu_t,\sigma_t^2). \tag{2.27}$$

where $f(D;\mu,\sigma^2)$ is the probability density function of a negative binomial distribution. Maximization of of the likelihood function was carried out using Differential Evolutionary algorithm (DE) [193] which was implemented in R package DEoptim [138].

As a measure of quantitative goodness of fit I calculated the coefficient of determinism (R^2) between the log-transformed model predictions (M_t), and the observed data (D_t) respectively. R^2, for a linear relationship, estimates the proportion of the observed variation explained by the best fitting line, and it is given as

$$R^2 = \left(\frac{CoV(M_t, D_t)}{\sqrt{V(M_t)}\sqrt{V(D_t)}} \right)^2. \tag{2.28}$$

Here $CoV(M_t, D_t)$, $V(M_t)$, and $V(D_t)$ are the covariance between the observed, and the predicted values, and the corresponding variances for the two respectively. For conducting the model validation, the first 80y time-series (approx. 85%) was used for model fitting, and parameter estimation. The subsequent 15 years of simulated data (approx. 15%) was reserved to conduct out-of-fit model validation.

Bifurcation analysis. To investigate the discrepancy of the true model parameters and their estimated MLE values, I conducted spectral density analysis by enumerating frequencies of oscillation for the simulated data and the corresponding best fitting deterministic model. Systematically, two control parameters - the basic reproductive number (R_0) and the amplitude of seasonality (β_1), were varied to quantify the change in the pre-vaccine era epidemic dynamics which was summarized by calculating the reciprocal of the dominant frequency, *i.e.*the dominant period of oscillation (in years). For every combination of control parameter, the model simulated for 1040y. Discarding the initial 1000y as transients, the dominant period was calculated for the subsequent 40y time-series.

2.3 Results

Simulated partially observed disease system. Assuming parameter values represented in the Table 2.1, we simulated a stochastic realization of the system described in Eq. 2.2-2.17. The system was initialized near the deterministic equilibrium such that $\Big(S_{t=0} = N/R_0,\ E_{t=0} = N(R_0 -$

$1)\omega\sigma/R_0,\ I_{t=0} = N(R_0-1)\omega\sigma/R_0(\omega+\gamma)(\omega+\sigma),\ R_{t=0} = N - S_{t=0} - E_{t=0} - I_{t=0},\ V_{t=0} = 0\Big)$. The first 100 y were discarded as transients, and the next 80 years was considered as the system of interest (Fig. 2.3). The system is in the state of quasi-equilibrium for the first 35y. In the absence of demographic stochasticity, and seasonality, we would expect the dynamics to remain stable at the initialized values. Instead, we observe noisy cyclical dynamics near the endemic equilibrium. Vaccination was introduced in the year 27. In response to the gradual roll out of vaccine coverage, we see a proportional decrease in the susceptibility of the population from around 9.25% until year 40 to 8.75% by the year 45 and the corresponding increase in the percent vaccinated during this period (Fig. 2.3A–B). Half a percent decrease in the susceptibility is enough to cause precipitous reductions in the infected individuals (Fig. 2.3D–E). This drop is not reflected as drastically in the number of true new cases, and the under reported observations of the number of cases (Fig. 2.3F-G).

In this hypothetical scenario, we have assumed that the effectiveness of the administered vaccine wanes over the period of time ($1/\delta = 110$ y) On average. Even with an immune duration that lasts for a period that is greater than the assumed life expectancy ($1/\omega = 1/80y$), progressively more number of individuals lose their vaccine derived immunity and reenter the susceptible class. We observed a significant replenishment (9.75%) in the population level susceptibility by the end of year 57 which is 0.5% more than the pre-vaccine levels (Fig. 2.3A). This rise in the susceptibility coupled with low but continued imports of infectious individuals, generates resurgent epidemics in the later vaccine era as reflected in (Fig. 2.3G). We summarized the vaccine induced reduction in transmission by calculating the time inhomogenous $R_0^V(t)$(Fig. 2.3G), which like the observed cases, underwent a reduction of approx. 50% at the zenith of the vaccine coverage.

Model fits and parameter estimates. Using trajectory matching, we fitted a deterministic dynamical model that is structurally identical to the process model used to simulate the incidence data. Among the model parameters used in the data generation step, we assumed 6 parameters to be unknown. Table 2.2 lists the set of assumed unknown parameters. Our de-

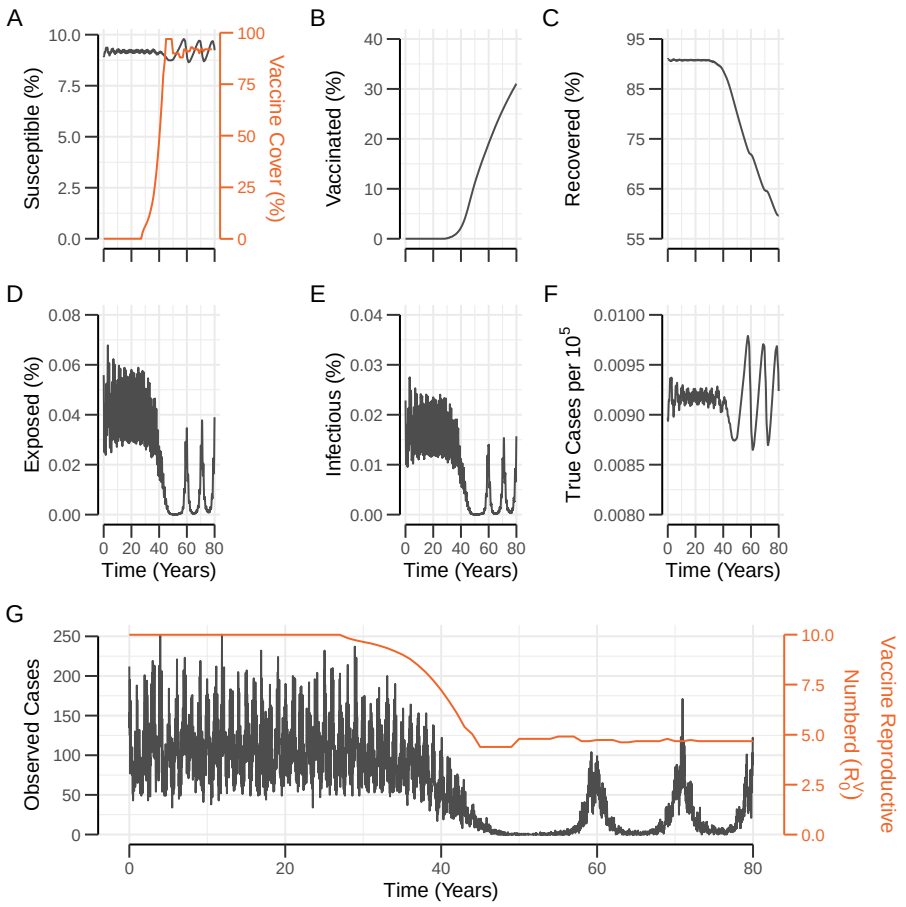

Figure 2.3: **Simulated System and Data**. Line plots represent a stochastic realization of the population dynamics of individuals in the system who are (A) susceptible, (B) vaccinated, (C) exposed, (D) Infectious, and (E) Recovered. (F) Represents the true number of cases per 10^5 individual. (G) shows the dynamics in the observed number of cases in the system. At strategically chosen set of parameter values, the system was simulated and allowed to reach an equilibrium. The first 100 y were discarded as transients and the remainder 80 y time-series was assumed to be data.

terministic model evaluated at the estimated MLE ($loglik = -17928.11$) statistically out performs the model when evaluated at the true parameter values ($loglik = -18329.72$). This discrepancy is characteristic of fitting a deterministic model to noisy dynamics produced from a stochastic process.

At the estimated MLE, our deterministic model fits strikingly well to the simulated data (Fig. 2.4A). During the pre-vaccine, and the early vaccine ear, we observed that the fit of the model remains roughly consistent. The initial decline in the number of cases caused due to the ramping up of the vaccine cover in the population is accurately tracked by our best fitting model. To identify regions of good and bad model performance, we evaluate the log density of the observation model at every time point (Fig. 2.4C–D). According to this diagnostic result, our model best performed in the epidemic troughs, while it's performance was wort while captures the peaks in the re-emergence era. Our model accurately reproduces the timing of reemergence of peaks, however, it over-estimates the peak sizes. Over-all, our best fitting model demonstrates high accuracy in predicting the observed dynamics ($R^2 = 91\%$, Fig. 2.4C). And although the error distribution is tightly clustered around the best fitting line, I observed a greater spread in the points among the more recent times. This result supports the qualitative fits where the model performs much better in the pre-vaccine and early vaccine eras than in reemergence era.

Table 2.2: Model parameter comparison

Parameter	True values	Estimated values
R_0	10	6.83
$1/\sigma$	13 d	25 d
β_1	0.1	0.21
$1/\delta$	110 y	40 y
ρ	0.005	0.0047
$1/\psi$	0.05	0.09
$loglik$	-18329.72	-17928.11

For our deterministic model, I estimated that duration of vaccine derived immunity that lasted approximately 40% of the real value($1/\delta_{True} = 110y,\ 1/\delta_{Estd.} = 40y$). I found that lower values of waning rate (δ), closer to the true parameter values, either did not capture the timing of the re-emergence or they simply did not the produce an epidemic. One other parameter that I unable unable to capture was the duration of latency ($1/\sigma_{True} = 13d,\ 1/\sigma_{Estd.} = 25d$).

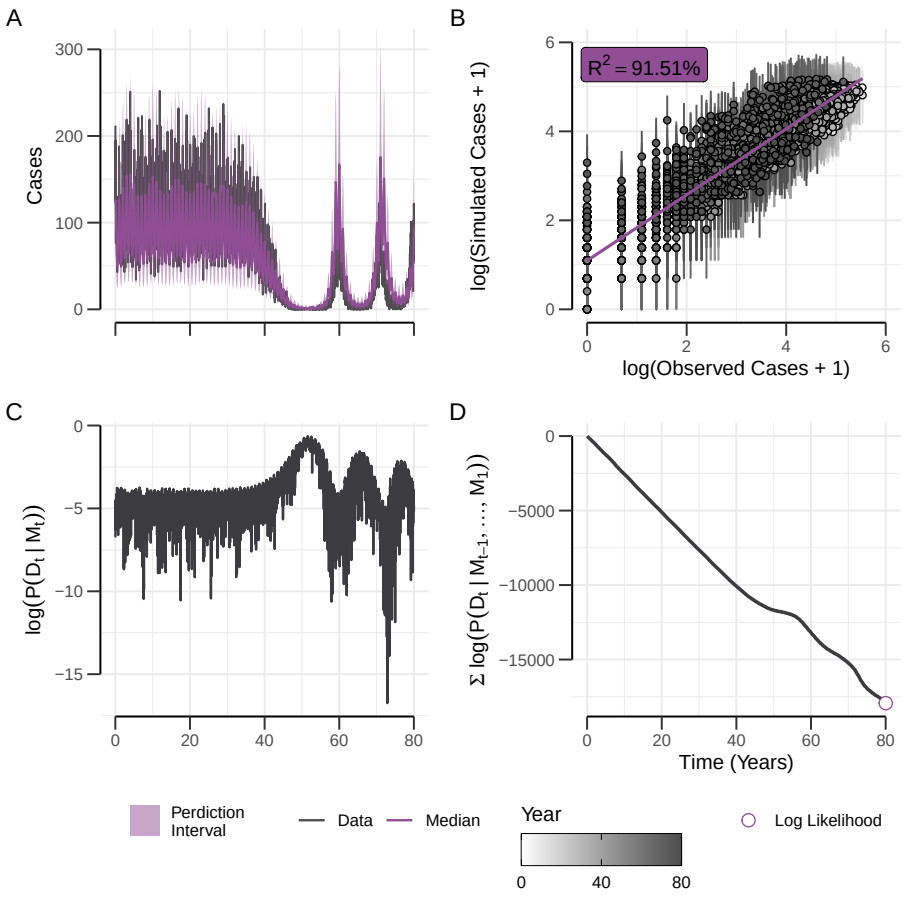

Figure 2.4: **Model-data agreement**. (A) Qualitative median model fits to the simulated data (0-40 years, purple) using the estimated MLE. Ribbon around the model trajectory represent the 95% prediction intervals. (B) Quantitative, log-linear model-data agreement at the MLE (Purple) assessed for the median trajectory. Error bars around the points represent the observation error distribution generated by simulating 1000 synthetic time series. (C) Log probability density, and (D) conditional log-likelihood evaluated for the best fitting model.

Out-of-fit model validation. A 15y subset of the simulated data (approx. 15% of observed time series) was reserved the assess the validity of the best fitting model (Fig. 2.5A). Over this duration of prediction horizon, I observed a precipitous drop in the explanatory power of our best-fitting model ($R^2 = 34.38\%$, Fig. 2.5B) during this epoch, as compared with the

dynamics in the with-in sample epoch. Compared with the dynamics in the observed data which are punctuated by 2 epidemic outbreaks, arranged roughly 10y apart, the model simulations predict high frequency annual outbreaks, modulated by low-frequency waves that are set roughly 10y apart. I observed that the discrepancy between the model prediction, and the observed data increased with the length of the prediction horizon. Here, the model appears to capture the general trends in the data with higher accuracy in the first 5y. As this the window of the prediction lengths, I observed an increasing degree of discrepancy betIen the model simulation, and the observed dynamics.

Bifurcation and spectral density analysis.

Majority of the dynamical landscape across combination of parameters appears to exhibit cycles with the dominant period of 1y. These annual dynamics are pronounced in regions with the high values of β_1 or R_0. Intermediate values of the two control parameters produce cyclical dynamics with multi-annual signatures. These periods ranged from 2y for intermediate values of $beta_1$ (approx. 0.5) and lower values of R_0 (approx 5.5). As the values transmissibility increases, ($R_0 = 7$), multi-annual cycles with a period greater than 8y are observed. The dynamics fluctuate between 1y and 4-8y cycles as the transmissibility is further increased ($R_0 >= 10$) (Fig. 2.6A, grey shaded regions).

Comparing the dynamic signature between the timer series of the simulated data and the best fitting model, both time series belonged to the region on the bifurcation diagram exhibiting annual epidemic regimes (Fig. 2.6A-C). Since, the estimation problem assumed the latency period within this system to be unknown, direct comparison between the of dynamics generated by the best fitting model to the dynamics of in the bifurcation diagram was not possible. To compare the two time-series, spectral density of the model simulated at the MLE ($1/\sigma = 25d$) was calculated separately. I observed very similar oscillatory behavior between between the data and the simulation generated from the best fitting model (Fig. 2.6A-C) both exhibiting a period of 1y.

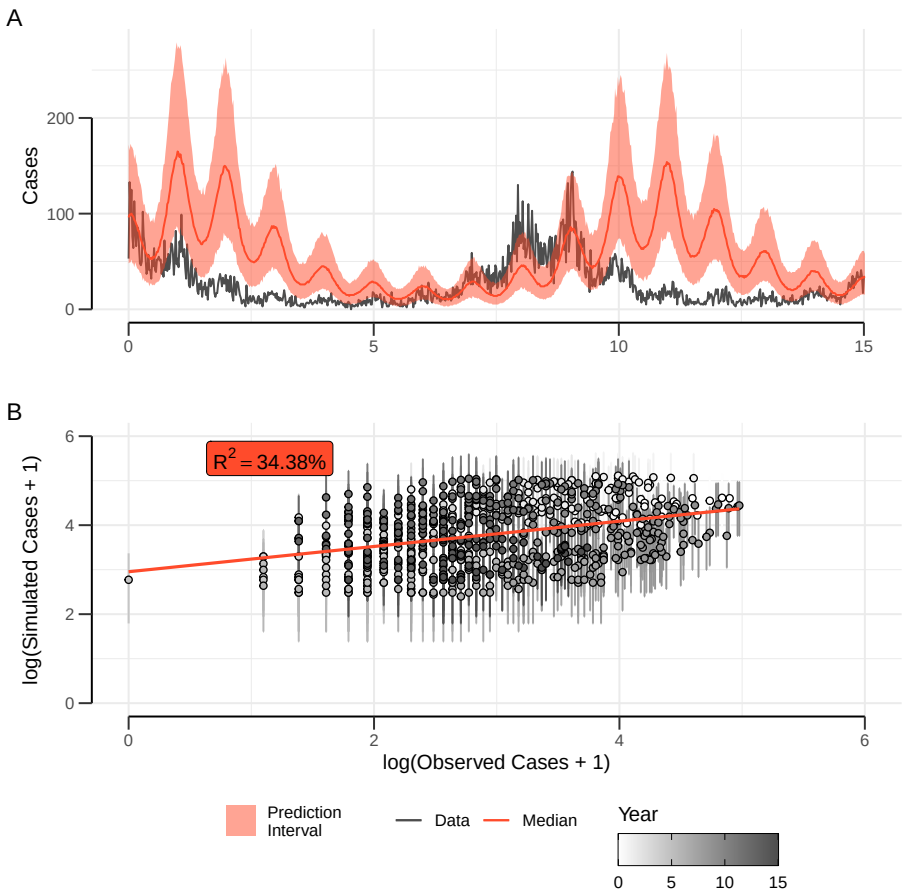

Figure 2.5: **Out of fit model-data agreement**. (A) Qualitative median model fits to the out of fit simulated data (41-80 years, orange) using the estimated MLE. Ribbons around the median model trajectory represent the 95% prediction intervals. (B) Quantitative, log-linear model-data agreement at the MLE (Orange) assessed for the median trajectory. Error bars around the points represent the observation error distribution generated by simulating 1000 synthetic time series.

2.4 Discussion

Sobering limitations in observing the data generating processes, introduce uncertainties that can be associated with system structure, demographic, extra-demographic and/or observation processes [95]. Mathematical models can be prove to be an indispensable tool to formulate and uncover

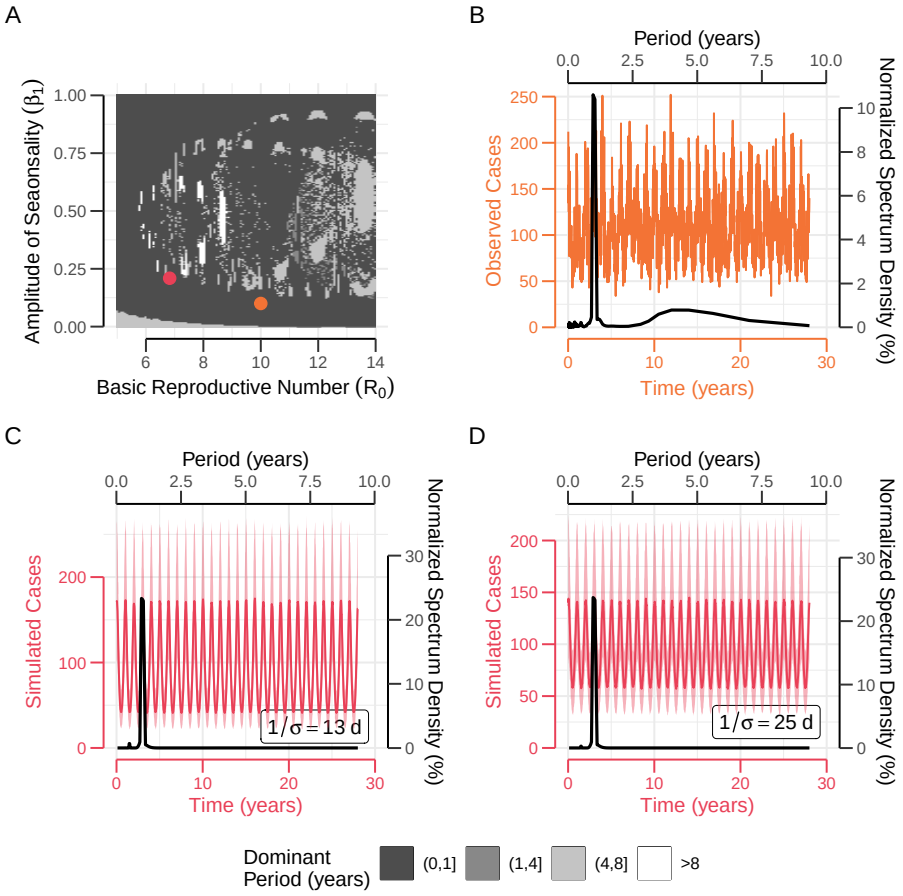

Figure 2.6: **Bifurcation and spectral density analysis**. (A) 2 dimensional bifurcation diagram for the pre-vaccine era representing the dominant period of the model dynamics as a function of amplitude of seasonality (β_1) and basic reproductive number (R_0). (B) Temporal dynamics in the synthetic data (left axis) and normalized spectral density for period of oscillation (right axis). Temporal dynamics in the simulated model (left axis) with 95% predictions intervals (ribbons), and normalized spectral density for the period of oscillation for (C) $1\sigma = 13$d and (D) $1\sigma = 25$d. The first 1000y of model simulations were discarded as transients. The subsequent 40y were considered for the calculation of the dominant period of the median simulated model realization.

mechanisms that drive observed dynamics. This is especially true in situations with paucity of access to the unobserved states of the system of

interest which is a crucial part of the puzzle that is often missing while investigating drivers of disease transmission.

The optimized best fitting deterministic model, was able to capture the observed the dynamics the within-sample data($R^2 = 91.51\%$). However, when assessing the model performance using the out-of of sample data, the statistical accuracy of our model prediction were highly diminished ($R^2 = 34.38\%$). From the quality of the model fit, in this era, it appears that the model performs relatively well in the first 5 years and begins to diverge rapidly from the data as the prediction horizon is expanded to first 10y and then 15y. This progressive deterioration in the explanatory power of the model can be attributed to long term transients in the deterministic model that ensue due to the time inhomogeneous roll-out of the vaccine dose. Slow decaying transients are a characteristic feature of deterministic models. Conversely, transients in the model with stochastic forcing, such as the one used to simulate the synthetic data for this study, are discarded within a much shorter time period, rapidly approaching a quasi-equilibrium [20]. The simulated dynamics from the fitted model there for exhibit high frequency oscillations interspersed with low-frequency long-term trends. Such dynamic modulations are absent in the simulated data.

For our best fitting model, all the model parameters that were used in the could not be recovered simulation of data (Table 2.2). One explanation for this discrepancy could be that dynamics that are good match to the observed data, generated using a deterministic model might require a very different parametric configuration than the one used in the data simulation using a stochastic model. To test this claim, a bifurcation analysis was using a two crucial parameters of the model – R_0 and β_1.

Interestingly, I observed substantial degeneracy in the effect of these parameters in generating model dynamics, such that a diverse combination of R_0 and β_1 were capable of generating annual cycles, a dynamical period dominant in the observed data. This collection included both the true ($R_0 = 10$ and $\beta_1 = 0.1$) and the estimated values ($R_0 = 6.83$ and $\beta_1 = 0.21$) of the parameters of interest. This means that its not that a very different parametric configuration might be required to generate dynamics similar to the observed data, in fact, a wide variety of combinations can generate

qualitatively similar dynamics. A more sophisticated analysis of the parameter values involving complete information based metrics such as the profile likelihood might be required to assess the identifiability of these model parameters comprehensively.

Fixing the vaccine coverage to the average value for the last 10y ($p = 92\%$), I calculated the the vaccine reproductive number (R_0^V, Eq. 2.24) for both the true ($R_0^V = 4.67$) and the estimated ($R_0^V = 4.74$) parameter values. This expected closeness in the values of R_0^V calculated from the true values and the estimated parameter values suggest that there is a trade-off in the parameter values of R_0 and probability of waning (ε^W). To reproduce the observed dynamics with a 37% reduced value of R_0, a higher value of ε^W, increased by 59%, is required to recapitulate the dynamics with great fidelity.

In this chapter, I make use of a hypothetical disease system to demonstrate fundamental concepts in interfacing incidence data with dynamical transmission models. Several communicable disease systems have be described as noisy oscillators [19]. Using a model that incorporated demographic stochasticity and temporally in-homogeneous rate of vaccine uptake, a synthetic time-series of incidence, representing a typical population with disease transmission and a gradual roll-out of vaccine was described. This analysis allowed to understand the consequences of utilizing deterministic models in systems influenced by demographic stochastic forcing. It shows that under varied dynamical assumptions about the underlying data-generating process, different explanations about the observed data are possible.

Further, through this analysis it was possible to identify the trade-offs which can limit the identifiability of certain parameters. Under such circumstances, it is recommended to partition the time series and estimate parameters separately. This is because subsets of observed data may encode different information about the underlying transmission process. In case of this study, it was mathematically shown that the basic reproductive number (R_0) and vaccine waning rate (δ) were not independently estimable. In this situation, the observed time-series could be split into two parts – pre-vaccine and post-vaccine eras. Then fitting a model the model to the

pre-vaccine era, one could estimate the value of R_0. Scaled estimates of δ, conditioned on the value of this pre-vaccine transmissibility, can then be quantified by fitting the model to the vaccine era while fixing the R_0 values constant during this epoch. Using completely known dynamical systems in this manner thus provides a crucial initial sanity check about the utility and performance of the hypothesized model while conducting mechanistic inference of infectious disease time series.

CHAPTER 3

INFORMING ANTIVIRAL EFFECTIVENESS FOR INFLUENZA A AND SARS-COV-2 BY QUANTIFYING WITHIN-HOST IMMUNE INTERACTIONS

3.1 Introduction

Antiviral therapies are among the most effective pharmaceutical interventions in the control and treatment of a variety of viruses [98]. These drugs function by interrupting the viral life cycle within host cells, with the ultimate aim of preventing infection, shortening the infectious period and curtailing the chain of transmission [99]. In addition to these direct effects, antivirals have been proposed as a candidate public health response both to contain outbreaks [81, 82], and as a strategy for prophylactic prevention of zoonotic spillover [58]. The efficacy and prescribed usage of any particular host-targeting antiviral depends on the cellular pathway that it manipulates in addition to the target viral kinetics [152, 163].

The infection process for respiratory viral infections begins with inhalation of aerosolized droplets emanating from an infectious host [154]. Disseminated viral particles then enter the upper respiratory tract and following a rapid replication phase in the mucosal epithelial cells, proliferate throughout the respiratory system. Viral production triggers a multifaceted cascade of rapid innate immune responses aiming for efficient control of the infection [141]. In devising antivirals, therefore, it is crucial to characterize the timing, mechanism and magnitude of the host's functional immune response relative to the virus replication cycle and proliferation [87, 99].

Animal infection experiments can be very useful in identifying the operative pathways and titrating the potential effectiveness of candidate drugs [2, 114, 144, 194]. There are several animal models available to assess infection, disease, transmission, and/or efficacy of interventions against influenza A virus and SARS-CoV-2, with each infection model having specific strengths and weaknesses. For influenza, ferrets and mice have been used since the first isolation of human influenza in the 1930s and dominate the experimental infection field (recently reviewed in [144]). Ferrets are naturally susceptible to H3N2 infection and efficient transmitters to naïve ferrets [23, 144]. The ferret model is hampered by a lack of reagents, however this drawback is slowly changing [2, 144]. Animal models for SARS-CoV-2 infection were rapidly explored and developed over the past 18 months. Like influenza, ferrets were quickly assessed as models of infection, disease and transmission (recently reviewed in [114]). As in the case for H3N2 infection, the ferret is naturally susceptible to SARS-CoV-2 infection and can transmit to naive animals via contact or aerosol [100]. The various animal infection model allied with mechanistic computational approaches can permit the testing of multiple hypotheses regarding viral proliferation and innate immunity. The predictions of such models can then be validated by subsequent concise *in vivo* infection studies.

Directly transmissible respiratory viruses like the influenza A H3N2 virus (H3N2), historically, and Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), more recently, have exacted a heavy toll on public health globally. Since its introduction into the human population following the 1968 influenza pandemic, H3N2 has accounted for substantial annual

deaths globally [154]. The estimated global disease burden for severe seasonal influenza cases is around 3-5 million cases annually with a median of 400,0000 recorded deaths [39, 86], with many of these attributed to H3N2. At the time of writing, the SARS-CoV-2 pandemic has already claimed over 5 million lives worldwide [206], inflicting an unprecedented strain on global systems, and especially healthcare [116, 207]. In addition to their substantial burden on morbidity and mortality, RNA viruses such as SARS-CoV-2 and H3N2 pose additional threats because of their potential for rapid divergent evolution [1]. Examples include the recent emergence of the highly transmissible B.1.617.2 (delta) and B.1.1.529 (omicorn) variants of SARS-CoV-2 [180, 187] and the repeated punctuated emergence of novel antigenic clusters of influenza viruses [22, 107, 188]. The combination of efficient transmission of these viruses, their potential for antigenic escape and the toll they exact on our population makes the potential use of antiviral therapies for their rapid and efficient control especially timely, along with other critical pharmaceutical interventions such as vaccination.

In this article, we examine the consequences for antiviral treatment on within-host dynamics of SARS-CoV-2 and H3N2 using a combination of data from animal infection experiments, mathematical models, and statistical inference. In the first phase, we present a suite of mechanistic models of varying degrees of complexity that aim to elucidate the potential role of different components of the host innate immune system during an infection. By comparing our best-fitting models, we note major differences in the relative contribution of the host immune system in modulating the infections of H3N2 and SARS-CoV-2, with stark differences in the within-host proliferation of the two viruses. Next, to understand the impact of antiviral timing, dose and mode of action, we incorporate antiviral dynamics into our fitted within-host model. Our numerical and analytical findings indicate efficient control requires virus-specific antiviral initiation time and effectiveness.

3.2 Methods

3.2.1 Infection Experiments

Procedure

Five to six-month-old, male and female (range 800-1200g) fitch ferrets (*Mustela putorius furo*, n=4 per group) were housed in pairs in the animal biosafety level 3 enhanced (ABSL3E) facilities at the University of Georgia. After a period of acclimation, ferrets were anesthetized with isoflurane and intranasally inoculated with H3N2, A/Singapore/INFIMH-16-0019/2016 strain (IRR FR-1590; SG/16) or SARS-CoV-2 virus (USA-WA1/2020, BEI Resources, NR-52281) diluted in 1.0 ml of Phosphate Buffered Saline (PBS) solution. Virus inoculum dose delivered to each ferret was 10^6 PFU (plaque-forming unit)/ml of IAV H3N2 or 10^6 FFU (focus forming units)/ml of SARS-CoV-2. Following inoculation, animals were monitored daily for clinical signs, including nasal discharge, sneezing, diarrhea, lethargy, increased respiratory rate and effort (congestion), cyanosis, neurological changes, and altered responses to external stimuli. Nasal washes, weights and temperatures were obtained every other day under light sedation. For nasal washes collection, ferrets were anesthetized with isoflurane and their nares irrigated with 3.0 ml of PBS. Infectious viruses were titrated in either Madin-Darby canine kidney (MDCK) or Vero E6 cells by PFU or FFU assays, respectively. A separate aliquot of nasal wash was processed for RNA isolation and subsequent quantification by qRT-PCR [4]. At the end of the study, infected animals were humanely euthanized.

Ethics Statement

All animal experiments were approved by the University of Georgia Institutional Animal Care and Use Committee (protocol number A2020 03-016). Animals were housed and maintained in accordance with the US Animal Act PL 99-158 and in accordance with guidelines stated in the “Guide for the Care and Use of Laboratory Animals”.

3.2.2 Within-Host Dynamics

Model of Infection and Innate Immunity

To systematically explore potential interactions between viral proliferation and the host innate immune responses, we developed a family of mechanistic models incorporating a number of distinct immuno-ecological hypotheses, as illustrated in Fig. 3.1.

Respiratory viruses such as H3N2 and SARS-CoV-2 commonly replicate in the epithelial cells of the upper respiratory tract [179], on target cells. In general, upon infection, cells can go through a transient eclipse phase [15] during which infectious virus is absent as virus genome undergoes intracellular replication and viral proteins are produced. Subsequently, viral particles are assembled within the infected cells, leading to the viral burst of infectious virions.

We assume target cells (denoted by T) are infected at the rate β_T and become productively infectious after a transient latency (the eclipse phase), lasting $\frac{1}{\delta_E}$ days. To be consistent with our experimental data sources, we assume nasal washes are composed of a mixture of viral RNA and viable virus. Specifically, we assume productively infectious cells (I) produce viral RNA (P) at rate ν and a subset (ε) of this population is assembled into viable virus particles (V).

Given the short temporal duration of these experiments, typically lasting no more than 8 days, we limit our investigation around the role of the host innate immune response. Particularly, we focus on examining the action of rapidly acting, cytokine-mediated immune responses such as type I interferons α, β (IFN-1) and type III interferon λ (IFN-3) [120, 150, 212]. This family of signalling molecules generally forms the majority of the host's initial broad sense response [141].

Three functions that are jointly associated with the release of IFN-1 and the IFN-III include

- **Increase in the prevalence of refractory epithelial cells**. Refractory cells can prevent viral entry. A high proportion of such cells can thus reduce the effective proliferation rate of the virus [164].

- **Inhibited viral shedding**. Cytokine production can promote hostile local conditions by enhancing environmental toxicity, inflammation or by recruiting resident macrophages that can attack extra-cellular viral particles. Depleting the extra-cellular viral population effectively reduces the viral production rate [89, 177].
- **Natural killer (NK) cell recruitment**. NK cells are cytotoxic lymphocytes that do not require activation to kill cells [18]. IFN-1 production can enhance the recruitment of lymphocytes such as natural killer cells, which can identify infected cells and rapidly destroy them by stimulating apoptosis thereby impeding within-host viral spread.

IFN-3, λ has similar functions to the IFN-1, but is generally considered to be produced earlier at mucosal surfaces and less potent than the activities of interferons α and β [113].

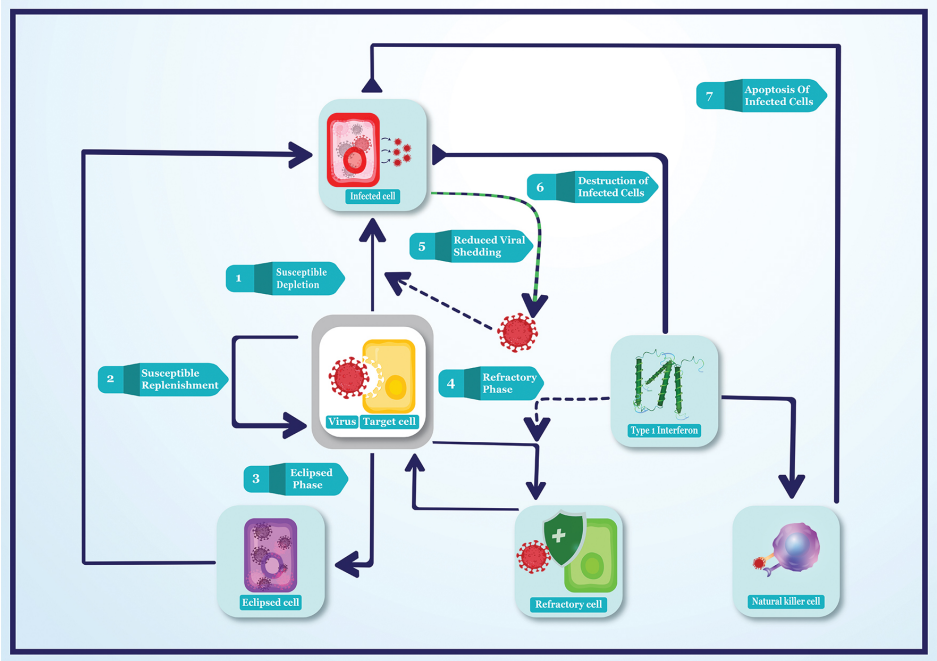

Figure 3.1: **Schematic of the focal hypotheses considered in this study.** All sub-populations involved in the host-virus interactions are colour coded. Arrows indicate promotion and flat-headed arrows indicate inhibition reactions describing cellular interactions.

Table 3.1: **Model compartments for the model of infection and immunity.**

Compartment	Definition
T	Target epithelial cells, invasion sites
E	Eclipsed epithelial cells
I	Infectious epithelial cells, replication sites
V	Viral titer
P	Viral RNA
R	Refractory epithelial cells
F	Type I (α & β) type III (λ) interferon
N	Natural killer cells

Hypothesis I. Susceptible cell Depletion. The simplest model incorporates epithelial target cells (T) that are infected by viable virions (V) at *per capita* rate β_T. Upon transmission, infected cells (I) produce novel viral RNA (P) at rate ν and viable viral particles (V) at a reduced rate ($\varepsilon\nu$, where $\varepsilon < 1$). Thus, in this model, the main driver of viral dynamics is the depletion of target cells.

$$\frac{dT}{dt} = -\beta_T VT, \tag{3.1}$$
$$\frac{dI}{dt} = \beta_T VT - \delta_I I, \tag{3.2}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.3}$$
$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.4}$$

Hypothesis II. Susceptible Replenishment. This is an extension of the model described by equations (3.1)–(3.4) and includes the replenishment of target cells. Specifically, we assume target cell dynamics are described by logistic growth with a maximum growth rate ν_T and a 'carrying capacity' of (κ_c).

Table 3.2: Model parameters with definition and units for the full model of infection and immunity. Parameters are grouped by the hypotheses and arranged in an order of increasing biological complexity.

	Parameter	Value/Range	Unit	Definition
Immune independent dynamics				
(H1) Susceptible depletion	κ_c	7×10^7 [29]	n_T	Target cell population size
	ν	≥ 0	$n_V n_T^{-1} d^{-1}$	Viral RNA production rate
	ε	[0,1]	-	Viable virus production rate
	δ_P	≥ 0	d^{-1}	Viral RNA clearance rate
	δ_V	≥ 0	d^{-1}	Viral clearance rate
	δ_I	≥ 0	d^{-1}	Infected cell clearance rate
	β_T	≥ 0	$n_T n_V^{-1} d^{-1}$	Infection rate
(H2) Susceptible replenishment	ν_T	≥ 0	$n_T d^{-1}$	Target cell production rate
(H3) Eclipsed phase	δ_E	≥ 0	d^{-1}	Eclipse cell conversion rate
Immune dynamics				
Type-1 interferon dynamics	ν_F	≥ 0	d^{-1}	Activation rate
	δ_F	≥ 0	d^{-1}	Deactivation rate
(H4) Refractory cell dynamics	ϕ	≥ 0	$n_T n_F^{-1} d^{-1}$	Production rate
	ρ	≥ 0	$n_R d^{-1}$	Reversion rate
(H5) Reduced viral replication	π	[0,1]	-	Sensitivity to interferon
(H7, H6) Natural killer cell dynamics	ν_N	≥ 0	$n_K n_V^{-1} d^{-1}$	Recruitment rate
	δ_N	≥ 0	d^{-1}	Clearance rate
	ψ	≥ 0	$n_I n_N^{-1} d^{-1}$	Attack rate
	ζ	≥ 0	$n_I n_F^{-1} d^{-1}$	Attack rate (implicit)

$$\frac{dT}{dt} = \nu_T T\left(1 - \frac{T+I+R}{\kappa_c}\right) - \beta_T VT, \tag{3.5}$$
$$\frac{dI}{dt} = \beta_T VT - \delta_I I, \tag{3.6}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.7}$$
$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.8}$$

Hypothesis III. An eclipse phase. This model extends the system described in Hypothesis I with the inclusion of eclipsed infectious cells (E). The assumption here is that within infected cells, there is a period of viral replication before the cell becomes productively infectious (akin to the latent class in population models [96]). Cells in the eclipsed phase (E) are converted to productively infectious cells (I) at rate δ_E.

$$\frac{dT}{dt} = -\beta_T VT, \tag{3.9}$$
$$\frac{dE}{dt} = \beta_T VT - \delta_E E, \tag{3.10}$$
$$\frac{dI}{dt} = \delta_E E - \delta_I I, \tag{3.11}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.12}$$
$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.13}$$

Hypothesis IV. IFN-induced refractory cells. Starting with the model described in equations (3.1)–(3.4), this hypothesis examines the impact of IFN-I and IFN-III (F) production, stimulated by the presence of infectious cells (I) leading to the generation of refractory target cells (R), which are uninfected target cells that are transiently resistant to viral attack. IFN-I and IFN-III molecules are produced at rate ν_F and bind to and convert target cells to refractory cells at rate ϕTF. Refractory cells are resistant to viral infection for an average of $1/\rho$ days.

$$\frac{dT}{dt} = -\beta_T VT - \phi TF + \rho R, \tag{3.14}$$
$$\frac{dI}{dt} = \beta_T VT - \delta_I I, \tag{3.15}$$
$$\frac{dR}{dt} = \phi TF - \rho R, \tag{3.16}$$
$$\frac{dF}{dt} = \nu_F I - \delta_F F, \tag{3.17}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.18}$$
$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.19}$$

Hypothesis V. IFN-induced reduction in viral shedding. This model extends equations (3.1)–(3.4) by incorporating an alternative antiviral action of IFN-I and IFN-III molecules. Specifically, the release of Type I and III interferons is implicitly assumed to reduce viral shedding by an infectious cell (I) by a factor $e^{-\pi F}$.

$$\frac{dT}{dt} = -\beta_T VT, \tag{3.20}$$
$$\frac{dI}{dt} = \beta_T VT - \delta_I I, \tag{3.21}$$
$$\frac{dF}{dt} = \nu_F I - \delta_F F, \tag{3.22}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.23}$$
$$\frac{dV}{dt} = \nu\varepsilon I e^{-\pi F} - \delta_V V. \tag{3.24}$$

Hypothesis VI. Natural killer cells. This model incorporates the impact of natural killer cells (NK cells, N) into the default model (equations 3.1–3.4). Here, we assume IFN-I and IFN-III release stimulates the recruitment of NK cells at rate $\nu_N F$. NK cells recruited to the site of infection kill infected cells at rate ψNI also leading to the clearance of the acting NK cells.

$$\frac{dT}{dt} = -\beta_T VT, \tag{3.25}$$
$$\frac{dI}{dt} = \beta_T VT - \delta_I I - \psi NI, \tag{3.26}$$
$$\frac{dF}{dt} = \nu_F I - \delta_F F, \tag{3.27}$$
$$\frac{dN}{dt} = \nu_N F - \psi NI - \delta_N N, \tag{3.28}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.29}$$
$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.30}$$

Hypothesis VII. Natural killer cells-2. Here, in contrast to the model presented in equations (3.25)–(3.30), we assume the release of IFN-I and IFN-III generates an inflammatory response leading to the apoptosis of infected cells (I) at rate ζF.

$$\frac{dT}{dt} = -\beta_T VT, \tag{3.31}$$
$$\frac{dI}{dt} = \beta_T VT - \delta_I I - \zeta FI, \tag{3.32}$$
$$\frac{dF}{dt} = \nu_F I - \delta_F F, \tag{3.33}$$
$$\frac{dP}{dt} = \nu I - \delta_P P, \tag{3.34}$$
$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.35}$$

Derivation of R_0

To summarize the within host transmissibility of a virus, we analytically derived the expression for basic reproductive number (R_0) for this system. R_0 for the within-host viral kinetics predicts the average number of infected cells generated by a typical infectious cell at the start of the infection [21]. We simplify the above model (Eq. 3.1-3.4) to derive the expression for (R_0):

$$\frac{dT}{dt} = -\beta_T V T, \tag{3.36}$$

$$\frac{dI}{dt} = \beta_T V T - \delta_I I, \tag{3.37}$$

$$\frac{dV}{dt} = \nu\varepsilon I - \delta_V V. \tag{3.38}$$

In the absence of virus, the within-host system has a stable equilibrium at (T^*, I^*, V^*)= (κ_c, 0, 0). We used the next generation operator [47] to derive an expression for R_0:

1. The compartments that contribute to the production of novel infections are I and V. $\mathscr{F} = [\beta_T V T,\ 0]$, is the vector of the rates at which new infections are generated. $\mathscr{V} = [\delta_I I,\ -\nu\varepsilon I + \delta_V V]$, is the vector of rates at which the virus flows through the system compartments.

2. For each of these vectors, $\mathscr{F}$ and $\mathscr{V}$, we defined the Jacobian matrices, evaluated at the virus-free equilibrium. F is such that every element $\mathrm{F}_{i,j} = \frac{\partial \mathscr{F}_i}{\partial Z_j}$, and V is such that every element $\mathrm{V}_{i,j} = \frac{\partial \mathscr{V}_i}{\partial Z_j}$, where $\mathrm{Z} = [\mathrm{I}, \mathrm{V}]$, is the vector of relevant state variables in generation of novel infected cells. This yielded $\mathrm{F} = \begin{bmatrix} 0 & \beta_T \kappa_c \\ 0 & 0 \end{bmatrix}$ and $\mathrm{V} = \begin{bmatrix} \delta_I & 0 \\ \nu\varepsilon & \delta_v \end{bmatrix}$.

3. The next generation operator is therefore given by $\mathrm{K} = \mathrm{FV}^{-1} = \begin{bmatrix} \frac{\beta_T \kappa_c \nu\varepsilon}{\delta_I \delta_V} & \frac{\beta_T \kappa_c}{\delta_V} \\ 0 & 0 \end{bmatrix}$. Solving the characteristic polynomial $|\mathrm{K} - \Lambda \mathrm{I}| = 0$ (where, I is the identity matrix) and identifying the dominant eigenvalue yields the expression for the basic reproductive number (R_0):

$$R_0 = \frac{\beta_T \kappa_c \nu\varepsilon}{\delta_I \delta_V}. \tag{3.39}$$

3.2.3 Sensitivity analysis

To identify model parameters critical to producing infection dynamics we conduct a sensitivity analysis on three model outcomes - (i) the peak viral load; (ii) the timing of the viral peak; and (iii) the slope of the viral decay. A sample of 500 parametric combinations is generated using a Sobol sampling scheme [190, 191] and the three model outcomes are calculated for each sampled parameter combination. The partial ranked correlation coefficient (PRCC) is calculated to quantify the association between model outcomes and individual parameter values. PRCC is an effective sampling metric to evaluate sensitivity of model outcomes on parameters [124]. We generate a PRCC bootstrapped distribution, of size 150, by sampling with replacement from the population of parameter configuration. Statistical significance of the PRCC is ascertained at 5% level of significance (LOS). Bonferroni's correction is applied to the LOS to account for error inflation associated with multiple comparisons.

3.2.4 Statistical inference and model selection

Observation model

For both H3N2 and SARS-CoV-2 viruses, data from ferret infection experiments is used to arbitrate among the competing models. Maximum likelihood estimation of unknown parameters is carried out by trajectory-matching in the R package *pomp* [102, 103]. Here we assume log-normally distributed observation error. We denote our data by $\{x\}_{n=1,t=0,r=1}^{2,T,4}$, where n represents the data type (viable virus and RNA), T represents the time course of infection, and r represents replicate animals. The likelihood function for these data is thus given by:

$$\mathscr{L}(\Theta) = \prod_{n\in V,P}\prod_{t=T}^{t}\prod_{j=1}^{n}\begin{cases} P\left(x_{n,t,r}|\iota_n Y_n(t,\Theta),\sigma_n\right) & \text{for } \iota_n Y_n(t,\Theta) > \mathrm{DT}_n, \\ P\left(x_{n,t,r}|DT_n,\sigma_n\right) & \text{for } \iota_n Y_n(t,\Theta) \leq \mathrm{DT}_n. \end{cases} \tag{3.40}$$

Here, ι_n is the data-type specific probability of under-sampling, $Y_n(t,\Theta)$ is the model solution of the data-type at time t, σ_n is the time invariant standard deviation of the log-normal distribution whose probability density function is given as $P(.|\Lambda_n)$ where Λ_n is the parameter vector of this distribution (see Table 3.3).

Viral infection data (titers and RNA) are right censored to account for the finite limit of the detection associated with viral quantification assays. Further, these detection thresholds (DT) differed between virus type (H3N2 and SARS-CoV-2) and data type (viable virus, RNA; Fig. 3.2). We account for right censoring in the data by incorporating a threshold in the observation model. If the sampled trajectories drop below the threshold DT, according to Eq. 3.40, the likelihood is evaluated at the data-type specific threshold DT_n. If not, the likelihood is evaluated at the sampled values, $Y_n(t,\Theta)$.

Table 3.3: Model parameters with definition for the joint observation model for viral titers and RNA.

Parameter	Range	Definition
ι_V	[0,1]	Virion sampling probability
σ_V	≥ 0	Log virion standard deviation
ι_P	[0,1]	Viral RNA sampling probability
σ_P	≥ 0	Log viral RNA standard deviation

Numerical optimization

Maximization of the likelihood (Eq. 3.40), is carried out using a differential evolution, stochastic optimizer routine offered by R package *DEoptim* [137]. For all hypotheses, a large population of parametric configurations (g = 500) is used to initialize the optimizer. This population is generated using a Sobol design to ensure efficient sampling of the parametric space [190, 191]. Trial mutants $v_{i,g}$ under this computational strategy are generated using three randomly chosen candidates from the original population such that $v_{i,g} = x_{0,g} + F(x_{1,g} - x_{2,g})$ [9]. We assume a high cross-over probability (0.9) and a moderate scaling factor (0.6) [84]. Owing to the complexity of the estimation problem, a very strict convergence

criterion is set to ensure convergence to a global maximum (steptol = 1500 iterations, reltol = 10^{-8}).

Quantifying goodness of fit

We use Akaike information criterion corrected for a small sample (*AICc*, n = 8 and 6 for IAV H3N2 and SARS-CoV-2, respectively) to assess the relative goodness of fit among the 7 competing models:

$$AICc = 2k - 2\log(\mathscr{L}(\Theta)) - \left(\frac{2k^2+2k}{n-k-1}\right). \tag{3.41}$$

Here, $\mathscr{L}(\Theta)$ is the likelihood function from Eq. 3.40 and k denotes the number of model-specific parameters. At a 5% level of significance, an *AICc* value of 2 was taken to signify a statistical difference in the comparative agreement among competing models. $\Delta AICc = 0$ was taken to denote the best mechanistic explanation of the observed dynamics.

Quantifying uncertainty in parameter estimates

The method of parametric bootstrap is used to estimate uncertainty in parameter MLEs. For this procedure 100 synthetic trajectories are simulated for each virus-model combination using the MLE. A global search optimizer, with less stringent convergence criterion (steptol = 500 iterations, reltol = 10^{-8}), is used to re-estimate model parameters for every synthetic time series. This generates a bootstrapped distribution of parameter estimates surrounding the MLE. 95% confidence bounds for parameter uncertainty are ascertained by calculating the 2.5^{th} and 97.5^{th} percentiles of the bootstrapped distribution of model parameters.

3.2.5 Effect of antivirals

Antiviral dynamics

Antivirals are chemical or biological agents that impede the within host transmission process. They may achieve this via different mechanisms:

(I) rendering naïve target cells partially resistant to infection; (II) inhibiting viral replication, or blocking viral fusion, or promoting virus neutralization, thereby reducing, on average, the viral output from productively infected cells; and (III) increasing the cytotoxicity associated with infected cells by assisting in the identification and flagging for resident antagonistic cells like macrophages and natural killer cells. To study the differential effect of these qualitatively distinct antiviral mechanisms, we construct a model incorporating an antiviral exhibiting first-order kinetics as shown in Eq. 3.42-3.49.

After the initiation of therapy at time t_{dose}, antiviral molecules infiltrate both susceptible target cells (T) and productively infected cells (I) at the rate β_A thus giving rise to uninfected target cells (T_A) and productively infected cells (I_A) in an antiviral phase. This allows for an explicit description of the effects of the antivirals on the transmission process. We allow for periodic administration of the antiviral at t_{dose} intervals.

$$\frac{dT}{dt} = -\beta_T VT - \phi TF - \beta_A AT + \rho R, \tag{3.42}$$

$$\frac{dI}{dt} = \beta_T VT - \beta_A AI - \delta_I I, \tag{3.43}$$

$$\frac{dR}{dt} = \phi TF - \rho R, \tag{3.44}$$

$$\frac{dF}{dt} = \nu_F I - \delta_F F, \tag{3.45}$$

$$\frac{dT_A}{dt} = \beta_A AT - \beta_T(1-\omega)VT_A, \tag{3.46}$$

$$\frac{dI_A}{dt} = \beta_T(1-\omega)VT_A + \beta_A AI - \frac{\delta_I I_A}{(1-\eta)}, \tag{3.47}$$

$$\frac{dV}{dt} = \nu\varepsilon(I + (1-\alpha)I_A) - \delta_V V, \tag{3.48}$$

$$\frac{dA}{dt} = -\delta_A A. \tag{3.49}$$

$$\text{where } A(t) = \begin{cases} A_{max} & t = x\delta(t, t_{dose}), \\ 0 & \text{otherwise.} \end{cases}$$

Here $\delta(t, t_{dose})$ is the Kronecker delta and x is an integer. Definitions on antiviral model compartments and parameters are provided in tables 3.4 & 3.5.

Table 3.4: **Model compartments for the model of antiviral action.**

Compartment	Definition
T	Target epithelial cells, invasion sites
I	Infectious epithelial cells, replication sites
V	Viral titer
R	Refractory epithelial cells
F	Type 1 interferon (α and β)
T_A	Target epithelial cells, anti-viral phase
I_A	Infectious epithelial cells, anti-viral phase
A	Anti-viral concentration

As described above, we consider three separate mono-therapies that represent qualitatively distinct effects on T_A and I_A. First, antiviral target cells (T_A) have reduced susceptibility to infection, denoted by $(1-\omega$; see Eq. 3.46). Second, the *per capita* baseline infectious period of I_A cells $(1/\delta_I)$ is reduced by a factor of $(1-\eta)$, thereby increasing the average cytotoxicity of antiviral-containing infected cells (see Eq. 3.47). Third, the average viral production per I_A cell is reduced by a factor $(1-\alpha)$ thereby reducing the viral burst size per productive cycle (see Eq. 3.48).

Antiviral performance indicator

Measure of therapeutic performance is ascertained using the virological efficacy index, which we define as the percent reduction in the area under the viral curve under treatment ($AUVC$) compared to the the area under the viral curve for the control ($AUVC_c$) [28]. Area under the curve in both cases was approximated using the trapezoidal rule implemented in the R package AUC [13].

$$\Gamma = 100\left(1 - \frac{AUVC}{AUVC_c}\right)\% \tag{3.50}$$

Table 3.5: Model parameters with definition and units for the full model of infection and immunity. Parameters are grouped by the hypotheses tested in this article and arranged in the increasing order of biological complexity. Individual model formulations are specified in Eq. 3.42-3.49

Parameter	Value/Range		Unit	Definition
	H3N2	*SARS-CoV-2*		
Non-immune dynamics (Fixed at the MLE)				
Susceptible depletion				
κ_c	7×10^7 [29]		n_T	Target cell population size
ν	9987.7	997.7	$n_V n_T^{-1} d^{-1}$	Viral RNA production rate
ε	4.94×10^{-1}	1.8×10^{-3}	-	Viable virus production rate
δ_V	8.56×10^5	1.17×10^{-1}	d^{-1}	Viral clearance rate
δ_I	5.90	4.04×10^1	d^{-1}	Infected cell clearance rate
β_T	2.37×10^{-4}	2.90×10^{-4}	$n_T n_V^{-1} d^{-1}$	Infection rate
Immune dynamics (Fixed at the MLE)				
Type-1 interferon dynamics				
ν_F	-	1×10^{-4}	d^{-1}	Activation rate
δ_F	-	3.21	d^{-1}	Deactivation rate
Refractory cell dynamics				
ϕ	-	1.41	$n_T n_F^{-1} d^{-1}$	Production rate
ρ	-	1.22	$n_R d^{-1}$	Reversion rate
Therapeutic Intervention (Varied)				
Antiviral dynamics				
A_{max}	1		n_A	Initial dose size
β_A	100		$n_T n_A^{-1} d^{-1}$	Binding rate
δ_A	1		d^{-1}	Decay rate
α	[0,1]		-	Reduction in viral replication
ω	[0,1]		-	Reduction in viral proliferation
η	[0,1]		-	Increase in cytoxicity
t_{dose}	[1,4]		-	Initiation time

3.3 Results

Viral dynamics

There are virus-specific differences in the quantified viral dynamics across individual animal replicates for H3N2 and SARS-CoV-2 (Fig. 3.2). H3N2 shows a more rapid viral take off that is consistent among individual animals, reaching an earlier peak (day 3) relative to SARS-CoV-2 that exhibits greater inter-individual variability in initial dynamics and reaches a maximum viral load typically on day 5. Both viruses fall below the detection threshold by day 11. Unlike H3N2, we observe a multi-modal infection curve for SARS-CoV-2. We examine whether sex-specific differences account for this variation in viral kinetics but find no discernible pattern (Fig. 3.5).

Parameter sensitivity to model outcomes

To explore how simulated dynamics respond to changes in model parameters, we conduct a systematic sensitivity analysis calculating partial rank correlation coefficients (PRCC) between combinations of parameter and three model outcomes. We find that nearly all of the PRCC values were significant ($p - value < 0.05$). However, many of the model parameters have small effect sizes ($PRCC < 0.1$). Among the model parameters, Initial viral load (V_0) and the basic reproductive number (R_0) have the largest *PRCC* associated with viral decay rate and the peak viral load (see Fig. 3.6). It should be noted that R_0 is a composite derived quantity and the effects of individual parameters (see Eq.3.39) are not considered individually mainly to assess the system sensitivity while reducing the dimension of the parameter space. We do not observe large effect sizes for immunity parameters ($PRCC < 0.2$, Fig. 3.6) in generating the chosen model outcomes. One reason leading to this result could be attributed to the fact that simple model outcomes, principally the initial peak size, timing and decay slope, are short-term phenomena that are driven by the processes associated with transmission and not immunity. Immune mechanisms tend to generate more complex dynamics on a longer time scale.

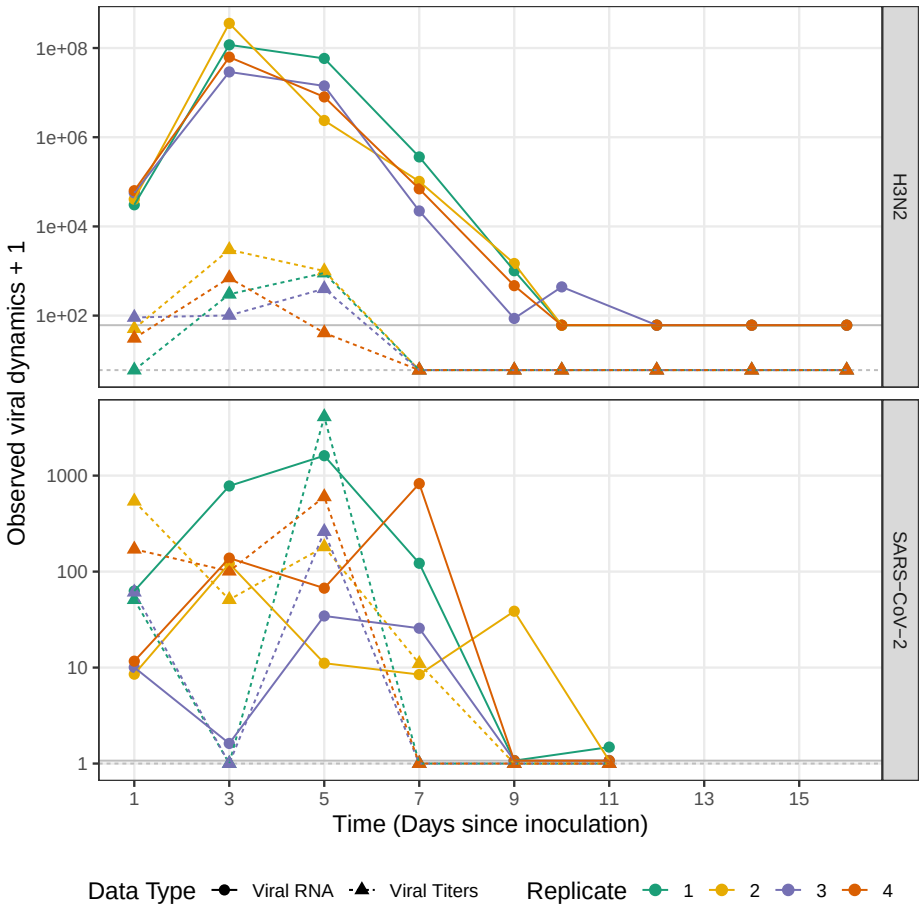

Figure 3.2: **Experimental data from ferret infection experiments.** Facets represent within-host viral dynamics of H3N2 influenza (top) and SARS-CoV-2 (bottom) collected from nasal washes of 4 individual ferrets (colors). Viral RNA data is represented using points and solid lines; Viral titer data is represented using triangles and dotdashed lines. Different detection threshold (dt) against virus-data type combinations are represented using solid and dotdashed lines for viral RNA and titer respectively. X-axis represents time in days since ferrets were inoculated.

Statistical inference with transmission models

We use the viral data obtained by both PFU/FFU and qRT-PCR assays (Fig. 3.2) in likelihood-based statistical inference to evaluate the evidence in support of each hypothesis (as represented in Eq. 3.1–3.35). As depicted in Fig. 5.2, all of proposed models can provide a good explanation of the

data. However, as indicated by ΔAICc values, we find individual models for H3N2 and SARS-CoV-2 that best fit the data (Fig. 5.2). We further quantify model-data agreement for the best-fitting model for each virus and data type: H3N2 (R^2 for viral RNA = 99.81%, viral titers = 85.22%; Fig. 3.7) and SARS-CoV-2 (R^2 for viral RNA = 92.68%, viral titers = 49.20%; Fig. 3.7).

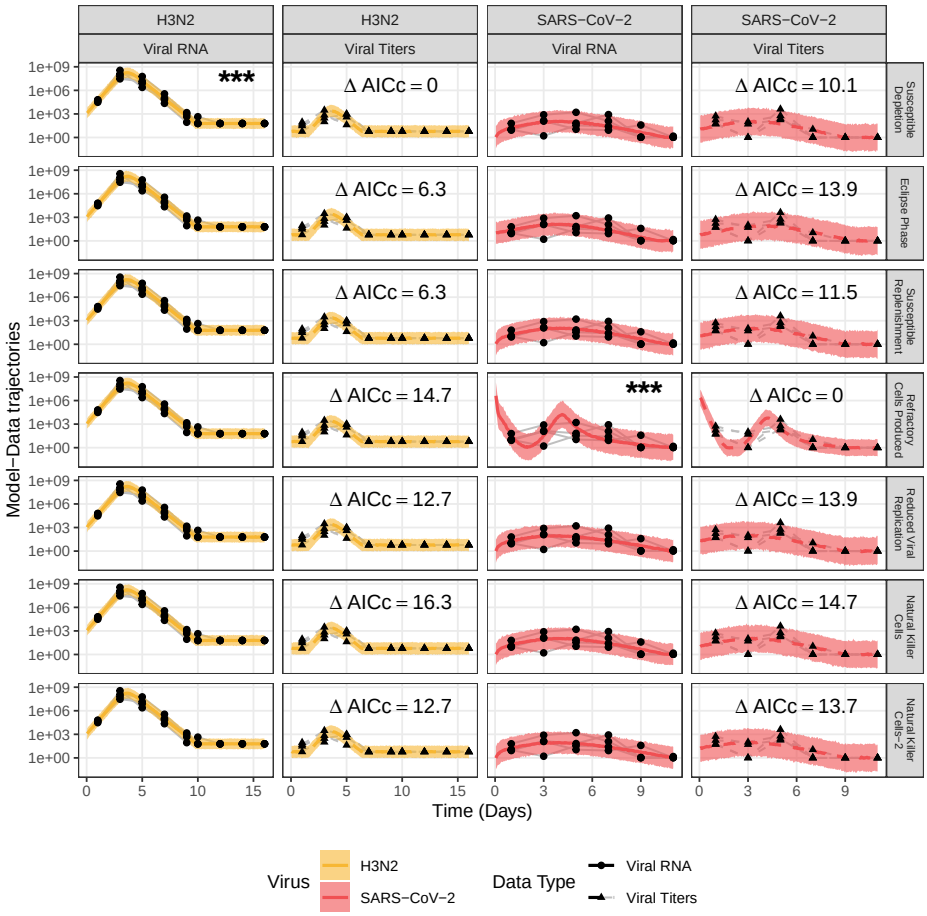

Figure 3.3: **Relative fits of within-host models to virus dynamics of H3N2 and SARS-CoV-2.** Models trajectories were fit to time series of virus titer (circles-solid lines) and RNA (triangles-dashed lines). Lines represent simulated median of the log-normal probability model and ribbons represent the 95% prediction intervals for H3N2 (golden) and SARS-CoV-2 (red) respectively. (***) denotes the best fitting model denotes the best fitting model to the observed dynamics ($\Delta AICc = 0$)

The supported hypothesis for the two viruses is different. In the case of H3N2, the most parsimonious explanation is the simplest model ("sus-

ceptible cell depletion") which is sufficient to explain the observed viral dynamics (ΔAICc $= 0$). This model has been previously documented to capture the single viral peak that is characteristic of H3N2 infections [79]. More surprising however is that all other models, initialized over a wide range of parameter hyperspace, essentially converged to the best-fitting model. Tables 3.6 and 3.7 present maximum likelihood parameter estimates (MLEs) and their corresponding confidence intervals for each hypothesis. Relative differences in AICc values of the competing models for H3N2 infection are ascribed to the penalties incurred on account of higher parametric degrees of freedom with the more complex dynamical explanations. We estimate a modest degree of intercellular transmissibilty for H3N2 ($R_0 = 1.64\ (1.27, 1.98)$), a high probability of sampling and low variability in both viral RNA ($\iota_P = 1.00\ (0.01, 1)$, $\sigma_P = 0.79\ (0.6, 0.94)$) and viral titer ($\iota_V = 0.66\ (0.26, 0.98)$, $\sigma_V = 0.90\ (0.68, 1.09)$).

Relative to H3N2, the analysis of SARS-CoV-2 data supports the "increased production of refractory cells" hypothesis as the most parsimonious explanation (ΔAICc $= 0$). We estimate a nearly 50-fold higher degree of intercellular transmissibility ($R_0 = 77.30\ (22, 163)$), compared to H3N2. Refractory cells are estimated to be produced at a rate of $\phi = 1.41\ (1.41, 5.28)\ n_T^{-1} d^{-1}$ but also a rapid reversion rate such that the refractory phase lasts for 0.82 (0.72, 1.37)($= \frac{1}{\rho}$) days. Similar to H3N2, we observe a low degree of sampling bias for both data types ($\iota_P = 1\ (0.12, 1)$, $\iota_V = 1\ (0.4, 1)$), however, a much higher degree of variability ($\sigma_P = 1.98\ (1.47, 2.51)$, $\sigma_V = 1.24\ (1.06, 1.64)$) is required to explain the variance in the observed infection dynamics. This, we propose might be a consequence of the high degree of individual-level variation exhibited by SARS-CoV-2 replicates unlike H3N2 (Fig. 5.2).

Therapeutic outcomes under differential antiviral interventions

The best-fitting models for each virus are used to assess the effects of different antiviral therapies (see Eq. (3.42)-(3.49)) on viral control and treatment. The model is fixed at a moderate value of antiviral reactivity ($\beta = 100\ n_I^{-1} d^{-1}$) and duration ($\frac{1}{\delta_A} = 1$ day) and we assume a daily pulsed administration of an antiviral that enhances cytotoxicity by killing

Table 3.6: **Maximum likelihood estimates (MLE)**. Model specific parameter MLE's were obtained by maximizing the likelihood function.

Parameter	Model						
	Susceptible Depletion	Eclipse Phase	Susceptible Replenishment	Refractory Cells Produced	Reduced Viral Replication	Natural Killer Cells	Natural Killer Cells-2
H3N2							
$\Delta AICc$	0.00e+00	7.00e+00	7.00e+00	1.50e+01	1.30e+01	1.50e+01	1.30e+01
R_0	1.64e+00	1.65e+00	1.65e+00	1.65e+00	1.65e+00	1.65e+00	1.65e+00
V_0	1.90e+03	1.72e+03	3.46e+03	1.71e+03	1.54e+03	1.89e+03	2.59e+03
ν_F	-	-	-	9.35e-07	7.46e-07	8.64e-05	5.82e-07
ν_N	-	-	-	-	-	7.52e+04	-
ν_T	-	-	4.53e-08	-	-	-	-
δ_E	-	3.98e+09	-	-	-	-	-
δ_F	-	-	-	6.68e+05	6.57e+05	9.83e+05	6.05e+05
δ_N	-	-	-	-	-	1.97e+04	-
δ_P	3.42e+02	3.69e+01	3.69e+01	3.68e+01	3.69e+01	3.69e+01	3.69e+01
π	-	-	-	-	4.61e-02	-	-
ϕ	-	-	-	1.53e+00	-	-	-
ψ	-	-	-	-	-	7.17e+04	-
ζ	-	-	-	-	-	-	1.26e+00
ρ	-	-	-	4.86e+00	-	-	-
ι_P	1.00e+00	1.00e+00	1.00e+00	1.00e+00	1.00e+00	1.00e+00	1.00e+00
ι_V	6.63e-01	7.91e-01	4.08e-01	8.67e-01	9.11e-01	7.74e-01	5.83e-01
σ_P	7.96e-01	7.96e-01	7.96e-01	7.96e-01	7.96e-01	7.96e-01	7.96e-01
σ_V	9.05e-01	9.13e-01	9.13e-01	9.14e-01	9.14e-01	9.13e-01	9.14e-01
SARS-CoV-2							
$\Delta AICc$	1.00e+01	1.40e+01	1.10e+01	0.00e+00	1.40e+01	1.50e+01	1.40e+01
R_0	1.00e+00	1.03e+00	1.01e+00	7.73e+01	1.01e+00	1.02e+00	1.01e+00
V_0	3.41e+03	4.43e+04	4.23e+03	1.00e+06	3.54e+03	4.62e+04	3.12e+03
ν_F	-	-	-	1.00e-04	4.60e-06	1.62e-05	1.00e-07
ν_N	-	-	-	-	-	1.69e+04	-
ν_T	-	-	3.31e-08	-	-	-	-
δ_E	-	1.25e+09	-	-	-	-	-
δ_F	-	-	-	3.21e+00	1.92e+04	5.62e+04	1.01e+05
δ_N	-	-	-	-	-	2.40e+03	-
δ_P	1.24e+00	8.94e+01	1.24e+00	2.70e+03	1.00e+00	1.25e+00	1.00e+00
π	-	-	-	-	9.03e-01	-	-
ϕ	-	-	-	1.41e+00	-	-	-
ψ	-	-	-	-	-	8.81e+04	-
ζ	-	-	-	-	-	-	1.00e+05
ρ	-	-	-	1.22e+00	-	-	-
ι_P	1.99e-05	4.82e-04	3.74e-05	1.00e+00	4.38e-05	7.56e-06	4.76e-05
ι_V	9.98e-01	9.96e-01	9.99e-01	1.00e+00	9.91e-01	9.90e-01	9.36e-01
σ_P	1.63e+00	1.66e+00	1.63e+00	1.98e+00	1.66e+00	1.63e+00	1.66e+00
σ_V	2.02e+00	2.09e+00	2.02e+00	1.24e+00	2.07e+00	2.02e+00	2.05e+00

the infected cells under antiviral influence(Fig. 3.4). The effectiveness of model outcomes is ascertained using Eq. 3.50. We then qualitatively demonstrate the relationship between the timing of the initiation of therapy and the efficacy of the antiviral, defined as the increase in the cellular toxicity (η). We observe that a smaller value of η is equally effective provided the intervention is initiated in the early phase of the infection. Owing to the higher values of within-host transmissiblilty ($R_0 = 77.3$) and cytotoxicty of cells infected with SARS-CoV-2 ($\delta_I = 44\ d^{-1}$), we observe that generally a more efficacious antiviral ($\eta \geq 0.5$) is required to produce an effect comparable in the therapeutic outcome in treatment of H3N2 ($R_0 = 1.64,\ \delta_I = 4\ d^{-1}$).

Surprisingly, when comparing between modes of antiviral action (as represented by the parameters ω, η and α), we find little noticeable difference in efficacy (Fig. 3.8). To explore this, we analytically derive an antiviral impact factor using the next generation operator [47]. For the full antivi-

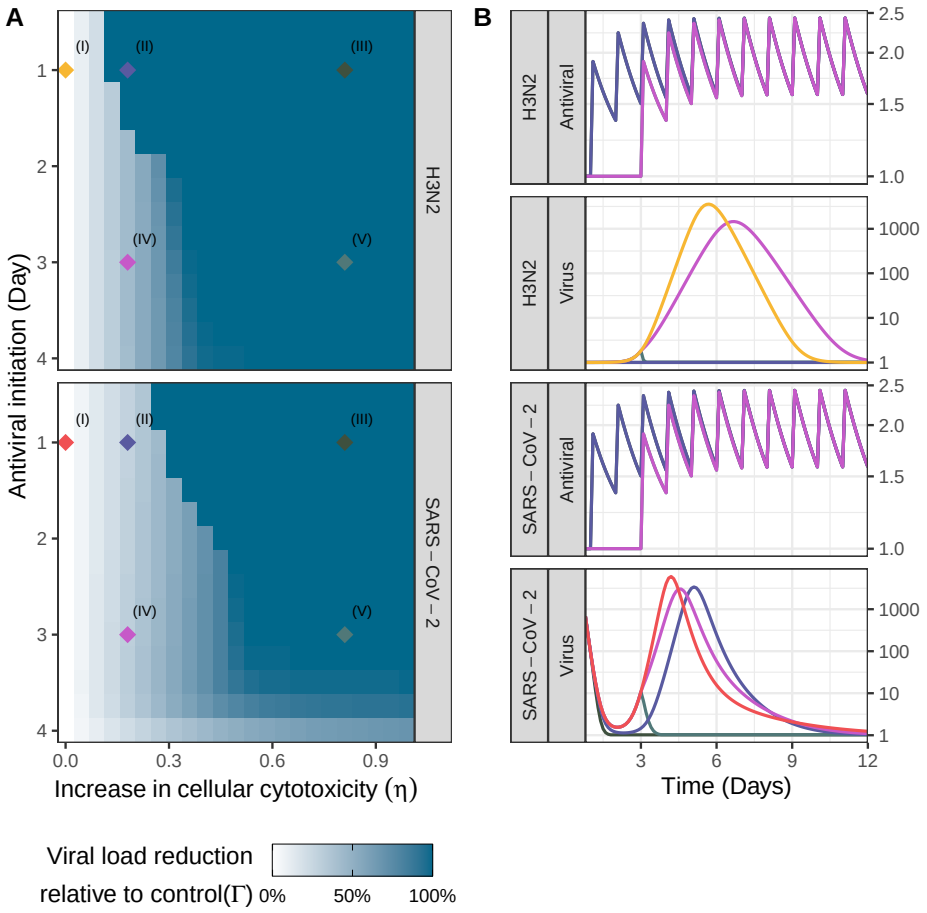

Figure 3.4: **Relationship between effectiveness and initiation time for antiviral enhancing cellular toxicity on the viral dynamics of H3N2 and SARS-CoV-2.** (A) Effectiveness of the antiviral therapy expressed as the reduction in of the area under the viral curve relative to the control region. (B) Simulated viral and anti-viral dynamics for regions (I-V) in (A). Control trajectories (region I, $\eta = 0$) of virus dynamics of H3N2 (golden) and SARS-CoV-2 (red) were generated using the parameter MLEs of best fitting models. For treatment (regions II-V), we assume a daily pulse of unit antiviral drug ($Amax = 1$) with moderate antiviral binding rate ($\beta_A = 100\, n_A^{-1} d^{-1}$) and duration ($\frac{1}{\delta} = 1$ day).

ral model described in Eq.3.42-3.49, we derive the within-host effective reproductive number (R_A, for details on derivation see Supplementary Information Ch.2),

$$R_A = R_{eff}\Big(1 - p_A(\alpha(1-\eta)+\eta) + f_0^A(1-\alpha)(1-\eta)(1-\omega)\Big). \tag{3.51}$$

Here, $R_{eff} = R_0 f_0$ is the effective reproductive number at the time of the antiviral administration. $p_A = \frac{\beta_A}{(\beta_A+\delta_I)}$, is the probability of getting bound by the antiviral; $f_0^A = \frac{T_0^A}{T_0}$, is the fraction of target cells bound by the antiviral (T_0^A) to the susceptible target cells (T_0). This expression (Eq. 3.51) calculates the average number of infected cells produced by a single infected cell in the instant following administration of the antiviral.

We define antiviral impact (ξ) as the reduction in the within host transmissibilty of the virus in the presence of an antiviral (R_A) as compared to a baseline (R_{eff}). Then from Eq. 3.51 we get:

$$\xi = \left(1 - \frac{R_A}{R_{eff}}\right) = p_A(\alpha(1-\eta)+\eta) - f_0^A\big((1-\alpha)(1-\eta)(1-\omega)\big). \tag{3.52}$$

From this equation, we observe that the antiviral impact depends not only on the efficacy of the drug *i.e.* the absolute values of α, η, and ω, but also is a product of both p_A (within-host reactivity and duration of the antiviral) and f_0^A (a proxy for the dependence of antiviral impact on the timing of the antiviral dose). For this simple case, we find that the effect of the modes of antiviral action appear multiplicatively and cannot be separated. This, we submit, is the main reason for the lack of qualitative differences among the modes of antiviral action reported in Fig. 3.8.

3.4 Discussion

In this study, we explore the potential role of different mechanisms in explaining the observed within-host viral dynamics of influenza A H3N2 and SARS-CoV-2 infections in ferrets. By formulating a suite of hypotheses

expressed as mathematical models, we use likelihood-based inference to arbitrate among these competing mechanisms. We find that the viral kinetics of H3N2 are most parsimoniously explained by the simplest model assuming only susceptible depletion. In contrast, the SARS-CoV-2 data support a more complex explanation involving the generation of refractory susceptible target cells stimulated by type I and III interferons. Further, we find that SARS-CoV-2 has a much higher rate of within-host proliferation (as summarized by R_0) than H3N2. This striking difference may, we propose, be a potential explanation for the different conclusion reached regarding the putative association with the host immune system. Given the inconsistent and stuttering initial phase of the viral curve for SARS-CoV-2 (Figs. 3.2 & 5.2), the model's preferred explanation is an initially large population of refractory cells that eventually reactivate susceptibility as the cytokine production declines due to reduced infectious cells. This mechanism introduces a brief delay in the dynamics with the infection of these newly susceptible target cells eventually contributing to viral proliferation. Previously, a similar finding was reported to explain the bi-modality and the prolonged plateau in peaks in observed equine influenza A (H3N8) [178] and human influenza A (H1N1) dynamics [12].

In contrast, the lower R_0 value for influenza stimulates a weaker interferon response. Interestingly, a lower proliferation rate for H3N2 thus facilitates rapid viral spread with termination of infection brought about by susceptible depletion. This conclusion is consistent with studies of viral dynamics at during the early stages of an infection in humans[37, 85]. We note, however, that our conclusion does not preclude the involvement of the host immune system in modulating H3N2 infection. We speculate that the dynamical simplicity of these data may affect the potential for identification of more complex mechanisms. Resolution of this issue will require the incorporation of additional parallel immunological signals in such analyses [131, 145].

Plaque-based assays (PBA) generate statistically accurate estimates of cellular infection burden by quantifying viral titers through the number of viable virions [126, 211]. However, with the manual assessment of plaque-forming units (PFU) often associated with PBA, data generated through

these methods may be noisy. A statistically precise measure of within-host viral burden, on the other hand, can be obtained using sampling methods based on quantitative polymerase chain reaction (qPCR) [147]. These methods generate data that can be orders of magnitude higher than the viable virus titer (*e.g.* For foot-and-mouth disease, the PBA:qPCR ratio is $\sim 1:1000$ [25]). Such discrepancies in the estimated viral titer can lead to an over-estimate of the proliferation potential of the virus or make comparison of different studies challenging. For this reason, our models simultaneously account for both qPCR and viral titer data.

To examine the potential control of these viruses with antivirals, we formulate models that permitted different modes of antiviral action. Antivirals may serve to: reduce susceptibility of target cells, lower viral production by infectious cells and increase infectious cell mortality. We find no qualitative difference among the three modes of action (see Fig. 3.8). A similar result has been discussed in while devising antiviral therapies for closely related coronaviruses [99]. Our theoretical analysis explained how antiviral effectiveness is determined by a product of the three individual antiviral modes of action. Hence, their mono-therapeutic effects are qualitatively indistinguishable, as we show in Eq. 3.52.

In this study we have attempted to understand the within-host interactions between invading viral pathogens and the host's innate immune system. Further, we qualitatively describe the differential action of antiviral therapies by leveraging on the pre-existing immune interaction for H3N2 and SARS-CoV-2 infections.

3.5 Supplementary Results

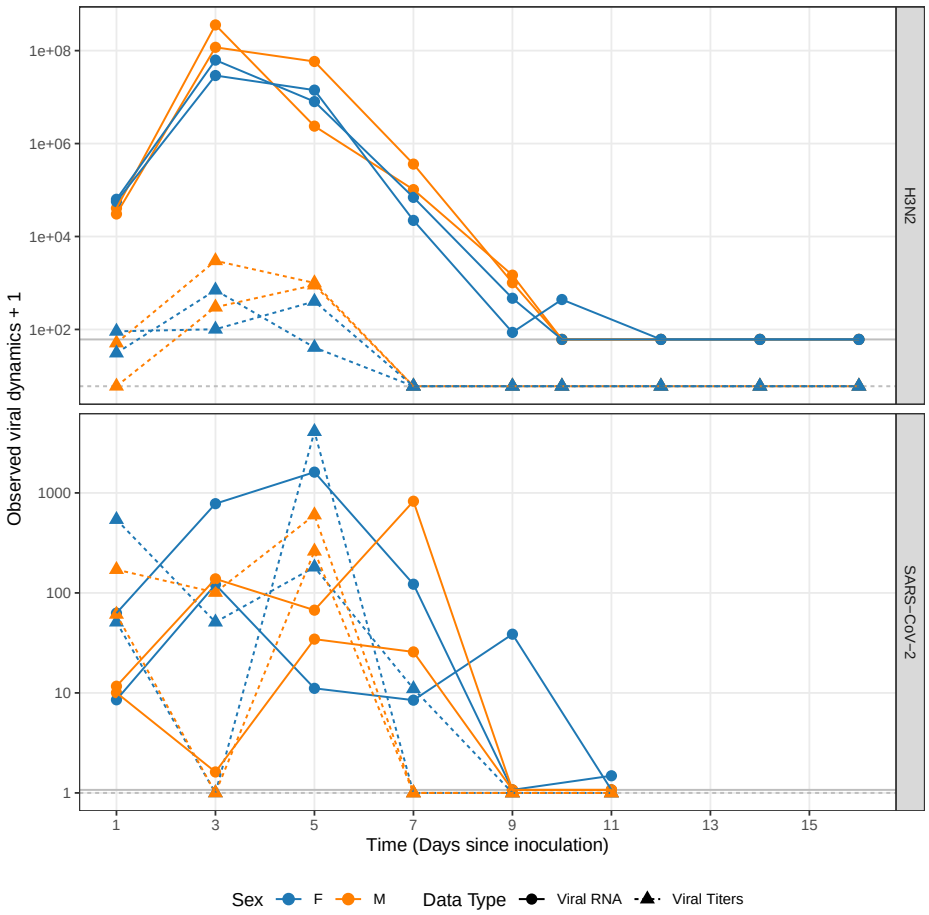

Figure 3.5: **Experimental data from ferret infection experiments stratified by the sex of the animal.** Facets represent within-host viral dynamics of H3N2 influenza (top) and SARS-CoV-2 (bottom). Viral RNA data is represented using points and solid lines; Viral titer data is represented using triangles and dotdashed lines. Different detection threshold (dt) against virus-data type combinations are represented using solid and dotdashed lines for viral RNA and titer respectively. X-axis represents time in days since ferrets were inoculated.

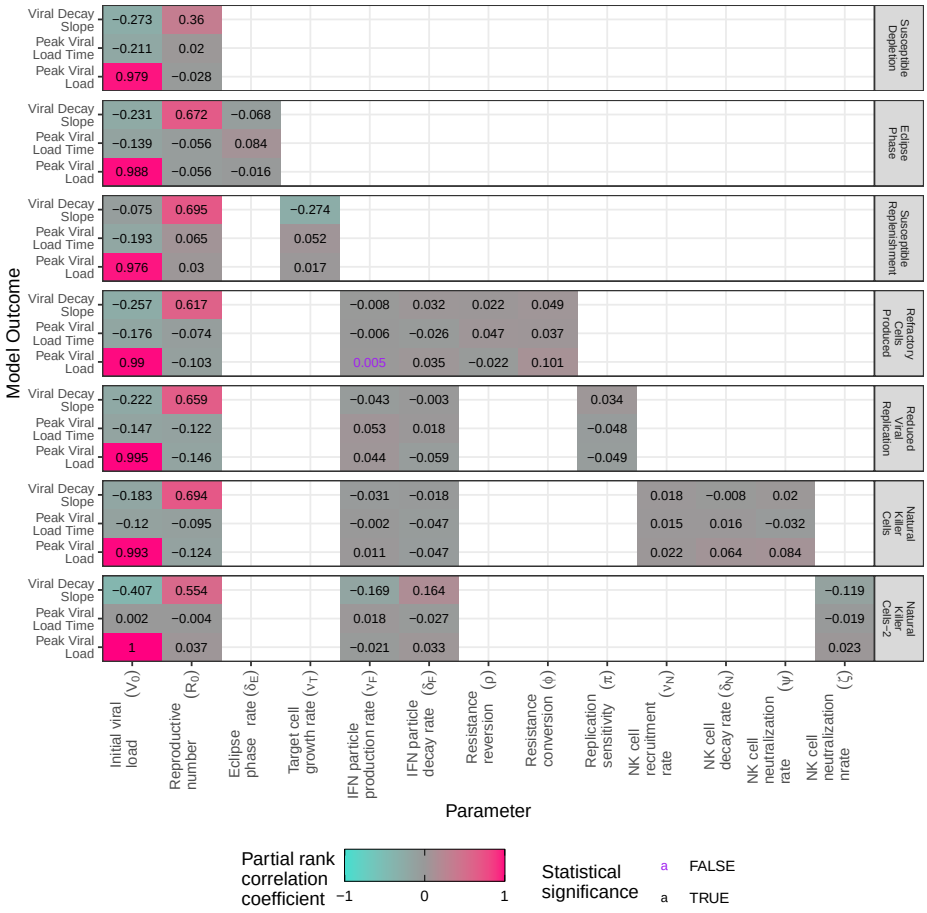

Figure 3.6: **Sensitivity analysis with hypothesis specific model parameters.** Partial rank correlation coefficients (PCC, color) were calculated to assess sensitivity of parameters in generating model outcomes. 5000 parameter configurations were generated using a Sobol sampling scheme. Bootstrapped confidence intervals were obtained by uniformly sampling with replacement and generating 150 samples. Statistical significance was ascertained at 5%, Bonferroni corrected, level of sigificance.

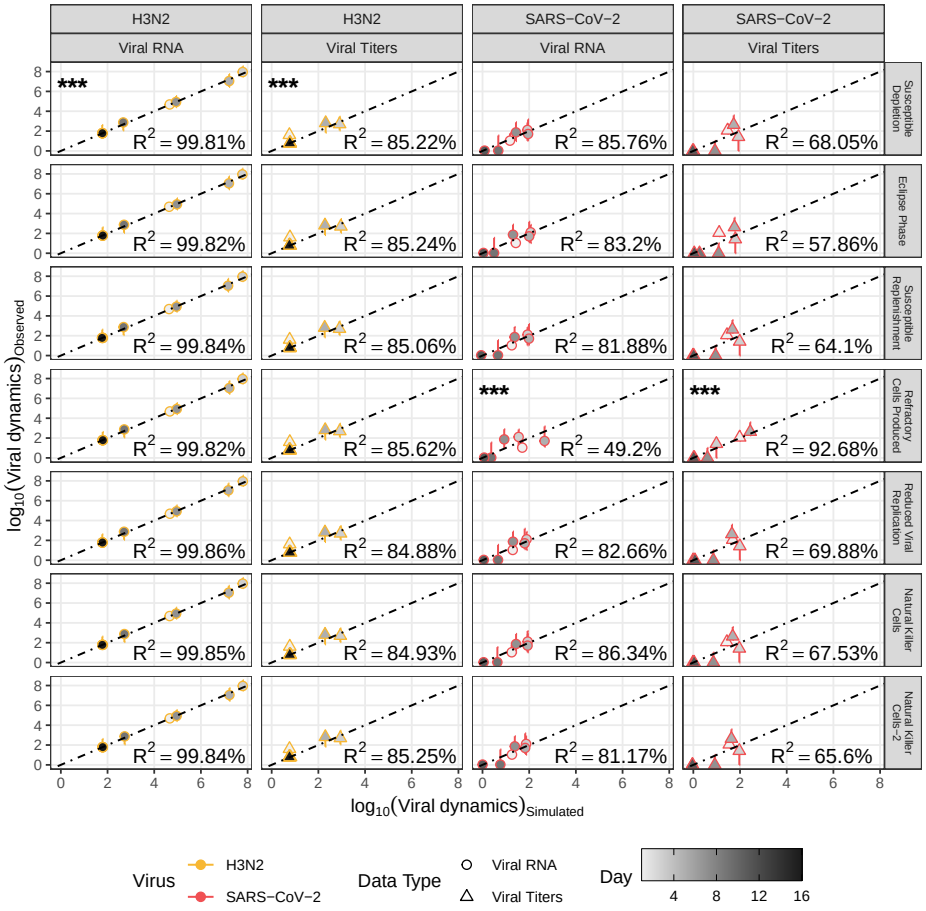

Figure 3.7: **Goodness of fit for virus-data-model combination.** Scatter plots show model-data agreement for viral RNA (empty circles) and viral titers (empty triangles) on a log_{10} scale. Error-bars represent individual level variation among replicates. Fill gradient represents time (in days) to demarcate the observed infection interval. Dotdashed line is the 1:1 reference line. For every model, coeficient of variation (R^2, reported inset) was calulated for each data-type at the medium of the replicate ditribution to quantify the proportion of variation explained by the models. (***) represents the best fitting model based on the AICc comparison.

Table 3.7: **Maximum likelihood estimates (MLE) and 95% confidence intervals**. Model specific parameter MLE's were obtained by maximizing the likelihood function. Uncertainty in the MLE was ascertained using parametric bootstrap.

	Model						
Parameter	Susceptible Depletion	Eclipse Phase	Susceptible Replenishment	Refractory Cells Produced	Reduced Viral Replication	Natural Killer Cells	Natural Killer Cells-2
				H3N2			
$\Delta AICc$	0.00e+00	7.00e+00	7.00e+00	1.50e+01	1.30e+01	1.50e+01	1.30e+01
R_0	1.64e+00 (1.27e+00, 1.98e+00)	1.65e+00 (1.20e+00, 1.93e+00)	1.65e+00 (1.28e+00, 1.87e+00)	1.65e+00 (1.22e+00, 2.05e+00)	1.65e+00 (1.19e+00, 1.97e+00)	1.65e+00 (9.54e-04, 1.65e+00)	1.65e+00 (1.26e+00, 1.96e+00)
V_0	1.90e+03 (7.91e+02, 1.03e+04)	1.72e+03 (5.06e+02, 7.11e+03)	3.46e+03 (7.69e+02, 7.48e+03)	1.71e+03 (1.86e+02, 5.66e+03)	1.54e+03 (4.96e+02, 5.99e+03)	1.89e+03 (1.89e+03, 3.92e+05)	2.59e+03 (9.45e+02, 5.58e+03)
ν_F	-	-	-	9.35e-07 (2.22e-07, 9.28e-05)	7.46e-07 (2.78e-07, 9.67e-05)	8.64e-05 (3.53e-05, 8.64e-05)	5.82e-07 (1.82e-07, 6.54e-03)
ν_N	-	-	-	-	-	7.52e+04 (7.40e+04, 7.52e+04)	-
ν_T	-	-	4.53e-08 (4.53e-08, 4.34e-02)	-	-	-	-
δ_E	-	3.98e+09 (2.45e+08, 9.80e+09)	-	-	-	-	-
δ_F	-	-	-	6.68e+05 (1.07e+00, 9.72e+05)	6.57e+05 (5.24e+03, 9.82e+05)	9.83e+05 (5.73e+05, 9.83e+05)	6.05e+05 (2.67e+05, 9.61e+05)
δ_N	-	-	-	-	-	1.97e+04 (1.80e+04, 1.97e+04)	-
δ_P	3.42e+02 (3.30e+00, 6.28e+02)	3.69e+01 (3.33e+00, 7.33e+01)	3.69e+01 (4.09e+00, 5.21e+01)	3.68e+01 (3.52e+00, 5.71e+01)	3.69e+01 (3.53e+00, 6.84e+01)	3.69e+01 (3.69e+01, 5.44e+05)	3.69e+01 (3.38e+00, 6.59e+01)
π	-	-	-	-	4.61e-02 (8.88e-03, 9.77e-01)	-	-
ϕ	-	-	-	1.53e+00 (1.15e-01, 9.40e+01)	-	-	-
ψ	-	-	-	-	-	7.17e+04 (7.17e+04, 7.35e+04)	-
ζ	-	-	-	-	-	-	1.26e+00 (2.20e-01, 1.33e+03)
ρ	-	-	-	4.86e+00 (8.10e-02, 4.70e+01)	-	-	-
ι_P	1.00e+00 (1.31e-02, 1.00e+00)	1.00e+00 (1.74e-01, 1.00e+00)	1.00e+00 (2.08e-01, 1.00e+00)	1.00e+00 (1.00e-01, 1.00e+00)	1.00e+00 (6.97e-02, 1.00e+00)	1.00e+00 (1.68e-01, 1.00e+00)	1.00e+00 (1.66e-01, 1.00e+00)
ι_V	6.63e-01 (2.58e-01, 9.77e-01)	7.91e-01 (4.40e-01, 9.94e-01)	4.08e-01 (3.99e-01, 9.79e-01)	8.67e-01 (3.77e-01, 9.92e-01)	9.11e-01 (5.09e-01, 9.91e-01)	7.74e-01 (3.65e-02, 7.74e-01)	5.83e-01 (3.02e-01, 9.92e-01)
σ_P	7.96e-01 (6.00e-01, 9.49e-01)	7.96e-01 (5.95e-01, 9.41e-01)	7.96e-01 (6.04e-01, 8.45e-01)	7.96e-01 (6.02e-01, 9.22e-01)	7.96e-01 (5.94e-01, 9.32e-01)	7.96e-01 (7.96e-01, 2.73e+00)	7.96e-01 (6.14e-01, 9.37e-01)
σ_V	9.05e-01 (6.76e-01, 1.09e+00)	9.13e-01 (7.06e-01, 1.08e+00)	9.13e-01 (7.82e-01, 1.00e+00)	9.14e-01 (6.58e-01, 1.08e+00)	9.14e-01 (7.16e-01, 1.08e+00)	9.13e-01 (9.13e-01, 2.31e+00)	9.14e-01 (6.86e-01, 1.11e+00)
				SARS-CoV-2			
$\Delta AICc$	1.00e+01	1.40e+01	1.10e+01	0.00e+00	1.40e+01	1.50e+01	1.40e+01
R_0	1.00e+00 (9.93e-01, 2.52e+00)	1.03e+00 (1.00e+00, 1.89e+01)	1.01e+00 (9.93e-01, 6.71e+00)	7.73e+01 (2.20e+00, 1.63e+02)	1.01e+00 (9.87e-01, 2.70e+00)	1.02e+00 (5.38e-03, 1.02e+00)	1.01e+00 (1.01e+00, 2.25e+01)
V_0	3.41e+03 (3.34e+03, 9.73e+05)	4.43e+04 (5.18e+01, 9.62e+05)	4.23e+03 (4.23e+03, 9.87e+05)	1.00e+06 (7.91e+03, 1.00e+06)	3.54e+03 (8.16e+02, 9.64e+05)	4.62e+04 (4.62e+04, 6.32e+05)	3.12e+03 (2.49e+02, 8.62e+05)
ν_F	-	-	-	1.00e-04 (3.16e-05, 1.00e-04)	4.60e-06 (8.69e-07, 9.88e-05)	1.62e-05 (1.62e-05, 6.48e-05)	1.00e-07 (1.00e-07, 8.68e-03)
ν_N	-	-	-	-	-	1.69e+04 (1.69e+04, 8.88e+04)	-
ν_T	-	-	3.31e-08 (3.31e-08, 7.37e+00)	-	-	-	-
δ_E	-	1.25e+09 (6.84e+08, 9.48e+09)	-	-	-	-	-
δ_F	-	-	-	3.21e+00 (3.21e+00, 5.57e+03)	1.92e+04 (4.23e+01, 9.79e+05)	5.62e+04 (5.62e+04, 7.48e+05)	1.01e+05 (1.01e+05, 9.56e+05)
δ_N	-	-	-	-	-	2.40e+03 (2.40e+03, 7.48e+04)	-
δ_P	1.24e+00 (1.24e+00, 9.64e+05)	8.94e+01 (8.94e+01, 9.89e+05)	1.24e+00 (1.24e+00, 9.46e+05)	2.70e+03 (2.70e+03, 8.29e+05)	1.00e+00 (1.00e+00, 9.43e+05)	1.25e+00 (1.25e+00, 3.86e+05)	1.00e+00 (1.00e+00, 9.52e+05)
π	-	-	-	-	9.03e-01 (6.18e-03, 9.87e-01)	-	-
ϕ	-	-	-	1.41e+00 (1.41e+00, 5.28e+01)	-	-	-
ψ	-	-	-	-	-	8.81e+04 (8.81e+04, 9.13e+04)	-
ζ	-	-	-	-	-	-	1.00e+05 (4.47e+01, 1.00e+05)
ρ	-	-	-	1.22e+00 (7.28e-01, 1.38e+00)	-	-	-
ι_P	1.99e-05 (1.99e-05, 9.25e-01)	4.82e-04 (4.82e-04, 9.56e-01)	3.74e-05 (3.74e-05, 9.32e-01)	1.00e+00 (1.12e-01, 1.00e+00)	4.38e-05 (4.38e-05, 9.61e-01)	7.56e-06 (7.56e-06, 7.09e-02)	4.76e-05 (4.76e-05, 9.72e-01)
ι_V	9.98e-01 (1.08e-01, 9.98e-01)	9.96e-01 (1.35e-02, 9.99e-01)	9.99e-01 (9.65e-03, 9.99e-01)	1.00e+00 (3.94e-01, 1.00e+00)	9.91e-01 (2.84e-01, 9.99e-01)	9.90e-01 (7.85e-01, 9.90e-01)	9.36e-01 (1.05e-01, 9.96e-01)
σ_P	1.63e+00 (1.11e+00, 2.03e+00)	1.66e+00 (1.25e+00, 2.02e+00)	1.63e+00 (1.15e+00, 2.01e+00)	1.98e+00 (1.47e+00, 2.51e+00)	1.66e+00 (1.12e+00, 2.11e+00)	1.63e+00 (1.63e+00, 1.69e+00)	1.66e+00 (1.16e+00, 2.04e+00)
σ_V	2.02e+00 (1.29e+00, 2.44e+00)	2.09e+00 (1.37e+00, 2.56e+00)	2.02e+00 (1.33e+00, 2.61e+00)	1.24e+00 (1.06e+00, 1.64e+00)	2.07e+00 (1.31e+00, 2.43e+00)	2.02e+00 (2.02e+00, 2.34e+00)	2.05e+00 (1.41e+00, 2.50e+00)

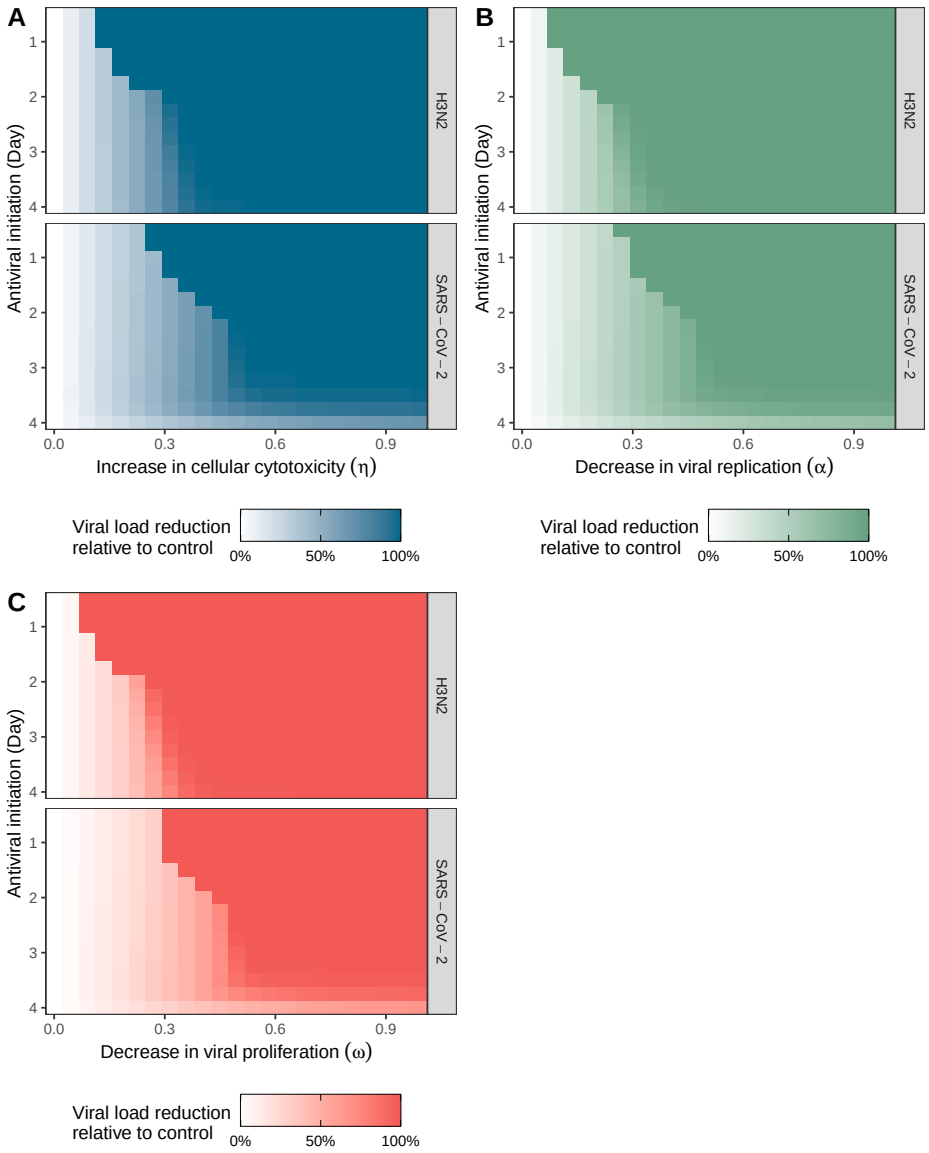

Figure 3.8: **Simulation study reflecting the relative action of antiviral effectiveness, initiation time and type on the viral dynamics of H3N2 and SARS-CoV-2 with a pulse dose administrative strategy.** (A, C) Effectiveness of the antiviral therapy expressed as the reduction in of the area under the viral curve relative to the control region. (B, D) Viral and anti-viral dynamics simulated for regions (1-5). Control trajectories, region (1), for the two viruses were generated using the parameter MLEs of two best fitting models and simulating dynamics for 12 days. For the remainder of regions, a moderate antiviral binding rate ($\beta_A = 100\, n_A^{-1} d^{-1}$) and duration ($\frac{1}{\delta} = 1$ day) were assumed during this study.

CHAPTER 4

UNRAVELING MECHANISMS DRIVING ANOMALOUS INFLUENZA B SEASONALITY IN MASSACHUSSETTS

4.1 Introduction

Historically[1], studies of seasonal influenza epidemics have primarily focused on influenza A viruses (IAVs) rather than influenza B viruses (IBVs) and, in particular, on the scientifically interesting and practically important phenomenon of antigenic evolution of the hemagglutinin (HA) glycoprotein in A/H3N2 viruses (see for example: [16, 108, 117, 159, 189]). The need to understand and predict population immunity to evolving seasonal IAVs has brought attention to the importance of individual-level histories of influenza virus infections [70, 149]. The sequence of IAV infections an individual experiences shapes their immune response to future IAV exposures. Thus, much attention has focused on understanding this immunological complexity and predicting the perpetual changes in the antigenic and spatial structuring of circulating IAVs. However, this emphasis on the ecology and evolution of IAVs has come at the expense of that of IBVs, despite

[1] The analysis presented in this chapter forms a crucial subset of the article *Anomalous influenza seasonality in the United States and the emergence of novel influenza B viruses* [22]. I was responsible for the writing of the article, conceptualizing and fitting dynamical models to the epidemiological time-series from Massachusetts. I have conducted parameter estimation and uncertainty quantification. I was also responsible for conducting the simulation study that is presented in following chapter.

awareness that these viruses contribute to the global burden of influenza morbidity and mortality [153, 195].

Consequently, the population biology of IBVs has, until recently [24, 118], received less attention, despite dramatic evolutionary changes over the past decade. Influenza B viruses first diverged into two lineages (Yamagata and Victoria) 30-40 years ago [51, 94, 173]. Recently, both lineages have exhibited higher evolutionary rates and several selective sweeps, but with different mechanisms of evolutionary change [201, 202]. Nucleotide deletions in the B/Victoria HA gene segment have been identified and shown to characterize divergent, co-circulating subclades. These nucleotide deletions are accompanied by epistatic mutations in other B/Victoria gene segments and interclade reassortment. In contrast, antigenic drift of the neuraminidase (NA) gene segment has been shown to be a main driver in recent B/Yamagata epidemic activity [202]. These evolutionary changes highlight that IBVs are capable of multiple adaptation strategies, yet we lack an understanding of their epidemiological consequences.

Here, we analyze the unusual 2019/2020 influenza season, hereafter referred to as the 2019 season (independent of the later SARS-CoV-2 introduction) using type specific time-series of reported cases of influenza A and B from Massachusetts. We use a combination of strategic transmission models and likelihood-based statistical inference to challenge alternative hypotheses to explain the observed changes in seasonal influenza dynamics. Ultimately, our modeling leads us to conclude that potential within host competition between IAVs and IBVs likely operate over too short a time scale to account for the observed shifts in influenza seasonality. Rather, we find that increased effective transmission of novel IBVs provides a parsimonious explanation for the early timing and relative dominance of these viruses in 2019.

4.2 Methods

4.2.1 Epidemiological data collation and sources

We downloaded weekly state-level subtype-specific influenza sample data from CDC FluView Interactive ILI and Viral Surveillance [61]. Starting in the 2016/2017 season, data from public health labs and clinical laboratories are presented separately (see [198] for additional information). We aggregated positive influenza A sample counts for the cases when subtyping was not performed or it was not possible to subtype the sample, into a single "no subtype" category. Finally, we converted epidemiological weeks (Epiweeks) into calendar dates. Our analysis incorporates data for the state of Massachusetts that reported sample counts in a particular week for both Type A and Type B (note that a count of zero samples is considered a reported count). We used the standard influenza season specification for the United States. The season starts with Epiweek 40 and ends with Epiweek 39 of the following calendar year.

4.2.2 Influenza transmission model

To explore the potential role of cross-immunity in determining the timing and magnitude of influenza A vs. B peaks, we developed a two-strain Susceptible Infectious Cross-protected Recovered Susceptible (SICRS) model (see Fig. 4.5; [95]). The population (of size N) is divided into 10 compartments dependent on infection and immune status. Susceptible (S) individuals have no immunity and can be infected with either type at rate λ_n, where λ_n is the force of infection of Type n ($n \in \{A, B\}$). Upon infection individuals move to the corresponding I_n compartment. Following the infectious period (with mean duration $1/\gamma_n$) individuals recover and move to the cross-protected compartment, C_n. After recovering, individuals initially have perfect immunity against reinfection with the same type, and partial cross-protection against infection with the other type. Cross-protection against Type m due to prior infection with n corresponds to a reduction in the infection rate, $(1-\chi_{mn})\lambda_m$. The cross-protection strength, χ_{mn}, takes values between 0 (no cross-protection) and 1 (perfect cross-protection). Cross-protection wanes with rate ϕ_n. Individuals who

have lost cross-protection against Type m but retain type-specific immunity move to R_n. Type-specific immunity also wanes, with rate w_n. If individuals experience an infection with the second type, m, before their type-specific immunity wanes they become infectious with Type m and move to the I_{mn} compartment. Upon recovery they move to the final model compartment, R, which corresponds to individuals with type-specific immunity to both types. Tables 4.3 and 4.1 summarize the model compartments and parameters respectively.

The force of infection for Type n depends on the number of individuals infectious with Type n and the rate cases are imported to the population from external sources (proportional to η_n),

$$\lambda_n(t) = \beta_n(t)(I_n + I_{nm} + \eta_n)/N. \tag{4.1}$$

Our model allows for seasonality in the transmission rate,

$$\beta_n(t) = \gamma_n R_0^{(n)} \left[1 + b_n \cos\left((2\pi t - t_0^n)/T\right)\right], \tag{4.2}$$

where $R_0^{(n)}$, b_n and t_0^n are respectively the basic reproductive number, amplitude of seasonality and peak day for Type n and T is the period seasonality (1 year).

The model dynamics are captured by a system of ordinary differential equations,

$$\frac{dS}{dt} = \sum_n \left\{-\lambda_n S + w_n R_n\right\}, \tag{4.3}$$

$$\frac{dI_n}{dt} = \lambda_n S - \gamma_n I_n, \tag{4.4}$$

$$\frac{dC_n}{dt} = -(1-\chi_{mn})\lambda_m C_n + \gamma_n I_n - \phi_n C_n, \tag{4.5}$$

$$\frac{dR_n}{dt} = -\lambda_m R_n + \phi_n C_n + w_m R - w_n R_n, \tag{4.6}$$

$$\frac{dI_{mn}}{dt} = (1-\chi_{mn})\lambda_m C_n + \lambda_m R_n - \gamma_m I_{mn}, \tag{4.7}$$

$$\frac{dR}{dt} = \sum_n \left\{\gamma_n I_{nm} - w_n R\right\}, \tag{4.8}$$

for $n \in \{A,B\}$. If $n = A$ then $m = B$ and vice versa.

For the likelihood-based inference, we generated the cumulative number of cases according to

$$\frac{dK_A}{dt} = \gamma_A(I_{AB} + I_A), \tag{4.9}$$

$$\frac{dK_B}{dt} = \gamma_B(I_{BA} + I_B). \tag{4.10}$$

For formal comparison with data, the weekly counts of new A and B cases were calculated, $Q_n(t) = K_n(t) - K_n(t-\delta)$ where $\delta = 1$ week. The deterministic model was implemented in R using the package 'pomp' [102]. Numerical simulations were performed using the resulting pomp object. Simulations were initialized at the endemic equilibrium of the non-interacting two-type model ($\chi_{BA} = \chi_{AB} = 0$) without seasonality ($b_A = b_B = 0$), which possesses an analytical solution [95]. The first 100 years were discarded as transient dynamics, before the subsequent 3 years were used in likelihood calculation.

Table 4.1: Parameters for Strategic Model

Symbol	Value	Definition
N	6.7×10^6	Total population
$R_0^{(n)}$	1–5	Reproductive number of Type n ($n \in \{A,B\}$)
$1/\gamma_A$	2.5 days [32]	Infectious period of Type A
$1/\gamma_B$	3.4 days [32]	Infectious period of Type B
$1/\phi_n$	1 month	Cross-protection duration after Type n infection
χ_{mn}	0–1	Cross-protection against m after infection with n
η_n	1 per day	Importation rate of Type n
$1/w_A$	4 years [189]	Duration of Type A-specific immunity
$1/w_B$	4, 10 years	Duration of Type B-specific immunity
b_A	0–1	Amplitude of seasonality for Type A
b_B	0–1	Amplitude of seasonality for Type B
t_0^A	day 40	Timing of seasonal peak for Type A
t_0^B	day 50	Timing of seasonal peak for Type B

4.2.3 Model fitting and comparison

Using type-specific data for Massachusetts [61] influenza seasons 2016-2018, we performed maximum-likelihood estimation of the parameters of our two-strain model to investigate whether there was support for a significant strength of cross-protection from viruses belonging to the other Type (A vs. B). We fitted two versions of our model, a non-interacting model in which the cross-protection parameters (χ_{AB} and χ_{BA}) were fixed

at zero and the full model in which these parameters were allowed to vary. We then performed a formal model comparison of these models (described below).

Maximum-likelihood estimates (MLEs) for the unknown parameters were found using trajectory matching. Weekly case counts were assumed to be subject to Poisson-distributed reporting error. The likelihood function for the data $\{x\}_{t=0}^{T}$ given model parameters Θ was

$$\ell(\Theta) = \prod_{n\in\{A,B\}} \prod_{t=0}^{T} P(x_{n,t}|\rho Q_n(t;\Theta)), \tag{4.11}$$

where ρ is the reporting probability, $Q_n(t;\Theta)$ is the simulated number of new cases at time t using parameters Θ and $P(\cdot|\lambda)$ is the Poisson probability mass function with parameter λ. Note that we estimated separate reporting probabilities for influenza Types A and B (ρ_A and ρ_B, respectively).

We employed a numerical optimizer routine for two tasks: i) explore the parameter space and find the MLEs for all unknown parameters and ii) verify whether the inclusion of cross-protection improved model fits to confirmed sample data. Trajectory matching was performed using the R package 'pomp' which allows for the construction and optimization of likelihood based objective functions [102]. The optimization routine was as follows:

1. A population of 500 parameter configurations were used to initialize a stochastic optimizer that implemented a hybrid-genetic algorithm to maximize the objective function. Guesses were generated using a Latin Hypercube (Sobol) design to ensure efficient sampling of the parameter space [190, 191].

2. The genetic algorithm routine was implemented using R package 'GA' [183]. This technique evolves the solution until a configuration that maximizes the objective function is found. Probability of random mutations was set to a low value (0.2) while the probability of crossover was set to a high value (0.8). Force of selection was assumed to be linear in fitness of the solution in the objective space.

3. At every significant increase in the objective, an increase in the fitness value by 10%, the solution was pushed through a deterministic optimizer method (L-BFGS) [128] to rapidly arrive at intermediate local maxima. We used a hybrid version of the optimizer to decrease convergence times to the the global maximum for each of the two models [184].

4. An exhaustive population of evolving solutions provided high initial fitness diversity and a convergence criterion of 500 evolutionary iterations ensured optimizer's convergence to a global maximum in the objective space [54].

We used Akaike Information Criterion (AIC) to assess the relative goodness of fit of the two hypotheses - non-interacting neutral model and a full two strain model. At 5% level of significance, an AIC value of 2 was taken to be a significant difference in the comparative agreements of the two models. Parameter estimates as well as their projected type-specific cases were used in this assessment. For the 2016-2018 era, models were allowed to reach equilibrium (burn-in period of 100 years) and subsequent 3 years were used in the fitting procedure. Parameter estimates corresponding to the best fitting model, Neutral, from the initial era were used to infer initial conditions of state variables for the 2019 season. This model was re-fit to estimate type-specific reproductive ratios for this season.

4.2.4 Estimation of Uncertainty around Parameter MLE

The method of parametric bootstrap was used to estimate 95% confidence intervals around the parameter MLE obtained in the previous step. This involved:

1. For each model in the two eras, MLEs obtained in the preceding section were used to simulate 1000 synthetic time series.

2. Starting at the MLE, a local search with an L-BFGS optimizer optimizer was conducted and parameter re-estimated for every time series.

3. This results in a bootstrapped distribution of parameter estimates around the MLE. Estimate sets corresponding to 2.5% and 97.5% percentiles of this distribution were taken to be as the 95% confidence bounds of the model MLEs

4.3 Results

In the 2019 season, influenza-like illness (ILI) reports rose in August (Fig. 4.2) exhibiting an earlier seasonal influenza take-off than any season since the 2009 H1N1 pandemic [44]. In temperate countries, influenza A epidemics tend to be larger and occur before influenza B epidemics [24]. The U.S. 2019 season, however, was characterized by an atypically large and early epidemic of influenza B, with the most severe cases observed in children [148].

4.3.1 Putative mechanisms driving atypical dynamics of influenza B in Massachusetts

Here we develop transmission models to explore distinct hypotheses for the epidemiological changes we have documented: one related to viral transmissibility and the other determined by within-host competitive mechanisms [42]. The recent global spread of novel B/Victoria clades and attendant increases in their diversity has raised the possibility that the HA amino acid deletions have resulted in higher transmission efficiency of these viruses [202]. Alternatively, changing epidemiology may be the result of individual-level competitive mechanisms, such as viral interference [42, 55], whereby an individual's innate immune response temporarily hinders the infection ability of other viruses [35]. Viral interference has been noted between influenza and non-influenza viruses [35, 106, 146] and has been suggested to play a role in the tendency of Type B influenza epidemics to occur later in the northern hemisphere than Type A epidemics within a particular influenza season [62]. Infection of IBV following an IAV within the same season has been observed in children, with a mean interval between diagnoses of 50 days [136]. Challenge experiments in ferrets suggest that, depending on the order of infections, viral interference is possible between

Table 4.2: Parameter estimates and goodness of fit.

	2016-2018		2019
Estd. Parameters	Cross-protection	Neutral	Neutral
R_0^A	2.54	2.59	3.07
R_0^B	2.30	1.56	2.41
χ_{AB}	0.0061		
χ_{BA}	0.95		
b_1^A	0.28	0.44	0.31
b_1^B	0.27	0.38	0.3
ρ_A	0.0038	0.0062	0.0032
ρ_B	0.0014	0.0023	0.0019
t_0^A	0.15	0.13	
t_0^B	0.18	0.15	
logLik	-3669.89	-3626.61	-631.54
AIC	7359.78	7269.21	1287.08
ΔAIC	90.56	0	

IAV subtypes as well as between IAVs and IBVs when the interval between infections is less than one week [112]. While evidence indicates that human antibodies can provide cross-protection across influenza B lineages [118], support for antibody-mediated protection across influenza types is lacking. Other immunological mechanisms that might provide a degree of cross-type protection include depletion of susceptible target cells, infection surviving epithelial cells [52], and T-cell responses [38, 110, 196]. Social distancing and isolation of symptomatic individuals during influenza epidemics may also limit the number of potential hosts during an outbreak of a co-circulating influenza type [95, 169, 170].

We find no support for cross-protection as a driving epidemiological mechanism when we use maximum likelihood (see Methods for details) to fit our strategic model to Massachusetts influenza case data for the 2016 to 2018 influenza seasons. In particular, we fit the two-type model (see Methods, *SI Appendix,* Table 4.3 and Fig. 4.5) and compare its explanatory power of the data to a null model that assumes transmission dynamics of IAV and IBV to be independent (see *SI Appendix,* Table 4.4 and Fig. 4.1).

As we describe in Table 4.2, the added complexity of the model assuming short-term viral interaction ("cross-protection" model) is not supported by the data (ΔAIC = 90.56).

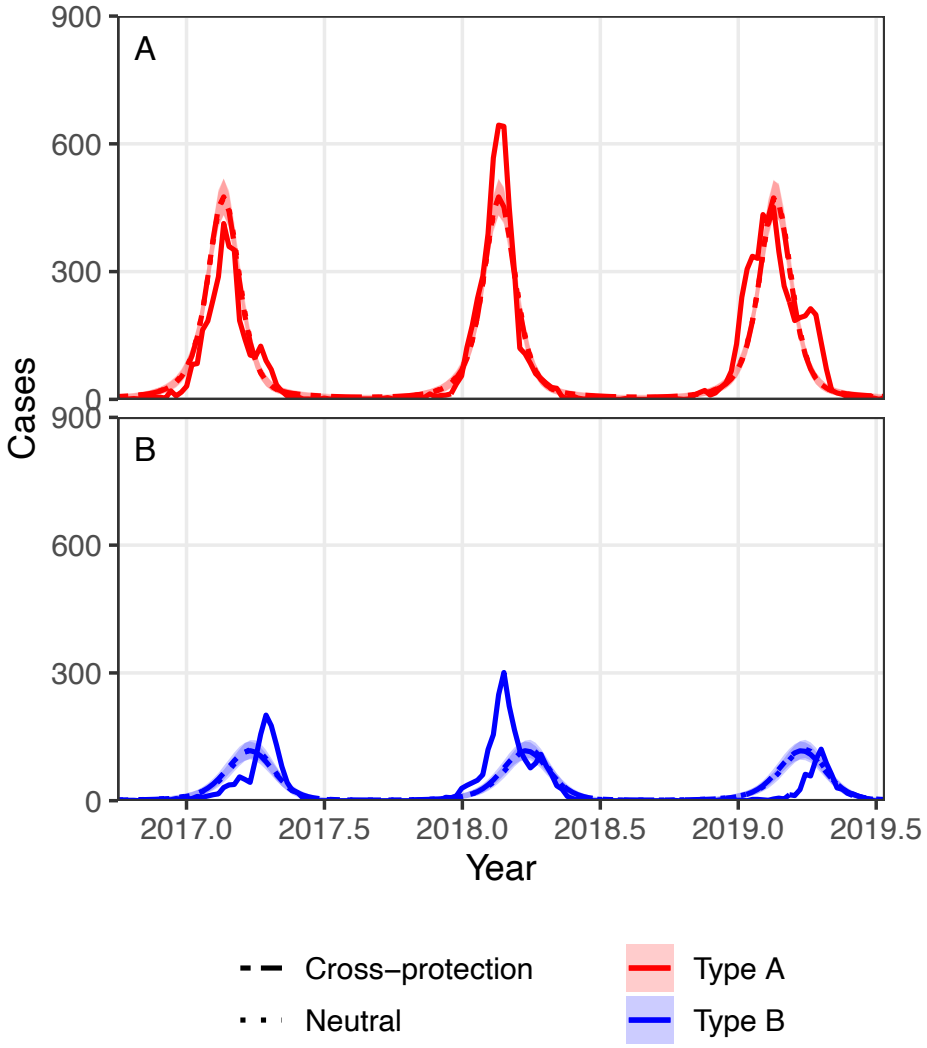

Figure 4.1: **Relative fits between epidemiological models for seasons 2016-2018.** Plots show matched type specific simulated trajectories - neutral (dashed) and cross-protection (dotted) models fitted to weekly case report data (solid) for MA. Ribbons represent Poisson observation error around model predictions.

To examine whether changes in virus transmission potential may underlie the observed epidemiological shifts, we fit the neutral model to incidence data for the 2019 season Fig. 4.2, Table 4.5). When comparing estimates based on 2016-2018 seasons to the 2019 season estimates, we observe an

18.5% increase in estimated R_0^A from 2.59 (95% CI: 2.59, 2.60) to 3.07 (3.06, 3.08) for season 2019 (see Table 4.2 and *SI Appendix*, Tables 4.4 and 4.5). Strikingly, we observe a 54.4% increase in the estimate of R_0^B from 1.56 (1.56, 1.58) in seasons 2016-2018 to 2.41 (2.40, 2.42) in season 2019. Note that a similar increase is found when R_0^B is estimated from Bayesian Skyride data (see Methods), with a rise from a value of 1.15 (95% CI: 1.11,1.20) in 2016 to 1.82 (1.71,1.96).

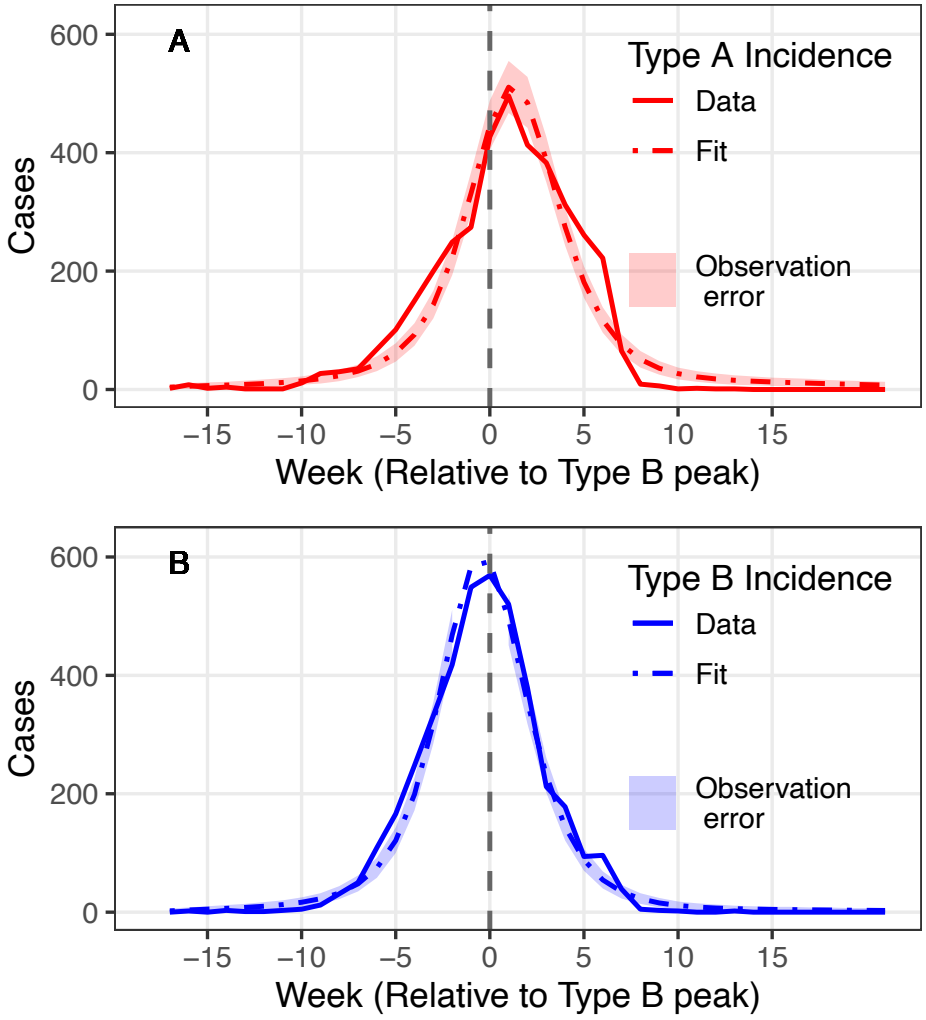

Figure 4.2: **Model fit for 2019 Influenza season.** Plots show matched type specific simulated trajectories for the neutral model (dashed) fitted to weekly case report data (solid) for MA. Ribbons represent Poisson observation error around model predictions.

As a final test of our key idea, we fit the neutral model to the data assuming no change in R_0^B across these four seasons (2016 to 2019). We find strong

support for a model with an increase in R_0^B at the outset of the 2019 season ($\Delta AIC = 530.42$; *SI Appendix,* Table 4.6).

4.3.2 Sensitivity to pathogen interaction

To explore the relationship between pathogen interaction and seasonal coupling between influenza A and B respectively, we develop a strategic, two-type mechanistic compartmental model (see Methods, *SI Appendix,* Fig. 4.5 and Table 4.3). The model structure permits an investigation of how potential changes in type-specific transmissibility and potential cross-type interference might influence the timing and magnitude of type-specific influenza epidemics. We consider a range of potential cross-protection parameter values (Fig. 4.3, Methods Table 4.1) and find that changes in type-specific basic reproduction numbers can, in principle, explain the observed changes in the relative timing of their peaks in the absence of cross-type protection. In fact, for a set ratio of type-specific reproduction numbers, changes in average cross-protection do not substantively change the lag between type-specific influenza peaks or differences in peak height (see Fig. 4.3A,B, scenarios b and c). When the Type A reproduction number (R_0^A) exceeds the Type B reproduction number (R_0^B) (Fig. 4.3 scenario a), we observe dynamics consistent with prior seasons in which Type A cases peak first and have a larger peak. By contrast, when the Type B reproduction number surpasses that of Type A, we observe an early and large peak in B cases (Fig. 4.3 scenario d) as observed in the 2019 U.S. influenza season. We note that such a peak is not observed when the duration of natural-immunity for Type B is high relative to that of Type A (*SI Appendix*, Fig. 4.6).

4.3.3 Union of dynamics in the type R_0 and susceptible backlog can generate an atypically early influenza B peak.

Given that our estimate of R_0^B for the 2019 season is lower than that of R_0^A, an increase in the basic reproductive number alone cannot account for the recent seasonal influenza dynamics. What, then, explains the unusual

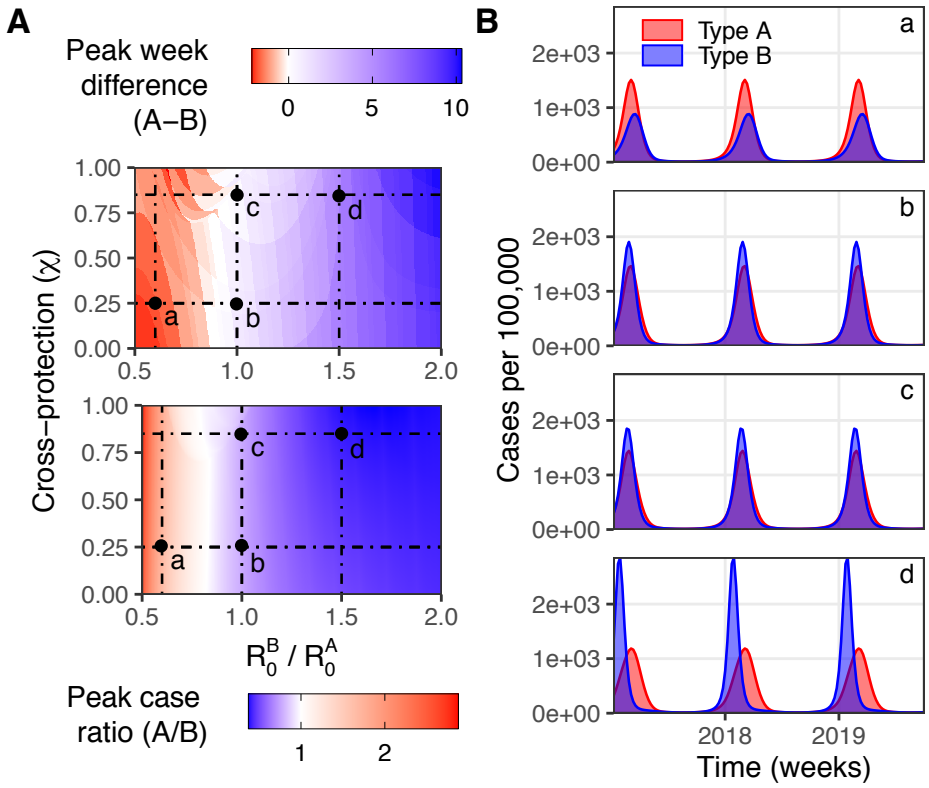

Figure 4.3: **Simulation study reflecting dynamic effects of cross-protection ($\chi_{AB} = \chi_{BA} = \chi$) and relative changes in type specific R_0s. A)** Depicts the difference in the epidemic phases of Types A and B (top) and the peaks ratios (bottom) of the two influenza types with relative changes of R_0^B with respect to R_0^A (x-axis). **B)** displays epidemic dynamics (cases per 100,000) for the two types resulting from parameter values selected at points (a), (b), (c) and (d) in **A**. R_0^A is fixed at 2, duration of cross protection is fixed at 1 month. Type-specific immunity is assumed to last 4 years for both types. See *SI Appendix*, Fig. 4.6 for results assuming that Type B-specific immunity lasts longer than Type A-specific immunity.

epidemiological patterns of the 2019 season? To answer this, we observe that simple transmission models describe the growth rate of an epidemic as

$$\frac{dI}{dt} = \gamma I\left(R_0\frac{S}{N} - 1\right), \tag{4.12}$$

where γ is the recovery rate. Therefore, in the early stages of an influenza season, epidemic take-off is determined by the combination of pathogen

transmissibility, as quantified by R_0, and the fraction of the population that is susceptible, $\frac{S}{N}$. Their product is commonly referred to as the *effective reproduction number*, R_{eff} [95]. We submit, therefore, that the unusually early and rapid rise in IBV cases in the 2019 season resulted from the documented substantial increase in R_0^B which coincided with a large fraction of the population susceptible to influenza B due to historically lower R_0^B values (Fig. 4.4). In concert, in the late summer months of 2019, these factors led to a higher relative *per capita* transmission rate for influenza B $\left(\frac{1}{I^B}\frac{dI^B}{dt} = 2.698\right)$ than for influenza A $\left(\frac{1}{I^A}\frac{dI^A}{dt} = 2.584\right)$, using (4.12), with S^B and S^A obtained from the fitted model (*SI Appendix*, Table 4.5 and Fig. 4.2).

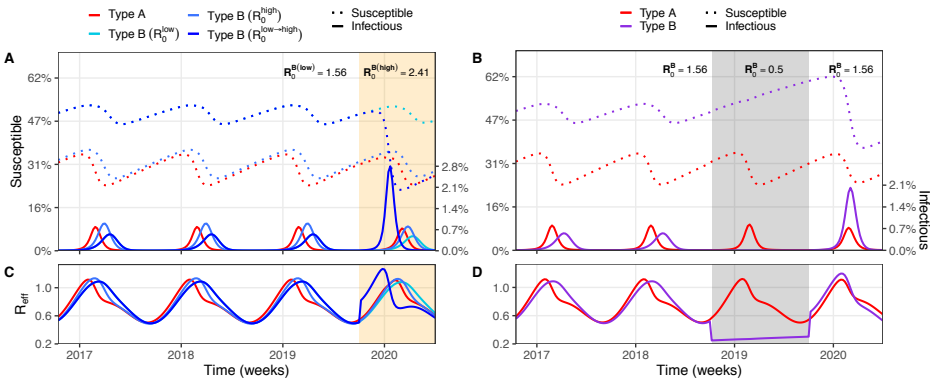

Figure 4.4: **Simulation-based illustration of the susceptible backlog hypothesis. A)** Simulation experiments demonstrating susceptible dynamics (dotted lines) and the corresponding relative timing and amplitude of influenza A (solid red line) and influenza B (solid blue lines). For influenza B, we depict three distinct scenarios: R_0^B is either low throughout ($= 1.56$), high throughout ($= 2.41$), or R_0^B starts low ($= 1.56$), but increases ($= 2.41$) at the start of the 2019 season (highlighted in orange). The associated effective reproductive numbers ($R_{\text{eff}} = R_0 \times \frac{S}{N}$) are presented in panel **C)**. Panels **B)** & **D)** present similar information to **A)** & **C)** but perform an alternative experiment, testing whether the absence of an influenza B outbreak in the 2018/2019 season highlighted in gray and resulting accumulation of susceptible individuals alone would explain the anomalous dynamics in influenza season 2019. Parameter values are presented in Table 4.2 and Table 4.5).

To illustrate our claim, we present a series of scenario analyses in Fig. 4.4. We show that an increase in R_{eff}^B in 2019 resulting from the combination of a higher R_0^B and a large susceptible fraction can capture the atypically early

and large influenza B outbreak (Fig. 4.4A,B), compared with no change in R^B_{eff} during seasons 2016-2019 when R^B_0 either remained low (1.56) or high (2.41). Further, there were relatively few IBV positive samples in the 2018 influenza season, therefore we examine whether a buildup of individuals susceptible to influenza B would explain the 2019 season patterns without an accompanying increase in R^B_0. This simulation could also be interpreted as a test of two additional ideas. We find that an accumulation of susceptible individuals leads to an unusually large outbreak in 2019, but does not affect the lag with the influenza A outbreak (Fig. 4.4C,D).

4.4 Discussion

We investigate mechanisms driving the intriguing the 2019 U.S. influenza season that is intriguingly atypical in relation to the previous seasons, with an uncharacteristically early surge in influenza Type B cases. Our exploration of transmission models reveals that the relative timing of the epidemics of these viruses is more sensitive to their respective transmission potential, as quantified by R_0, than to any changes to cellular- or immune-mediated interactions. It is important to emphasize that our results do not argue against such within-host effects; instead, our models indicate that these virus-virus interactions do not leave a strong dynamical footprint in population-level incidence data. This is because of the combination of modest R_0s for both virus types leading to a relatively small fraction of the population affected and the short-term nature of any influenza interference between virus types (potentially lasting only a few days [112]). Our working hypothesis instead is that the early arrival and large epidemic of influenza B cases in the 2019 season can be explained by the combined effects of influenza B with higher transmissibility than prior IBVs [202] and the availability of an unusually large susceptible pool due to historically lower R^B_0 values in past seasons, as illustrated in Fig. 4.4A. This proposed explanation is consistent across our likelihood-based model fitting and our strategic modeling.

Our analysis is limited by the number of years of data available. Reporting procedures changed dramatically following the 2009 H1N1 pandemic, and

we have restricted our analysis to include seasons after the pandemic. Studies connecting subsequent infections with Type A and Type B influenza at the individual-level, such as [136], are needed to fully determine whether cross-protection between types occurs and if so, the typical duration of protection.

A key outstanding question is how influenza virus evolution may change the type-specific age distribution of cases (or positive samples) [201]. Influenza infection histories are more likely to include one or more prior infections with an influenza B virus for those in older age-groups. It is possible that dominating lineages play a role in determining the proportion of positive samples that are Type B. Important avenues for future research include considering the roles of lineage determined disease severity, reporting probability, and age-specific susceptibility.

Changes in the timing of Type A and Type B influenza peaks have implications for optimal vaccination strategies [24], particularly given that the effectiveness of influenza vaccines is thought to wane over a period of approximately six months [57]. The apparent connection between evolution of influenza lineages and early timing of influenza B cases suggests that measures of evolutionary change could be a valuable source of information for designating the timing of vaccination campaigns. Type-specific differences between susceptibility to infection and severe disease could also help identify high risk age-groups for targeted vaccination. Thus, it may be beneficial to identify at-risk age-groups based on the expected timing of type-specific peaks in infections. Increased viral surveillance and characterization will be needed to effectively determine the optimal timing of vaccination, which may differ between age-groups [56]. Early awareness of the circulation of a novel influenza clade (such as the B/Victoria subclade V1A.3 identified in the 2019 season) could motivate early initiation of vaccination of younger age-groups that may be at a higher risk of infection. This knowledge would be helpful for physicians and public health officials, particularly when other viruses with varying effects across age groups (e.g., SARS-CoV-2) are co-circulating.

4.5 Supplementary Results

Table 4.3: Model State Variables

Compartment	Definition
S	Susceptible individuals with no prior immunity
I_A	Infectious with Type A; no prior immunity
I_B	Infectious with Type B; no prior immunity
C_A	Recovered from A; cross-protected against B
C_B	Recovered from B; cross-protected against A
R_A	Recovered from A; no cross-protection
R_B	Recovered from B; no cross-protection
I_{BA}	Infectious with B; prior immunity to A
I_{AB}	Infectious with A; prior immunity to B
R	Recovered from A and B
K_A	Cumulative cases of Type A
K_B	Cumulative cases of Type B

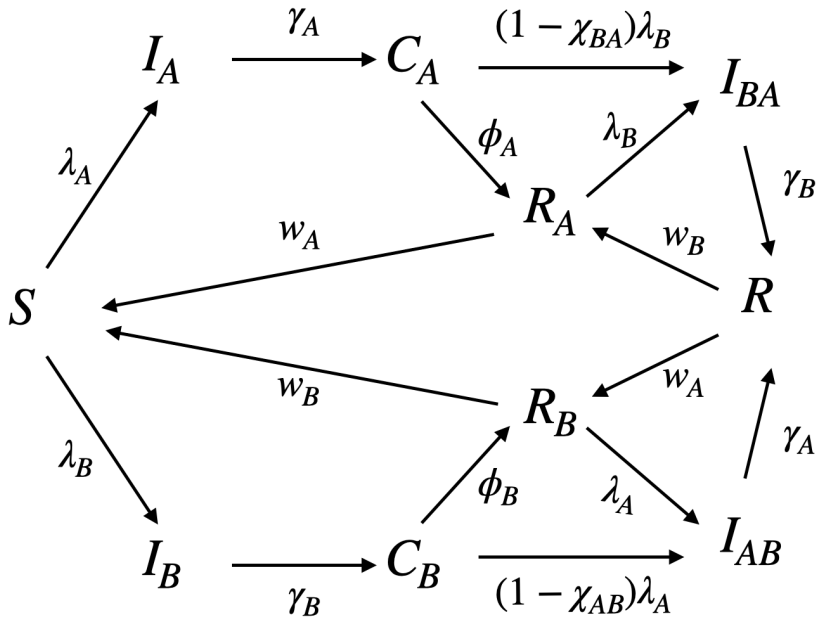

Figure 4.5: **Two-type influenza transmission model schematic.** Definition of rate parameters and state variables are provided in Tables ?? S1 respectively.

Table 4.4: Parameter estimates and goodness of fit (2016-2018 seasons). Cross-protection and neutral models were formally contested to assess relative goodness of fit using ΔAIC values. Models were integrated until equilibrium (burn-in period of 100 years) and subsequent 3 years were used in the fitting procedure. Simulated trajecteories were fit to weekly, type-specific, incidence data for influenza for the state of Massachusetts.

	2016-2018	
Estd. Parameters	Cross-protection	Neutral
R_0^A	2.539 (2.536,2.559)	2.591 (2.585,2.601)
R_0^B	2.305 (2.301, 2.327)	1.562(1.559, 1.578)
χ_{AB}	0.00605 (0.00604, 0.00743)	
χ_{BA}	0.953 (0.951, 0.955)	
b_1^A	0.28 (0.27, 0.29)	0.4419 (0.4404, 0.4427)
b_1^B	0.2712 (0.2707, 0.2732)	0.376 (0.374,0.377)
ρ_A	0.00377 (0.00377, 0.00379)	0.00620 (0.00618, 0.00622)
ρ_B	0.0014 (0.0013, 0.0014)	0.0023 (0.0023, 0.0024)
t_0^A	0.155 (0.154,0.156)	0.132 (0.131, 0.133)
t_0^B	0.184 (0.183,0.185)	0.152 (01.151, 0.154)
logLik	-3669.89	-3626.61
AIC	7359.78	7269.21
ΔAIC	90.56	0

Table 4.5: Parameter estimates and goodness of fit (2019 season). The best fitting model (neutral) from the previous era, Table 4.4, was used to estimate type specific epidemiological parameters. The model was integrated to reach equilibrium (burn-in period of 103 years) using parameter MLE from the previous era. Eventual sizes of model compartments from this simulation were fixed as initial conditions for 2019 season. Model parameters not estimated above were fixed at the MLE from the previous era. Simulated trajectories were fit to weekly, type-specific, incidence data for Influenza for the state of Massachusetts.

	2019	
Estd. Parameters	Flu A	Flu B
R_0	3.074 (3.062, 3.083)	2.4093 (2.397, 2.419)
b_1	0.308 (0.306,0.309)	0.293 (0.292, 0.294)
ρ	0.0032 (0.0031,0.0033)	0.00192 (0.00191,0.00199)
logLik	-408.5	-223.04

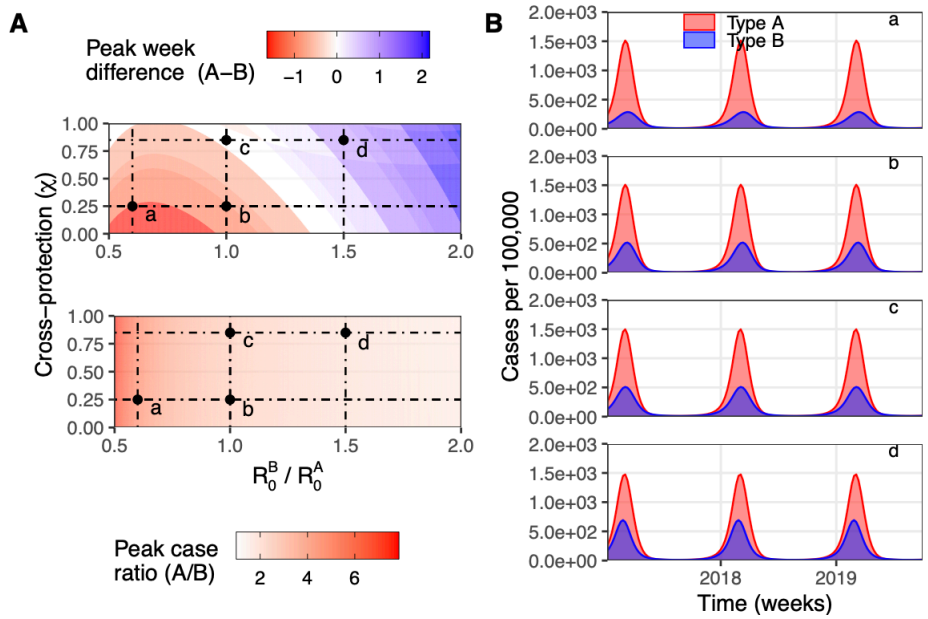

Figure 4.6: **Simulation study demonstrating dynamic effects of cross-protection and changes in type-specific R_0s when the duration of natural immunity differs between influenza types.** Type-specific duration of natural-immunity for Type A and Type B was set to be 4 years and 10 years respectively. **A)** Shows changes in the peak case ratio (top) and phase-difference in the peak weeks (bottom) of the two influenza types with relative changes of R_0^B with respect to R_0^A (x-axis) and cross-protection (χ). **B)** displays epidemic-dynamics (cases per 100,000) of the two influenza types using parameter values at points (a), (b), (c), and (d) in **A**. R_0^A is fixed at 2 and the duration of cross-protection ($\frac{1}{\phi}$) is set to 1 month. χ is assumed to be symmetric across the two influenza types.

Table 4.6: Parameter estimates and goodness of fit. The neutral model was fit to seasons 2016-2019 to assess relative goodness of fit against the best fitting model (neutral with season dependent parameter values, Tables 4.4 and 4.5) using ΔAIC. Simulated trajectories were fit to weekly, type-specific, incidence data for influenza for the state of Massachusetts.

Estd. Parameters	2016-2019 Neutral
R_0^A	4.104
R_0^B	2.871
χ_{AB}	
χ_{BA}	
b_1^A	0.281
b_1^B	0.160
ρ_A	0.00334
ρ_B	0.00124
t_0^A	0.269
t_0^B	0.225
logLik	-4535.358
AIC	9086.720
ΔAIC	530.436

CHAPTER 5

DISENTANGLING THE CAUSES OF MUMPS REEMERGENCE IN THE US

5.1 Introduction

The sustained re-emergence of mumps in multiple high-vaccine coverage countries, including the US [14], Canada [157], England [75, 92], South Korea [151], and Denmark [125], over the last two decades is puzzling from a population biological perspective, and has raised concerns about the adequacy of current vaccines [45]. Mumps is an infectious disease caused by an RNA virus of the family *Paramyxoviridae* [160]. While infection in school-aged children is typically mild, with symptoms including inflammation in the salivary glands and fever [160], in post-pubescent individuals infection can cause severe disease, including encephalitis [192], deafness [80], and infertility [132, 210]. Prior to routine infant immunization, infection with mumps was viewed as a childhood rite of passage [140], with serological studies from the pre-vaccination era suggesting that by 14–15 years of age, 90% of the population were typically seropositive to mumps antibodies [65, 133]. In countries like the US, routine immunization has been highly successful at substantially reducing mumps incidence by over 99.9% as compared to the pre-vaccine era incidence [45], and it

was once considered a candidate for global elimination [65]. However, the recent resurgence of mumps in some highly vaccinated populations has cast doubt on the feasibility of this goal [162].

The exact nature of vaccine failure underlying this recent re-emergence remains contested [72, 115, 208]. The two leading hypotheses are that (1) vaccinal immunity against mumps transmission wanes over time [115, 208], and (2) genetic mismatch between vaccine strains and presently circulating viruses has reduced vaccine efficacy [71, 72]. There is empirical evidence in support of both hypotheses [88, 203] and given the contrasting policy implications associated with each mechanism, it remains important to pin down the underlying cause of mumps resurgence. With waning vaccine effectiveness, periodic booster immunizations are necessary [115, 168]. For a mismatched vaccine, on the other hand, boosters with the same vaccine are typically ineffective and an updating of the vaccine is called for [50, 121, 165, 213].

In support of the waning hypothesis, longitudinal serological studies have shown that vaccine-derived antibodies decay over time [174], however there is uncertainty about how this maps onto waning of protective immunity at the population level since serological correlates of protection have not been identified [115]. Epidemiological factors consistent with waning immunity include a shift in the age-distribution of reported cases to the older age-groups (typically 18-25y) [14]. Waning of vaccine efficacy serves to diminish the population level immune protection by bolstering the susceptible pool and increasing the chances of successful pathogen invasion and persistence [77, 129, 167].

In addition, there is evidence suggesting that vaccination might be providing incomplete protection (referred to as "leaky" protection [129]) against presently circulating strains. Mumps viruses exhibit substantial phylogenetic diversity, with 13 known genotypic variants identified globally [93]. Historical vaccines (such as the original Jeryl Lynn vaccine developed in the US) are typically derived from prior circulating genotype A viruses. However, the recent resurgent epidemics in the UK, Canada, and the US have all been caused by genotype G viruses, indicative of genotype replacement [208]. Furthermore, a molecular analysis of genotypes A and G

has identified significant differences in 8 major sites associated with the B-cell response. The authors of this study posit that this antigenic mismatch may drive the reduced effectiveness of the vaccine against the genotype G viruses [72].

Using the reemergence of mumps in the US as a case study, in this paper, we aim to determine the putative mechanism of vaccine failure underlying national level mumps re-emergence by performing likelihood-based statistical inference using dynamical transmission models and an age-stratified time-series of national level mumps incidence. Transmission models have proven useful in drawing ecological inferences about re-emergence of numerous vaccine preventable childhood infections including measles [73, 134], pertussis [48], influenza [22], and diptheria [43]. Here, we construct an age-structured *SEIR* mechanistic model informed by demographic and vaccine uptake data to test the competing hypotheses regarding vaccine failure. We conduct formal hypothesis testing using age-structured time-series of national-level US mumps incidence between the years of 1977-2018 sourced from the Centers for Disease Control and Prevention [142, 161]. We find that the model with waning vaccine immunity best explains the dynamics in the observed data. Our model is effectively able to reproduce crucial characteristics such as the timing of recent epidemics, and the gradual increasing trend in the age-distribution of cases to the older age-groups.

5.2 Methods

5.2.1 Case Reports And Co-variate Data

We acquired mumps incidence data from two separate sources. Annual tables on the age-specific case reports of mumps were acquired from the Morbidity and Mortality Weekly Reports (MMWR) published by the Centers of Disease Control (CDC) [142]. Initially, MMWR pdf tables were digitized to a computer readable format. Further, the age-specific case reports were aggregated to generate six time series with roughly 5 year intervals during 1977-2018 (Fig. 5.1A–5.1C). Weekly aggregated case reports were downloaded from the Project Tycho database [161]. These

data were spatially grouped to generate a high-frequency time-series of national-level mumps case reports during 1967-1985 (Fig. B.2B). We temporally aggregated data from the weekly time series to generate the average spatial distribution of mumps in the US during the two major epidemics in the vaccine era (Fig. 5.1D–5.1E).

Data on demographic covariates were obtained from the US census website [33]. We processed the demographic time series to generate (1) age-specific estimates of national level population sizes consistent with the age groupings of the reported mumps cases, (2) estimates of age-specific migration rates (see Supplementary Information Ch.4 for details Fig. 5.1A–5.1C), and (3) national-level estimates of the annual birth rates for the US (Fig. B.3B, B.3D & B.3E). Age-specific contacts can be non-homogeneous *e.g.* infants come more in contact with their (adult) parents than other infants. On the other hand, teenagers make more physical contacts with other teenagers of a comparable age. To account for the assortative mixing among the various age-based sub-populations, we used the contact matrix reported in the POLYMOD study [135] for the UK. This allowed us to include a social contact structure that is closely comparable to the US in our analysis [48]. This contact matrix was transformed to correct for reciprocity, such that the total daily contact rates had a symmetric structure, and to match age distribution for the US (Fig. B.1, see Supplementary Information Ch.4 for details). National level coverage estimates of mumps vaccines, for the neonatal and booster schedule, were downloaded from the World Health Organization (WHO) website [90]. Missing early coverage for these time series were imputed for both doses (Fig. B.2C; Fig. B.3C–B.3D; see Supplementary Information Ch.4 for details).

5.2.2 Models Of Mumps Re-emergence

To investigate the potential mechanism of vaccine failure in explaining the re-emergence of the mumps virus, we formulated an age-structured Susceptible-Exposed-Infectious-Recovered (*SEIR*) transmission model [96]. The population, with total size $N(t)$, was divided into 6 compartments. Individuals enter the first vaccinated compartment (V^1) at the rate $(1-\alpha)p(t)\nu(t)$, where $p(t)$ is the annual infant immunization coverage in the

year t, and α is the probability of primary vaccine failure, and $\nu(t)$ is the *per capita* birth rate. The remainder of the newborns are assumed to be susceptible (S) who can get infected with either mumps genotypes at rate λ_n, where λ_n is the force of infection of mumps genotype n ($n \in \{A, G\}$). Following infection, individuals move into the corresponding exposed compartments, E^n , remaining latently infected for a period of $\frac{1}{\sigma}$ days, before moving to the infectious compartments, I^n. Ultimately, the infection is cleared following a mean duration of $\frac{1}{\gamma}$ days, and individuals acquire long-term immunity, as they enter the recovered (R) compartment. Vaccine-derived immune protection in this system lasts for a total duration of $1/\delta$ years after which, individuals re-enter the S class. Here we considered a model with variable (specifically 1 and 3) sub-classes for the vaccinated sub-population in order to generate a Erlang-distributed waiting time with shape parameter x [5, 96, 205]. When $x = 1$, we recover the exponential distribution and as x increases the distribution becomes increasing peaked about $(1/\delta)$ with the variance declining according to $\frac{1}{x\delta^2}$. It was desirable to generate a distribution of vaccinated that was distinct from the exponential distribution and yet have a significantly minimal increase in the computational overhead during simulation. As a pragmatic decision, three vaccinated sub-classes were thus used. Vaccinated individuals are also capable of acquiring the infection with a force of infection reduced by a genotype-specific factor ε^n which is typically referred to as the vaccine leakiness [77, 123]. The model accounts for the importations of infectious individuals of genotype n at the rate ι^n per year. Fig. B.4 depicts the vaccine imperfection model structure.

The genotype-specific force of infection is proportional to the number of individuals in the I^n compartment and the seasonality of transmission,

$$\lambda^n(t) = qC\left[1 - \beta_1 sin\left(\frac{2\pi t}{T}\right)\right]\frac{I^n(t)}{N(t)}. \tag{5.1}$$

Here q, C, β_1 are the probability of infection given contact, the average daily contact rate, and the amplitude of seasonality respectively, and T is the period of seasonality (assumed to be 1 year). The seasonal forc-

ing is assumed to be sinusoidal with the fixed phase of seasonality. The model dynamics are represented using the following system of differential equations

$$\frac{dS}{dt} = \Big(1-(1-\alpha)p(t)\Big)\nu(t)N(t) + 3\delta V^3 - \Big(\sum_n \lambda^n(t) + \mu(t)\Big)S, \tag{5.2}$$

$$\frac{dV^1}{dt} = (1-\alpha)p(t)\nu(t)N(t) - \Big(3\delta + \sum_n \varepsilon^n\lambda^n + \mu(t)\Big)V^1, \tag{5.3}$$

$$\frac{dV^2}{dt} = 3\delta V^1 - \Big(3\delta + \sum_n \varepsilon^n\lambda^n + \mu(t)\Big)V^2, \tag{5.4}$$

$$\frac{dV^3}{dt} = 3\delta V^2 - \Big(3\delta + \sum_n \varepsilon^n\lambda^n + \mu(t)\Big)V^3, \tag{5.5}$$

$$\frac{dE^n}{dt} = \lambda^n(t)\Big(S + \varepsilon^n V\Big) - \Big(\sigma + \mu(t)\Big)E^n + \iota^n, \tag{5.6}$$

$$\frac{dI^n}{dt} = \sigma E^n - \Big(\gamma + \mu(t)\Big)I^n, \tag{5.7}$$

$$\frac{dR}{dt} = \gamma\sum_n I^n - \mu(t)R. \tag{5.8}$$

Here, $\varepsilon_n = \varepsilon$ when $n = G$, and 0 otherwise; $\iota^n = \iota$ when $n = G$ and $t = t_{intro}$, 0 otherwise. The model described above distills the essential reactions representing various imperfection traits. For complete details on the full age-structured model, compartments, and parameter definitions refer to Supplementary Information Ch.4, and Tables B.1,B.2.

For statistical inference with the case data, we kept track of new cases generated by the model. This was achieved by calculating a piece-wise continuous integral over a time period comparable to that of the observed data (1 year), with the expression for new cases over the annual interval $[t-1,t)$ given by

$$C_t = \int_{t-1}^{t} \gamma\sum_n I^n(s)ds. \tag{5.9}$$

The model was implemented in R package "pomp" [102].

5.2.3 Reproductive Numbers And Vaccine Impact

We derived an expression for the basic reproductive number (R_0) for the age-structured transmission model using the next generation method (NGM) [47] (see Supplementary Information Ch.4 for details). As R_0 quantifies pathogen transmissibility in a fully susceptible population (i.e. without vaccination), we derived an analoguous measure of transmissibility in presence of the vaccine, which we denote R_p. Using these two reproductive numbers, we further derived an expression for the population-level vaccine impact, ξ, defined as the decrease of transmissibility of the virus in the presence of the vaccine ($p > 0$) relative to the transmission in absence of the vaccine ($p = 0$) [123],

$$\xi = 1 - \frac{R_p}{R_0}. \tag{5.10}$$

5.2.4 Parameter Estimation and Hypothesis Testing

We utilized age-stratified time series of annual mumps case reports (see Fig. 5.1B) to perform maximum-likelihood estimation of model parameters to ascertain relative support among 4 qualitatively distinct hypotheses, namely: (1) The “no loss” hypothesis assumes that the vaccine is perfect ($\varepsilon = 0$, $\delta = 0$); (2) the “exponential waning” hypothesis assumes that the the vaccine derived immune protection wanes over time with exponentially distributed duration and that there is no vaccine leakiness ($\varepsilon = 0$); (3) the “Erlang waning” hypothesis is the same as the exponential waning hypothesis, but assumes that the duration of immunity is Erlang distributed with shape parameter $x = 3$ (see above); (4) the “leaky” hypothesis assumes that at time $t = t_{intro}$ a novel genotype is introduced to the population against which the vaccine provides leaky immunity ($\varepsilon > 0$) . For the leaky hypothesis, we also analysed the impact of introducing the novel genotype to differing age classes. Moreover, for each hypothesis, we also investigated the effect of 4 different shapes of booster uptake immediately following

its roll-out. A formal comparison of all these models was conducted as described below.

Trajectory matching was employed to find maximum-likelihood estimates of the unknown model parameters. As the data included cases without recorded age ("unstructured cases") as well as age-structured cases, we constructed a reporting model capable of fitting to both data types by including a time-varying probability of age being recorded, η_t, calculated from the proportion of case reports that were age-structured (see Fig. B.3A). We assumed that the observation errors in both age-structured and unstructured annual case data followed a normal distribution allowing for over-dispersion in the reporting process [83]. Specifically, for age-structured data, the mean and variance in the number of reported cases at time t in age group i that depend on the age-specific reporting probability, ρ_i, and overdispersion parameter ψ_i, are given by

$$\mu_{i,t} = \rho_i \eta_t C_{i,t}, \tag{5.11}$$

$$\sigma^2_{i,t} = \mu_{i,t}(1 - \rho_i \eta_t + \psi_i^2 C_{i,t}), \tag{5.12}$$

for $i \in \{[0,5),[5,15),[15,25),[25,40),\geq 40\}$. The unstructured data are distributed similarly, with mean and variance

$$\mu_{u,t} = \rho_u(1-\eta_t)\sum_i C_{i,t}, \tag{5.13}$$

$$\sigma^2_{u,t} = \mu_{u,t}\left(1 - \rho_u(1-\eta_t) + \psi_u^2 \sum_i C_{i,t}\right), \tag{5.14}$$

where ρ_u and ψ_u are the reporting probability and overdispersion parameter for unstructured cases. The likelihood function for the observed age-structured data $D_{i,t}$ and unstructured data $D_{u,t}$, conditioned on the model parameter vector Θ was given as

$$\mathscr{L}(\Theta) = \prod_t \left(f\left(D_{u,t}; \mu_{u,t}, \sigma^2_{u,t}\right) \prod_i f\left(D_{i,t}; \mu_{i,t}, \sigma^2_{i,t}\right) \right), \tag{5.15}$$

where $f\left(x; \mu, \sigma^2\right)$ is the probability density function of a normally distributed random variable with mean μ and variance σ^2.

Maximization of the likelihood function (Eq. 5.15), was carried out using a differential evolutionary (DE) algorithm. It is a stochastic optimization approach that is known to provide an efficient convergence to the global optimum for real-valued objective functions [66](see Supplementary Information Ch.4 for details). DE was implemented using the R package DEoptim [138]. The latent state-variables were initialized near an endemic equilibrium. The first 100 years were discarded as transient dynamics and subsequent 36 years (1977-2012) were used in the calculation of likelihood. To compare the relative goodness of fit among the four competing hypotheses, we used the Akaike information criterion (AIC), defined as

$$AIC = 2p - 2\log\left(\mathscr{L}(\Theta)\right), \tag{5.16}$$

where p is the number of free model parameters. At 5% level of significance, an AIC value of 1.92 was taken to be a statistically significant difference between model fits.

5.2.5 Uncertainty Estimation

We used parametric bootstrapping to estimate confidence intervals for MLEs. This involved the following: 1) For each model, MLEs obtained in the preceding section were used to simulate 1,000 synthetic time series. 2) Starting at the MLE, a local search with a Nelder-Mead optimizer [143] was conducted and the parameters estimated for every time series. 3) This results in a bootstrapped distribution of parameter estimates. Estimate sets corresponding to 2.5^{th} and 97.5^{th} percentiles of this distribution were taken to be as the 95% confidence bounds of the model MLEs.

5.2.6 Model Performance

A subset of the case-reports data (6 years, 2013-2018) was reserved to test the performance of the model by conducting out-of-sample prediction. Model performance was quantified using the coefficient of determination [46], R^2, between the observed case records (D) and the log-transformed

median simulated reported cases (M),

$$R^2 = \left[\frac{CoV(D,M)}{\sqrt{V(D)V(M)}} \right]^2, \tag{5.17}$$

where $CoV(D,M)$, $V(D)$ and $V(M)$ are co-variance between D and M, variance of D and variance of M respectively. We computed values of R^2 for each separate age group during both the within sample (pre-2013) and out-of-sample infection prediction epochs.

We also compared how the resulting age-distribution of the the observed case data compared with the simulated cases from the best fitting model using the symmetrized Kullback-Leibler divergence [111]. For this analysis, we ignored the unstructured subset of reported cases due to their diminishing presence in the recent epidemics (see Fig. 5.1B). Age distributions for observed and simulated cases were constructed by normalising the data for each week, i.e. $A^D_{i,t} = D_{i,t} / \sum_i D_{i,t}$ and $A^M_{i,t} = M_{i,t} / \sum_i M_{i,t}$. The Kullback-Leibler divergence for each time point was then given by

$$\overline{D}^{KL}_t = \frac{1}{2} \left[\sum_i A^D_{i,t} \log\left(\frac{A^D_{i,t}}{A^M_{i,t}} \right) + \sum_i A^M_{i,t} \log\left(\frac{A^M_{i,t}}{A^D_{i,t}} \right) \right]. \tag{5.18}$$

We quantified uncertainty in our estimate of $\overline{D}^{KL}_t$ using bootstrap sampling. We sampled 1000 simulated replicates from our reporting model and calculated $\overline{D}^{KL}_t$ for each replicate by replacing $A^D_{i,t}$ in Eq. 5.18 with the simulated data.

5.3 Results

After the roll out of infant vaccination (see Fig. B.3C), the US experienced a pronounced decline in mumps incidence, down to a nadir of 10% of pre-vaccination levels in 1984 (Fig. 5.1A). However, this decline was interrupted in the mid-1980s, when the US experienced resurgent outbreaks, with cases concentrated in the 5-15 age group (Fig. 5.1B and 5.1C) [14] with a geographic concentration centered around the mid-western states (Fig. 5.1D) [78]. In response, a nationwide vaccine booster campaign was initiated in 1989 (see Fig. B.3D for estimated vaccine uptake data),

targeting 4 year-olds [140]. Subsequently, mumps incidence further declined until a second large outbreak in 2006, which has been followed by sustained transmission ever since (Fig. 5.1A). The majority of post-2006 cases have been aged over 15y (Fig. 5.1C), and the midwest has again been a geographic focus (Fig. 5.1E). Concerningly most cases have occurred in individuals who have completed the recommended two-dose vaccine schedule [139].

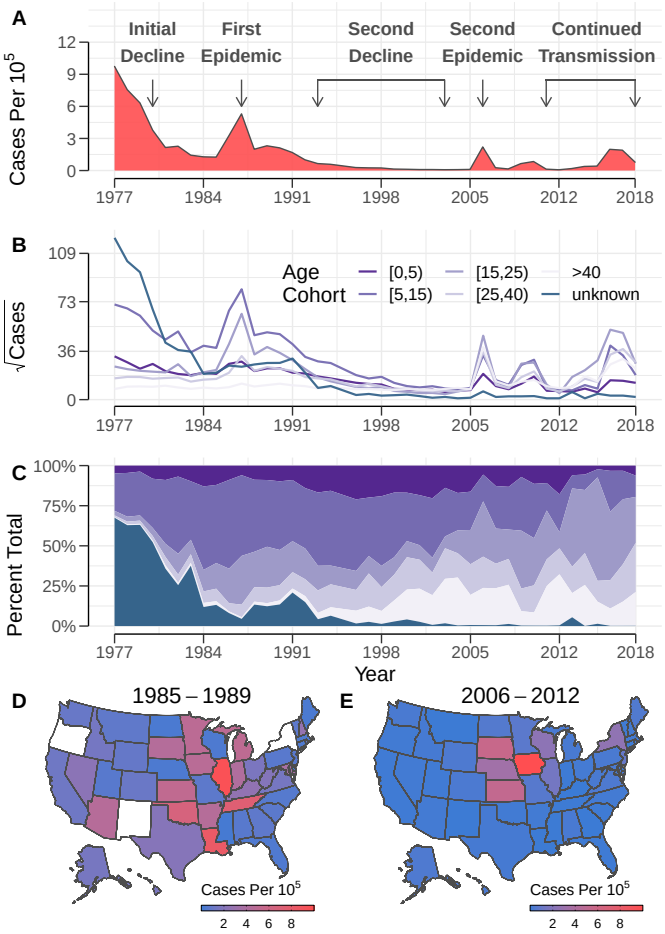

Figure 5.1: **Mumps distribution in the United States.** (A) Total reported mumps cases per 10^5. (B) Age-stratified case reports across 5 age cohorts (Purple gradient). (C) Age-distribution of mumps case reports (Purple gradient);(D–E) Spatial distribution of average mumps case reports per 10^5 during the two reemergence eras, 1985-1989 (panel D) and 2006-2012 (panel E).

5.3.1 Formal hypothesis testing

As outlined above, two distinct mechanisms of vaccine failure have been proposed to explain the resurgence of mumps in highly vaccinated countries: (1) transient vaccinal protection (the "waning" hypothesis) and (2) imperfect immunity against genotype G (the "leaky" hypothesis). To formally arbitrate among these putative drivers of mumps re-emergence, we formulated an age-structured, two-genotype transmission model that captures each mechanism of vaccine failure (see Methods, Eqs. B.13–B.22). We further investigated two models of immune waning. In the first, the duration of vaccinal immunity was assumed to be exponentially distributed (i.e. with no characteristic age at which immunity is lost). In the second, motivated by the observation that recent cases were primarily 15-25yo, we modelled the duration of vaccinal immunity using an Erlang distribution, which permits a modal age at which immunity is lost. All hypotheses were compared to a "no-loss" model which assumed perfect, permanent vaccine-derived immunity. Where possible parameter estimates were obtained from the literature, with unknown epidemiological and immunological parameters estimated via trajectory matching executed using likelihood-based statistical inference (see Methods for details). Finally, due to missing vaccine uptake data in the period covering the roll-out of booster doses (1989-2000), we considered 4 qualitatively distinct shapes of booster uptake (constant, sigmoid, convex and concave – see Supplementary Information Ch.4 for full details).

Among the competing models, we found the exponential waning model provided the best explanation for the observed data, as quantified using the Akaike information criterion (AIC) (see Table 5.1). For this model, we estimated a basic reproductive number, R_0 of 13.90 (95% CI: 8.02, 28.20), consistent with previous estimates [8]. While highly variable among individuals, on average vaccine-derived protection lasts for 111.45 y (95% CI: 111.4 y, 3×10^4 y). This implies that from a population of individuals who are vaccinated, many of the individuals lose immunity within a short period of time where as there is another subset of individuals who remains protected for a long period of time. On an average the distribution of such a population appears to be exponential with an average estimated to the cur-

Table 5.1: Table of parameter maximum likelihood estimates and derived quantities

	Model			
Parameter/Quantity	No Loss	Waning (Exponential)	Waning (Erlang, N = 3)	Leaky
ΔAIC	207.288	0	26.853	10.87
AIC	3006.822	2799.534	2826.387	2810.404
$\log\left(\mathscr{L}(\Theta)\right)$	-1484.411	-1379.767	-1393.194	-1384.202
R_0	42.414	13.992	31.371	22.199
R_p	0.484	5.72	18.69	0.76
ξ	0.989	0.591	0.404	0.966
β_1	0.185	0.036	0.341	0.735
σ^{-1} (Days)	12.022	24.998	16.894	24.999
δ^{-1} (Years)	-	111.456	56.457	-
ε	-	-	-	0.028
t_{intro}	-	-	-	2000
Booster shape	Constant	Sigmoid	Sigmoid	Constant

rent MLE. Of the booster uptake shapes, the sigmoidal trajectory provided the best fit to the data. To compare the performance of vaccination across models, given the different mechanisms of failure, we followed McLean & Blower [129] and Magpantay *et al.* [48, 123] and derived a quantification of the vaccine impact (ξ), defined as the reduction in population-level virus transmissibility in the presence of the vaccine. For the exponential waning model and vaccination coverage fixed at the average for 2008–2018 (91.7% and 91.6% for the neonatal and the booster respectively), ξ was estimated to be 59.1% (95% CI: 56.33%, 99.38%).

The remaining models were ranked based on the difference in AIC relative to the exponential waning model (ΔAIC). The next best performing model was the leaky model, with $\Delta\text{AIC} = 10.87$. Compared with the waning model, we estimated a higher value of R_0 for the leaky hypothesis (22.199; 95% CI: 22.19, 34.48). This model explains the mumps resurgence in early 2000s as a result of the introduction of a new genotype in 2000 that caused the erosion of herd immunity. Due to the limited size of the outbreak relative to pre-vaccination levels, our estimates of vaccine leakiness were highly constrained, $\varepsilon = 0.03$ (95% CI: 0.01, 0.05). Consequently, the leaky model predicts that the vaccine retained a high population-level vaccine impact, $\xi = 96.66\%$ (95% CI: 94.69%, 98.14%). Crucially, while the

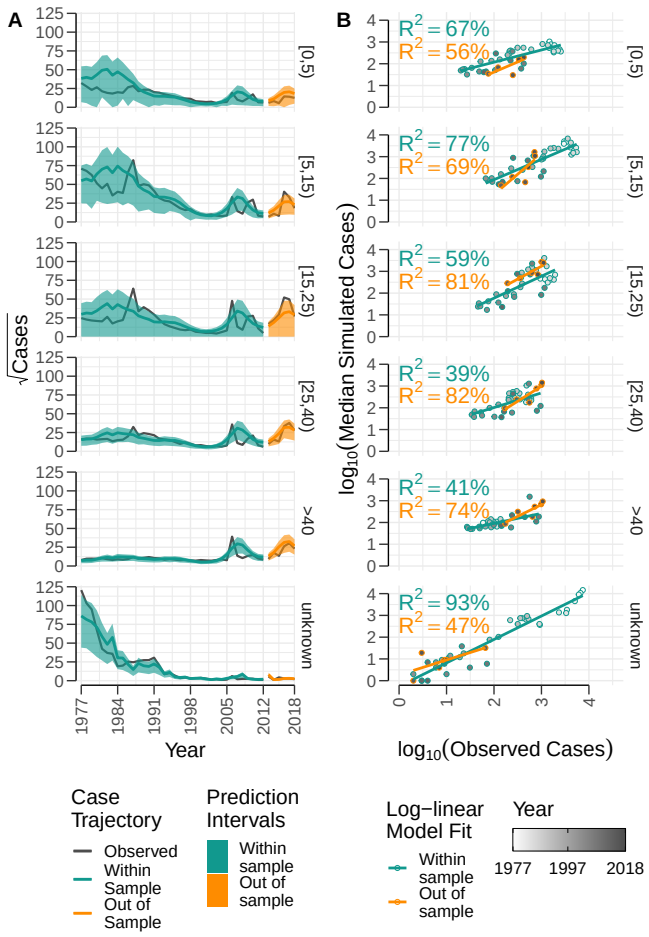

Figure 5.2: **Model-data agreement.** (A) Age-specific qualitative model fits to within sample time-series (1977-2012, seagreen) and the corresponding 5 year out of sample prediction (2013-2018, orange) for the waning hypothesis. Ribbons represent 80% prediction intervals for the two prediction epochs. (B) Age-specific log-linear model agreement for within sample fit (seagreen lines and hollow circles) and out of sample predictions (orange lines and hollow circles). Statistical performance of the model was assessed using coefficient of variation (R^2, reported inset) for the two prediction epochs. Year of record is represented using a continuous grey gradient.

leaky hypothesis correctly predicts disease resurgence from 2006 onwards, the predicted age-profile of cases lacks the peak in 15-25yo individuals found in recent data. Instead, cases are more evenly distributed across age groups (Fig. 5.6–5.7). The remaining two models (Erlang waning and

no loss) both performed substantially worse with ΔAIC = 26.9 and 207.3, respectively.

5.3.2 The waning model provides a good explanation of recent mumps epidemiology.

In addition to having the lowest AIC score, time series simulated from the exponential waning model are able to reproduce key epidemiological features of mumps data, including (1) the initial vaccine era decline, (2) a resurgence in the 1980s and (3) a second resurgence in the mid-2000s (Fig. 5.2A). Two dynamical features that this model struggles to capture are (1) the peak and duration of the mid-80s re-emergence, and (2) the distinct outbreaks in 2006 and 2010, which the model combines into a single multi-year outbreak. Predictive accuracy was limited during the first re-emergence era (Fig. 5.2A). This was likely due to over-dispersion in the reporting errors as suggested by our parameter estimates (Fig. 5.8C). To validate our fitted model, we withheld the final five years of the time series (2013-2018). Strikingly, our model accurately captures the multi-annual dynamic signature, and correctly predicts the epidemic in 2017.

We quantified the predictive performance of the exponential waning model using the coefficient of determination (R^2). To isolate predictive characters of the model across the six age-specific time series, we estimated R^2 independently for each age cohort (Fig. 5.2B). We found that the model best captures the dynamics associated with the unstructured cases (those with missing age information) ($R^2 = 93\%$). We observed the lowest R^2 in the [25, 40) y age cohort where the model explained 39% of observed variation. In other age classes, we report relatively high values of R^2 we report relatively high values of R^2 with 67%, 77%, 59% and 41% for [0,5) y, [5,15) y, [15,25) y, and $>$ 40 y respectively. Surprisingly, for the majority of age groups, the prediction in the out-of-sample epoch, out-performs the prediction in the training data (Fig. 5.2B). Better predictability in the out-of-sample epoch might be a associated with dynamically simple trajectories of observation data during this time period as compared to the dynamics realized in the within-sample epoch. Moreover, if the predic-

tion horizon was to expanded to include more data, we would expect a relatively lower value of model accurcy in the this region.

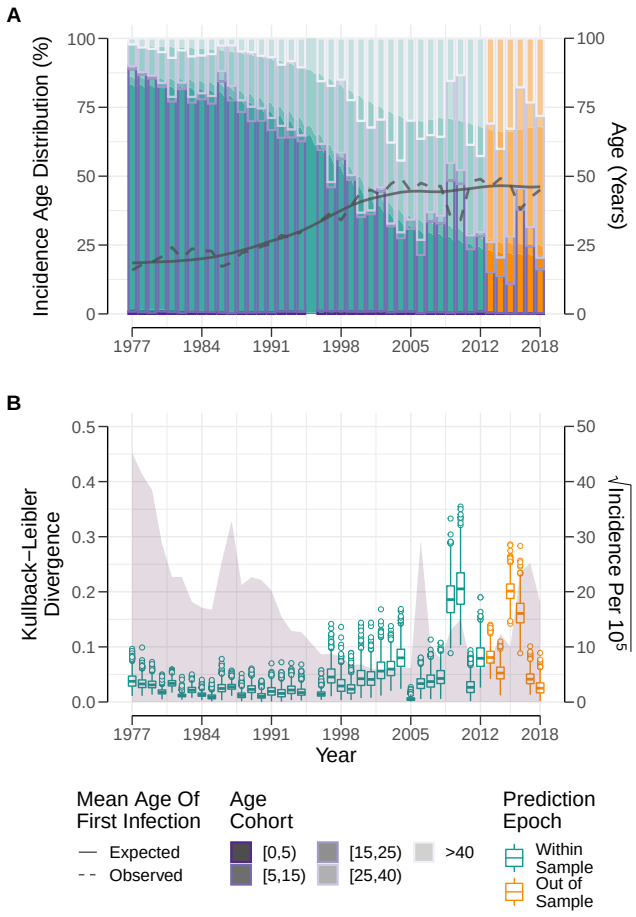

Figure 5.3: **Quantitative comparison of mumps incidence age distribution.** (A) Qualitative comparsion between expected, assuming the model of waning immunity, and the observed incidence per 10^5 age-distribution (purple gradient, left, y-axis). Simulated incidence is broken down into the two prediction epochs - within sample (seagreen gradient) and out of sample (orange gradient) as in figure 5.2. Lines represent comparison of temporal shifts in the mean age of infection (right, y-axis) under the waning model (dotted) and observed incidence (dashed). (B) Quantitative agreement between observed *v/s.* expected age-distribution of incidence using the Kullback-Leibler Divergence ($\overline{D}^{KL}$, left y-axis) through time. Boxplots represent bootsrapped distribution of $\overline{D}^{KL}$ calculated by comparing observed age-distribution to 1000 synthetically generated time-series under the estimated observation noise. The area plot represents an estimate of age-aggregated mumps incidence per 10^5 (right, y-axis).

In addition to exploring the performance of the model at recapitulating the time series for each individual age group, we also examined how well our model captures the age profile of cases. A salient recurring feature common to the majority of the recent epidemics is the upward shift in the mean age of infection in the reported cases relative to the pre-vaccine or early vaccination era [6, 27, 142]. We find our model accurately tracks observed trends in the mean age of infection calculated (Fig. 5.3A). To further explore how well our fitted model reproduces the full age distribution of cases, we compared the dynamics of the estimated true age distribution to the model simulated age distribution (Fig. 5.3A, left y-axis). The waning model was able to reproduce the dynamics in the age-distribution with strikingly good overall accuracy. To quantify relative agreement between the observed and the expected age distributions, we calculated the Kullback-Leibler Divergence ($\overline{D}^{KL}$, Fig. 5.3B). We observed a systematic increase in the bootstrap variance of the $\overline{D}^{KL}$ over time, becoming pronounced in the most recent years when incidence has been low. We speculate that this is likely due to the increased impact of demographic stochasticity during eras of low prevalence.

5.3.3 Long term trends in population immunity explain the US mumps resurgence.

The results of our statistical inference allow us to reconstruct changes in the immune profile of the US population after the introduction of vaccines (Fig. 5.4). Using the fitted model, we estimated a long average duration of vaccine derived immune protection of 111.5y, 95% CI: (111.4, 3×10^4)y. While this is longer than the typical human lifespan, our results favour an exponential distribution for the duration of immunity. Consequently, an average duration of 111.5y implies that 80% of individuals remain immune for at least 20 years after their last dose and that less than 50% remain immune after 80 years. Importantly, our results show that herd immunity is not possible using the existing vaccine, as the fraction of vaccinated individuals who remain immune drops below 93% (the approximate herd immunity threshold calculated via $1 - 1/R_0$) within 8 y of the last admin-

istered vaccine dose. These findings clearly identify a role for age-specific boosting schemes.

By incorporating vaccine uptake and demographic data, our model is also able to reconstruct the proportion of the population with vaccine immune protection by age cohort through time (Fig. 5.4B), and the number of individuals who lose immunity each week (Fig. 5.4E). Our results suggest that proportion of the US population over 25 who are susceptible to mumps (composed of individuals whose immunity has waned or have not received the full two-dose schedule vaccine) has gradually risen over time reaching around 16% by 2000 (Fig. 5.4D). From our reconstructed susceptible profile we can calculate the effective reproductive number over time (Fig. 5.4C), finding that the effective reproductive number remained just above 1 (characteristic of an endemic disease) until the 1985 resurgence, after which it remained below 1 until the late 90s when it entered an oscillatory phase driven by susceptible build up punctuated by recurrent outbreaks (see also Fig. 5.4F).

5.3.4 Sensitivity of the relative prevalence to the duration of immunity.

To evaluate the sensitivity of our results, we performed a simulation study to investigate how the frequency of breakthrough infections relative to naïve infections is affected by the duration of vaccine protection. Due to the difficulty in interpreting the average immune duration when it is exponentially distributed, we instead quantify vaccine protection terms of the proportion of the vaccinated population who have lost immunity by age 18, which we denote P_{18}. Complementing the results shown in Fig. 5.4A, we find that P_{18} exceeds 6% even when the average immune duration is 200 years.

Our model allowed us to assess the impact of waning immunity on the individual level, and quantify the ratio of infections in individuals whose vaccine immunity has waned (I_W) to those in unvaccinated individuals (including non-vaccinated and those with primary vaccine failure) (I_S), which we refer to as the relative prevalence ratio (RPR). We calculate the

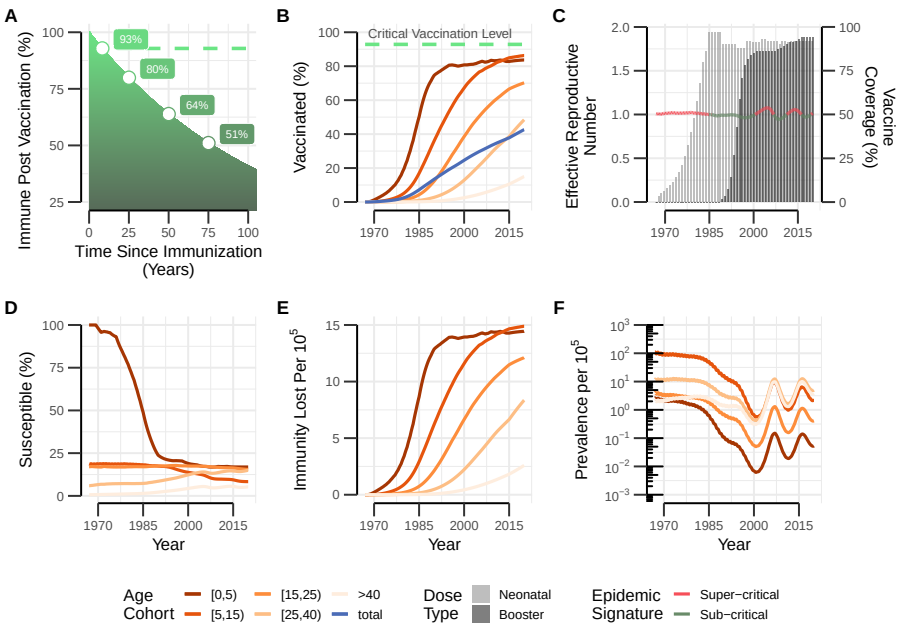

Figure 5.4: **Reconstructed population immune profile and impact on mumps resurgence.** (A) Proportion of vaccinated individuals retaining immunity over time for the exponential waning. The average duration of immunity was fixed to $1/\delta$ =111.5 years (i.e. the maximum likelihood estimate). The green dashed line indicates the approximate critical vaccination level for the exponential waning model. Here the critical vaccination threshold was approximated using $1-1/R_0$. (B) Percent of the population with vaccine immune protection by age cohorts over time. (C) Dynamics in the effective reproductive number highlighting epidemic transitions between supercritical (red) and subcritical (green) phases (left, y-axis). Annual vaccination rates of neonatal (light green) and booster (dark blue) doses (right, y-axis). (D) Time series of age-specific susceptibility profiles.(E) Number of individuals who lose immunity each week per 10^5. (F) Weekly mumps prevalence per 10^5 individuals in each age cohort ($10^5 * I_a/N_a$)

RPR separately for each each age cohort based on the time since the last vaccine dose received by vaccinated individuals.

Apart from the [0, 5) cohort, the RPR was always greater than 1, implying more infections in previously vaccinated individuals than the unvaccinated (Fig. 5.5A). For the younger age groups $\big([0, 5)$ y and $[5, 15)$ y$\big)$, increasing

the average duration of immunity (i.e. decreasing P_{18}) always resulted in a drop in the RPR. For long-lasting immunity, there was a general trend that as the population aged, RPR rose due to waning immunity providing a mechanism for susceptible recruitment unavailable to the unvaccinated compartment. Interestingly, unlike in the case of the younger age classes, we observed a non-monotonic effect of waning intensity on the RPR of the older age-groups $\left(>15 \text{ y}\right)$. For these older age cohorts, under a rapid waning scenario, the RPR was comparable to a low waning intensity. This, we posit, is a consequence of contesting flows of waning and aging among the individuals in the population. For the scenario with long lasting immunity (Fig. 5.5A, purple region), most of the vaccinated individuals remained protected against infection during their life-time. Conversely, with a vaccine derived immunity that was short-lived (Fig. 5.5A, orange region), individuals rapidly lost protection and were soon infected. The aging in relation to the transmission process in this scenario occurs at a substantially lower rate. Susceptible individuals who have lost their immunity get infected, and have recovered from the infection by the time they graduate from an age class. This effect was exacerbated with the wider age-cohorts, generating a greater overlap of the RPR for a vaccine with longer immune duration compared with a scenario where the immunity wanes rapidly. Highest values of RPR, for these older age classes were therefore produced at intermediate values of immune loss probability. This phenomenon has also been documented in case of other disease systems like pertussis [48] and in the recent and ongoing pandemic of the novel SARS-CoV-2 [76].

It is important to note that since we are considering the ratio of I_W/I_S, and not taking into account the fraction of the population who received the vaccine, the RPR can exceed 1 indicating that more infections are taking place in individuals whose immunity has waned than were unvaccinated. Indeed, the RPR for $P_{18} = 1$ (which corresponds to a completely ineffective vaccine) reflects the vaccine uptake.

Clinically, the observation of infections predominantly among previously vaccinated individuals can be misconstrued to imply that vaccines are ineffective at controlling transmissible pathogens. To rebut this claim, we

calculated the population level vaccine impact as a function of varying waning intensities. Overall, we observed that a decrease in the waning intensity substantially decreases infection prevalence (Fig. 5.5C) confirming that vaccines which impart long-lasting immunity are indeed impactful in the control of pathogens like mumps, and that having a larger proportion of infections represented in the vaccinated sub-population is merely an expected artifact of the waning immunity at a population level. For our system, however, we estimated a relatively modest vaccine impact (59.1%) at the population level. We speculate that this might be a joint outcome of high viral transmissibility ($R_0 = 14$) and a substantial primary vaccine failure probability ($\alpha = 0.054$).

5.4 Discussion

The recent re-emergence and continued persistence of mumps in countries with high estimated vaccine coverage has been perplexing, casting doubt on global health goals [122]. A major obstacle in this context is the absence of serological correlates of protection against mumps. Indeed, it still remains unclear what components of the host immune system are adaptively immunogenic against subsequent infections and what constitutes an infectious dose for a successful transmission event for entry and establishment of an infection [185].

In the absence of straightforward immunological indicators of protection, we attempted to disentangle the putative mechanisms underlying the re-emergence of mumps in the US by statistically contending compartmental models describing mechanisms for vaccine failure and their effect on the circulation of the mumps virus. Among the models contested, we found that the exponential waning model is most consistent with the observed data.

Importantly, our exponential waning model reconciles population-level immunological trends with seemingly contradictory individual-level clinical observations. Specifically, while a long average immune duration of immunity is estimated and around 50% of individuals retain immunity for their entire lives, we estimate 11.77% (95% CI: 0.1%, 11.81%) of vacci-

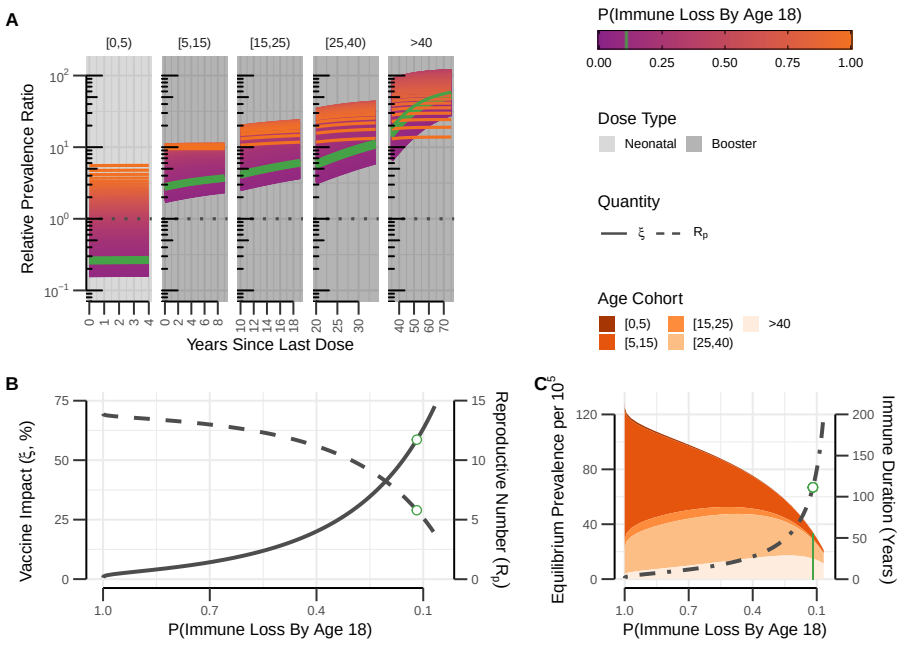

Figure 5.5: **Age-specific immunity post vaccination** (A) Relative prevalence ratio $\left(I_i^w(t)/I_i^s(t)\right)$ of infection after vaccination across the five age-cohorts (facet columns), as a function of varying probability of immunity loss by the age of 18 years (colour gradient). Estimated MLE of immune duration was converted to probability of immune loss (indigo); Time underwent a reset to 0 years after administration of neonatal dose (grey background) and booster dose (black background). (B) Population level vaccine impact (solid lines) and reproductive number (dashed lines) calculated at 91.7% and 91.6% neonatal and booster vaccine coverage respectively (fixed at an average for 2008-2018) as a response, and (C) Stable-state prevalence distribution across five age-cohorts as a function of varying probability of immune loss by age 18. Relationship between duration of immunity and Probability of immune loss by age 18 are represented on the secondary axis (dotdashed lines). Covariate values were fixed at the last known value in the year 2018. Dynamics were simulated for 300 y and final values in the infectious compartments values we taken to be prevalence per 10^5

nated individuals lose immunity by the age of 18 y. Given persistent high coverage, this explains the paradoxical clinical observation that most infections occur in vaccinated individuals. Furthermore, our results explain the resurgent dynamics and the shift in the age profile of cases. As the

immune profile shifts from being mostly due to natural infection (which we take to be lifelong, consistent with a lack of documented reinfections) to vaccine-derived, the impact of waning immunity rises. This leads to a growing pool of older susceptible individuals with waned vaccinal immunity (who in the pre-vaccine era would likely have natural immunity) among which the virus can spread. Such multi-decadal trends can result in the initial achievement of herd immunity (when only younger individuals have received vaccines and most older individuals possess natural immunity) before it is lost (as the vaccinated individuals age and their immunity wanes).

We found that the leaky scenario is not consistent with the available epidemiological data. While a small but non-zero leakiness parameter was able to generate outbreaks matching the timing and size of post-2006 resurgence, our results show that if leakiness is the mechanism of re-emergence then the rise in incidence should be approximately uniformly realized across all age-classes. Assuming a constant contact structure and an imported genotype with transmissibility identical to the previously circulating genetic variant, these dynamics should stabilize to the pre-vaccine age-distribution of cases. Instead, we documented a gradual increase in the mean age of first infection in the observed mumps dynamics. This pattern is consistent with the waning hypothesis. The limited spread and the age profile of cases during recent resurgences place constraints on the possible role of immune mismatch.

Although our exponential waning model successfully explains the long-term trends in mumps epidemiology in the US, it fails to capture some of the finer temporally resolved characteristics of the data in the two resurgence eras (1985-1989 & 2000-2012). Particularly, the average simulated dynamics (1) underestimate the peak and overestimate the duration of the mid-80s re-emergence, and (2) they combine outbreaks in 2006 and 2010 into a single multi-year outbreak. In both scenarios, we believe model-data disagreements are the result of spatial heterogeneity and stochasticity in the disease transmission process (when incidence is very low [96]), which our deterministic and spatially aggregated model is not designed to capture. In both the resurgent eras - 1985–1989, and 2006-2012, mumps incidence

in mid-western states surrounding Illinois and Iowa respectively contribute disproportionately to overall cases observed (Fig. 5.1D–E). Another model limitation is our parsimonious assumption that vaccine leakiness (i.e. the probability that a mismatched virus can infect a vaccinated individual) does not depend on age. Given the leaky model is outperfomed by the exponential waning model, it would be interesting to examine whether its performance improves if the impact of immune evasion depends on either age or the time since the last vaccine dose.

Our findings are qualitatively consistent with those presented in two recent studies exploring drivers of re-emergence of mumps in the vaccine era. Lewnard & Grad successfully reproduce the age-distribution of observed cases in the 2006 mumps epidemic using a model of waning efficacy fitted to vaccine trial data [115]. Similarly, Wohl *et al.* [208] found no evidence in support of immune evasion among the mumps genotypes that dominate circulation in the northern US and concluded that waning vaccine immunity is the plausible explanation for mumps resurgence. Interestingly, by doing the careful book-keeping of susceptible recruitment dynamics at the national level, we arrive at a very similar conclusion, albeit with a much longer estimated duration of vaccine-derived immunity than reported in these prior studies. Our results bolster the confidence that the waning hypothesis is indeed the most plausible driver of the ongoing resurgence of mumps. Our study, like the previous ones implicates the roll-out of a third booster dose of mumps vaccine [60, 109], and recommends periodic targeted immune boosting of immunity among college going adults [31].

Our analysis suggests waning immunity provides a parsimonious explanation for the resurgence of mumps in the US. We found that, due to the combination of waning immunity and primary vaccine failure, robust herd immunity is unable to be achieved with present vaccine schedules. Indeed, our results suggest that the present vaccine requires booster doses every 8 years to achieve and maintain herd immunity. An area of priority for future research is the use of empirically-validated models such as ours to identify optimal, age-targeted vaccination schemes with the view to control, and eventually eliminate mumps circulation in the US.

5.5 Supplementary Results

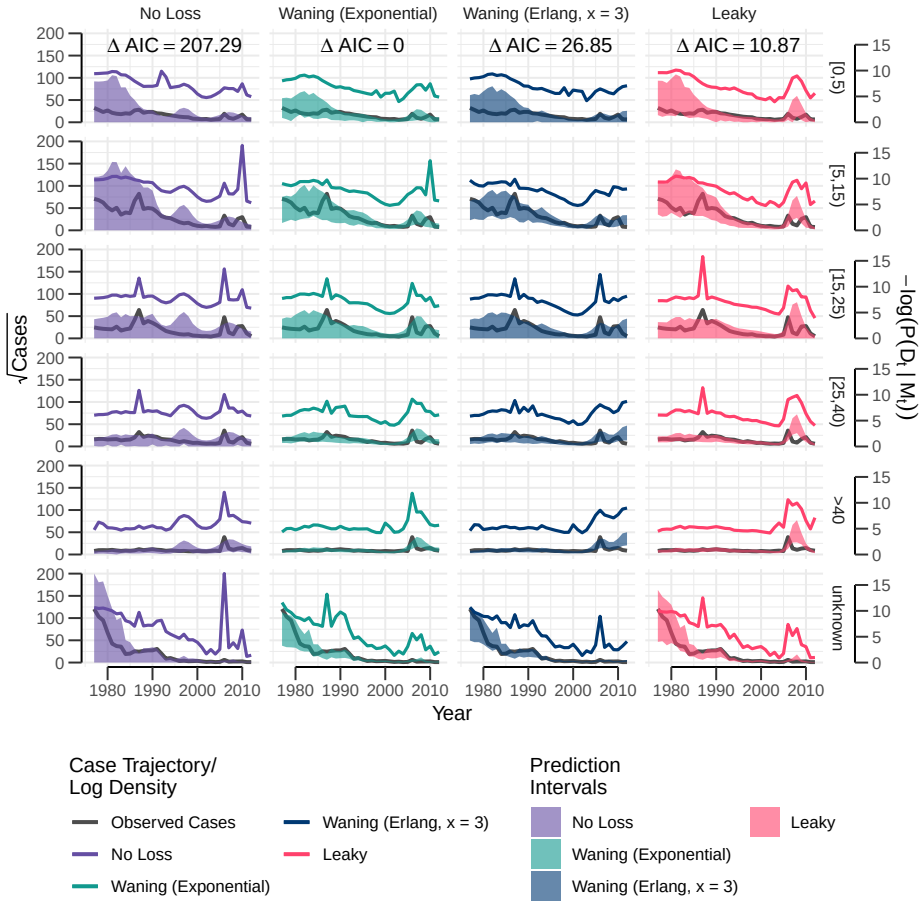

Figure 5.6: **Relative Model Fit.** Qualitative agreement of age-specific observed cases of mumps (facet rows, black) to 4 models of vaccine imperfection (facet columns) - No Loss (violet), Waning (Exponential, seagreen), Waning (Erlang, x = 3, darkblue), and Leaky (pink). Colored solid lines represent negative log probability density for the four models at the MLE. Ribbons represent 80% prediction intervals from the model simulation. Relative quantitative model-data agreement has been conveyed using Akaike information criterion (ΔAIC, reported inset) among various models.

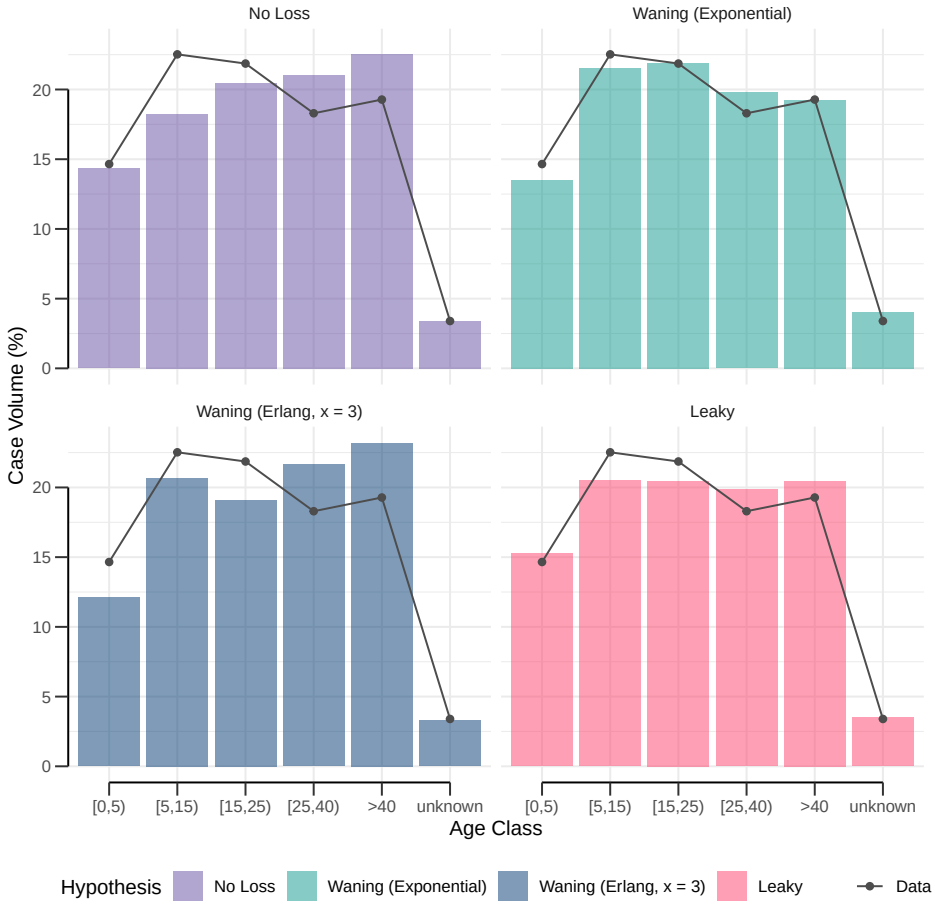

Figure 5.7: **Relative distribution of average mumps cases.** Barplots represent the mean age-distribution of cases during 2000-2018, calculated using the median trajectories for the 4 models of vaccine imperfection – No Loss (violet), Waning (Exponential, seagreen), Waning (Erlang, x = 3, darkblue), and Leaky (pink). Solid lines represent observed case distribution over the same time period.

Table 5.2: 95% Bootstraped Confidence Intervals

	No Loss		Waning (Exponential)		Waning (Erlang, x = 3)		Leaky	
Parameter/Quantity	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
R_0	35.5052	43.9744	8.0209	28.2023	27.2806	51.1304	21.1990	34.4816
R_p	0.4318	0.5652	0.1109	8.5538	8.2783	25.3014	0.5804	1.6774
ξ	0.9857	0.9898	0.5633	0.9938	0.2422	0.7883	0.9469	0.9814
δ^{-1} (Years)	-	-	111.4000	37357.4656	31.1158	102.2840	-	-
ε	-	-	-	-	-	-	0.0125	0.0483
β_1	0.0214	0.6431	0.0000	0.1348	0.0001	0.8518	0.3842	0.8423
σ^{-1} (Days)	10.2939	13.4371	12.7643	66.9263	8.1108	279.6526	20.5322	93.9482

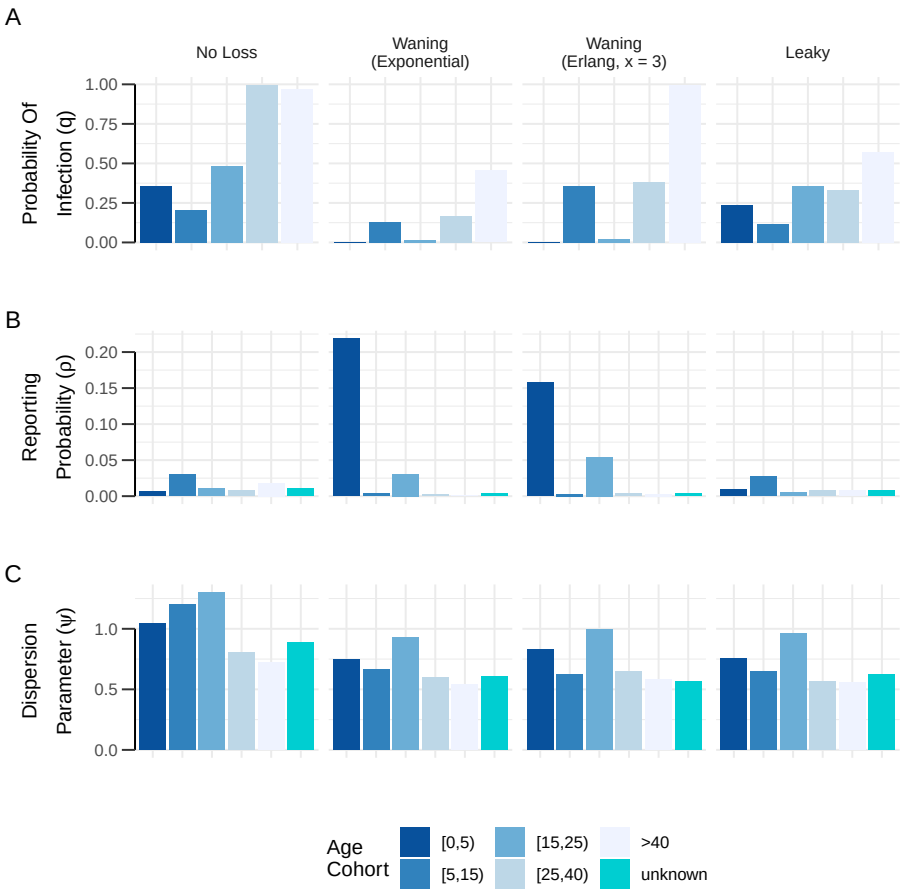

Figure 5.8: **Age-specific parameter estimates of Process and Observation model.** Faceted barplots represent maximum likelihood estimates of probability of infection (q), reporting probability (ρ) and dispersion parameter (ψ) corresponding to the four hypotheses of mumps reemergence (columns). Age-class has been reported using a blue gradient.

CHAPTER 6

CONCLUSION

Studying infectious diseases is especially challenging on account of the three main reasons: (1) Most of the mechanistic processes that determine the transmission dynamics of a pathogen remain hidden from view in a way that renders straightforward inference very difficult *e.g.* readily available epidemiological data are usually time-series of case incidence for a disease spreading through the population. Commonly these time-series can be thought of as an under-reported subset of the underlying disease prevalence sampled over a particular censual interval [19, 91]. This under-reporting and the irregularity in the sampling can be associated with the sampling effort that is proportional to disease prevalence and severity [101].

Until the pre-pandemic years, however, it was difficult to acquire datasets on temporally varying statistical estimates of immunity, exposure, and/or infectivity at the population level. (2) Many of disease systems have inherent heterogeneities that can generate conditions for effective proliferation and long term in persistence of a pathogen, *e.g.*, some individuals contribute disproportionately towards generating secondary infections within a population [119, 209], assortative social mixing among age cohorts within a population [135], asymmetric geographic coupling among cities, states, countries, [41], cellular biological processes, dictated by the circadian rhythm, births, deaths, migration, environmental conditions are more or less inconstant over time [166]. (3) Infectious diseases are

mainly a multi-scale process that operate at the cellular level but can have cascading effects observed at the planetary scales [69].

In this work I have employed a combination of mathematical models and likelihood-based statistical methods to conduct inference on the complex, partially observed systems of infectious diseases. To demonstrate the width and the flexibility of this inferential paradigm, I have investigated 3 problems surrounding viral infections that are set at distinct scales of biological organization. This enabled me to formalize heterogeneities that are intrinsic to a particular scale of biological organization. Their incorporation within a theoretical model allowed me to understand their role in driving dynamics across these disease systems.

For all the problems described in this work, I began by formulating ecological hypotheses about mechanisms that were plausibly generative of the observed dynamics. These hypotheses were formalized within mathematical models incorporating varying complexities and structures, thereby allowing me to describe a multitude of channels that could have led to production the observed data. For the within-host analysis, investigating the differential role of the host's cellular biology and immunology in controlling viral dynamics of the SARS-CoV-2 and H3N2, I formulated several models involving mechanisms about cellular production of target cells, role of temporal lags between viral infection and production, and multiple functional mechanisms of the innate immune system including reduction in target cell susceptibility, or increased cytoxicity of the infected cells. When assessing the atypical 2019 influenza B season in Massachusetts, I investigated the the role of short-term facilitative interaction, among influenza types, contested against the inherent increase in the type specific transmissibilty between these viruses. Finally, while investigating the recent mumps resurgence in the US, necessitated the formulation of hypotheses of vaccine imperfection and their role in stoking the recent, ongoing epidemics of mumps in the US. I assessed the relative agreement of simulated dynamics with observed data, generated from various models vaccine imperfection.

To introduce realism in the simulated dynamics, models were informed by incorporating time inhomogeneous forcing, technically referred to as

time-varying covariates. These external forcing sources were in the form of seasonal forcing in case of the interaction between influenza A and B. In case of mumps reemergence, model dynamics were augmented by introducing age-specific contact rates and population demography incorporated as covariates. To account for the variation in the age-specific population sizes, covariates corresponding to various demographic processes like, birth, deaths migration were also incorporated within the model formulation. In addition to this, variability in the reporting protocols was also formulated by introducing a time varying reporting process, specified within the observation model. Incorporating covariates to inform models allowed to disentangle the effect of the actual mechanism of interest from the effect of variation due the extrinsic factors *e.g.*, demography, seasonality on the model dynamics. This resulted in improving the model-data agreement as the model dynamics could better track the variation in the observed data.

Depending upon the type of problem, an appropriate statistical model to denote the observation process was chosen in conducting model selection. Statistical models were also chosen based on their agreement with the observed data. Within the analysis of the mumps reemergence, it was observed that a model that allowed for higher variability in the observation was necessary to provide a good fit to the data. This might be because, in the vaccine era, there is low incidence of mumps which was explained better with a model with a small but non-zero value of dispersion. Conversely, A Poisson observation model was selected while analysing the atypical seasonality of influenza B.

To arbitrate among the numerous hypotheses and identify a likely mechanism underlying the observed time-series, a goodness of fit measure was quantified by maximizing the likelihood function. Interestingly, for the within-host viral time-series, this analysis led to very different mechanisms being selected for the two viruses. For SARS-CoV-2, a substantial role of the innate immune system in controlling the infection process was identified. Whereas, a parsimonious model of susceptible depletion was selected to be the best fitting explanation in driving the infection dynamics for influenza. Very different within-host, virus-specific $R_0 s$ were estimated for

the two viruses, with SARS-CoV-2 possessing nearly 10^2 times the proliferation potential than H3N2. One reason for the selection of simpler model of infection for influenza might be the lack of the data pertaining to other systems states, especially covariates capturing the temporal variation immune molecules, such as type-I interferon, that drive the innate immune response. Analyzing parallel data streams depicting unobserved immune compartments might help improve the inference about the role these immune mechanisms in modulating H3N2 infection in this ferret model.

Furthermore, it might also be interesting to conduct a challenge experiment with similar infection protocol, in genetically modified, animal subjects with diminished ability to produce an interferon response. This might provide a robust test of validity in support of the results generated from this modeling study [34]. For this modified experiment, if the type-I interferon related immune responses are indeed important as observed in cases a much larger overall virus load should be recorded for infection with SARS-CoV-2 infection. On the other hand, no changes in the viral dynamics, on average, would be observed in these immuno-compromised animals for H3N2 infection. In this manner, using the ability of the models to generate counterfactual scenarios can also be utilized in extending and designing novel experiments.

The results from trajectory matching can depend on the optimization routine used to maximize the objective function. For the problem of the atypical influenza season, statistical evidence favoring cross-protection hypothesis between the two influenza sub-types was not found. Like in the second chapter, where all complex models collapse to the simple model, the cross-protection model should have collapsed to the neutral model, thus only incurring a penalty for bearing additional parameters. However, the optimization software that was used in this analysis prohibited from specifying perfectly 0 values to the parameters [184]. To adjust for that, a very small, yet non-zero value of cross-protection (χ) and rate of recovery when in the cross-protective class (ϕ) was specified as the lower bound for research. This violated the nesting of the neutral model within the cross-protection model. To supplement the result from the trajectory matching,

a bifurcation analysis based on relative transmissibility (R_0^B/R_0^A) and intensity of cross-protection (χ) was also conducted. This analysis also arrived at a similar conclusion that over the short time-scale, as considered in this analysis, intrinsic transmissibility differences among influenza types have a much higher impact on the system dynamics than cross-protection intensity. Thus models can be used to assess the system sensitivity to control parameter values. Such a multi-faceted analysis can thus reinforce our confidence in the conclusions of the study.

This work provides a compendium of mathematical models and likelihood-based statistical inference, applied to exploration of complex mechanisms driving infectious disease transmission. The methods described in this document can be extended in a straightforward manner to tackle problems involving integrating effects of multiple scales of biological organization.

BIBLIOGRAPHY

1. Achenbach, J., Johnson, C. Y., Sun, L. H. & Shammas, B. *'Goldilocks virus': Delta vanquishes all variant rivals as scientists race to understand its tricks* Aug. 2021. https://www.washingtonpost.com/health/delta-variant-most-prevalent/2021/08/08/d1017f0e-f558-11eb-9068-bf463c8c74de_story.html.
2. Albrecht, R. A. *et al.* Moving forward: recent developments for the ferret biomedical research model. *mbio* **9,** e01113–18 (2018).
3. Altizer, S. *et al.* Seasonality and the dynamics of infectious diseases. *Ecology letters* **9,** 467–484 (2006).
4. An, D. *et al.* Protection of K18-hACE2 mice and ferrets against SARS-CoV-2 challenge by a single-dose mucosal immunization with a parainfluenza virus 5–based COVID-19 vaccine. *Science advances* **7,** eabi5246 (2021).
5. Anderson, D. & Watson, R. On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67,** 191 (1980).
6. Anderson, R. M. & May, R. M. Directly transmitted infections diseases: control by vaccination. English. *Science* **215,** 1053–1060 (Feb. 1982).
7. Anderson, R. & May, R. *Population biology of infectious diseases. Report of the Dahlem workshop on population biology of infectious disease agents Berlin 1982, March 14–19* (Springer, 1982).
8. Anderson, R. M., Anderson, B. & May, R. M. *Infectious diseases of humans: dynamics and control* (Oxford university press, 1992).

9. Ardia, D., Boudt, K., Carl, P., Mullen, K. M. & Peterson, B. G. Differential Evolution with DEoptim: An Application to Non-Convex Portfolio Optimization. *R Journal* **3,** 27–34 (2011).

10. Arulampalam, M. S., Maskell, S., Gordon, N. & Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing* **50,** 174–188 (2002).

11. Ashby, B. & Gupta, S. Sexually transmitted infections in polygamous mating systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368,** 20120048 (2013).

12. Baccam, P. *et al.* Kinetics of Influenza A Virus Infection in Humans. *Journal of Virology*. https://journals.asm.org/doi/full/10.1128/JVI.01623-05 (Aug. 2006).

13. Ballings, M. *Package 'AUC'* PhD thesis (UGent, 2013).

14. Barskey, A. E., Glasser, J. W. & LeBaron, C. W. Mumps resurgences in the United States: A historical perspective on unexpected elements. *Vaccine* **27,** 6186–6195. ISSN: 0264-410X. http://www.sciencedirect.com/science/article/pii/S0264410X09011803 (2018) (Oct. 2009).

15. Beauchemin, C. A. & Handel, A. A review of mathematical models of influenza A infections within a host or cell culture: lessons learned and challenges ahead. en. *BMC Public Health* **11,** S7. ISSN: 1471-2458. http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-11-S1-S7 (2020) (2011).

16. Bedford, T. *et al.* Integrating influenza antigenic dynamics with molecular evolution. *elife* **3,** e01914 (2014).

17. Bharti, N. *et al.* Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334,** 1424–1427 (2011).

18. Biron, C. A., Nguyen, K. B., Pien, G. C., Cousens, L. P. & Salazar-Mather, T. P. NATURAL KILLER CELLS IN ANTIVIRAL DEFENSE: Function and Regulation by Innate Cytokines. en. *Annual Review of Immunology* **17,** 189–220. ISSN: 0732-0582, 1545-3278.

http://www.annualreviews.org/doi/10.1146/annurev.immunol.17.1.189 (2021) (Apr. 1999).

19. Bjørnstad, O. N. & Grenfell, B. T. Noisy clockwork: time series analysis of population fluctuations in animals. *Science* **293,** 638–643 (2001).

20. Black, A. J. & McKane, A. J. Stochastic formulation of ecological models and their applications. *Trends in ecology & evolution* **27,** 337–345 (2012).

21. Bonhoeffer, S., May, R. M., Shaw, G. M. & Nowak, M. A. Virus dynamics and drug therapy. *Proceedings of the National Academy of Sciences* **94,** 6971–6976 (1997).

22. Borchering, R. K. *et al.* Anomalous influenza seasonality in the United States and the emergence of novel influenza B viruses. *Proceedings of the National Academy of Sciences* **118** (2021).

23. Bouvier, N. M. Animal models for influenza virus transmission studies: a historical perspective. *Current opinion in virology* **13,** 101–108 (2015).

24. Caini, S. *et al.* The epidemiological signature of influenza B virus and its B/Victoria and B/Yamagata lineages in the 21st century. *PloS one* **14** (2019).

25. Callahan, J. D. *et al.* Use of a portable real-time reverse transcriptasepolymerase chain reaction assay for rapid detection of foot-and-mouth disease virus. *Journal of the American Veterinary Medical Association* **220,** 1636–1642 (2002).

26. Callaway, E. *Beyond Omicron: what's next for COVID's viral evolution* 2021.

27. Canada, P. H. A. o. *Government of Canada* Nov. 2021. https://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-4-active-vaccines/page-14-mumps-vaccine.html.

28. Canini, L. & Perelson, A. S. Viral kinetic modeling: state of the art. *Journal of pharmacokinetics and pharmacodynamics* **41,** 431–443 (2014).

29. Cao, P. *et al.* Innate Immunity and the Inter-exposure Interval Determine the Dynamics of Secondary Influenza Virus Infection and Explain Observed Viral Hierarchies. en. *PLOS Computational Biology,* 28 (2015).

30. Cao, Y., Gillespie, D. T. & Petzold, L. R. Avoiding negative populations in explicit Poisson tau-leaping. *The Journal of chemical physics* **123,** 054104 (2005).

31. Cardemil, C. V. *et al.* Effectiveness of a third dose of MMR vaccine for mumps outbreak control. *New England Journal of Medicine* **377,** 947–956 (2017).

32. Carrat, F. *et al.* Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American journal of epidemiology* **167,** 775–785 (2008).

33. *Census* https://www.census.gov/data/data-tools.html.

34. Chabas, H. *et al.* Evolutionary emergence of infectious diseases in heterogeneous host populations. *PLoS biology* **16,** e2006738 (2018).

35. Chan, K. F. *et al.* Investigating viral interference between influenza A virus and human respiratory syncytial virus in a ferret model of infection. *The Journal of infectious diseases* **218,** 406–417 (2018).

36. Chatterjee, A., Vlachos, D. G. & Katsoulakis, M. A. Binomial distribution based τ-leap accelerated stochastic simulation. *The Journal of chemical physics* **122,** 024112 (2005).

37. Christoforos, H. *et al.* Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. *Journal of The Royal Society Interface*. https://royalsocietypublishing.org/doi/10.1098/rsif.2016.0289 (June 2016).

38. Clemens, E. B., Van de Sandt, C., Wong, S. S., Wakim, L. M. & Valkenburg, S. A. Harnessing the power of T cells: the promising hope for a universal influenza vaccine. *Vaccines* **6,** 18 (2018).

39. Cozza, V. *et al.* Global seasonal influenza mortality estimates: A comparison of 3 different approaches. *American journal of epidemiology* **190,** 718–727 (2021).

40. Danchev, V. & Porter, M. A. Neither global nor local: Heterogeneous connectivity in spatial network structures of world migration. *Social Networks* **53,** 4–19 (2018).

41. Danon, L., Read, J. M., House, T. A., Vernon, M. C. & Keeling, M. J. Social encounter networks: characterizing Great Britain. *Proceedings of the Royal Society B: Biological Sciences* **280,** 20131037 (2013).

42. DaPalma, T., Doonan, B. P., Trager, N. M. & Kasman, L. M. A systematic approach to virus–virus interactions. *Virus Research* **149,** 1–9 (Apr. 2010).

43. Al-Dar, A. A., Al-Qassimi, M., Ezzadeen, F. H., Qassime, M., Ghaleb, Y., *et al.* Diphtheria resurgence in Sada'a-Yemen, 2017–2020. *BMC infectious diseases* **22,** 1–6 (2022).

44. Dawood, F. S. *et al.* Interim estimates of 2019–20 seasonal influenza vaccine effectiveness—United States, February 2020. *Morbidity and Mortality Weekly Report* **69,** 177 (2020).

45. Dayan, G. H. & Rubin, S. Mumps Outbreaks in Vaccinated Populations: Are Available Mumps Vaccines Effective Enough to Prevent Outbreaks? en. *Clinical Infectious Diseases* **47,** 1458–1467. ISSN: 1058-4838, 1537-6591. https://academic.oup.com/cid/article-lookup/doi/10.1086/591196 (2019) (Dec. 2008).

46. Devore, J. L. *Probability and Statistics for Engineering and the Sciences* (Cengage learning, 2011).

47. Diekmann, O., Heesterbeek, J. & Roberts, M. G. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface* **7,** 873–885 (2010).

48. Domenech de Cellès, M., Magpantay, F. M., King, A. A. & Rohani, P. The impact of past vaccination coverage and immunity on pertussis resurgence. *Science translational medicine* **10,** eaaj1748 (2018).

49. Dopico, X. C. *et al.* Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nature communications* **6,** 1–13 (2015).

50. Du, X. *et al.* Mapping of H3N2 influenza antigenic evolutionin China reveals a strategy for vaccine strain recommendation. *Nature Communications* **3,** DOI: 10.1038 –ncomms1710. https://www.nature.com/articles/ncomms1710.pdf (May 2019).

51. Dudas, G., Bedford, T., Lycett, S. & Rambaut, A. Reassortment between influenza B lineages and the emergence of a coadapted PB1–PB2–HA gene complex. *Molecular biology and evolution* **32,** 162–172 (2015).

52. Dumm, R. E. *et al.* Non-lytic clearance of influenza B virus from infected cells preserves epithelial barrier function. *Nature communications* **10,** 1–15 (2019).

53. Edwards, A. W. F. *Likelihood* (CUP Archive, 1984).

54. Eiben, A. E., Smith, J. E., *et al. Introduction to evolutionary computing* (Springer, 2003).

55. Elveback, L. & Fox, J. Illness, Oral Poliovirus Vaccine, And Interference. *Archives of environmental health* **9,** 724–726 (May 1964).

56. Ferdinands, J. M., Alyanak, E., Reed, C. & Fry, A. M. Waning of influenza vaccine protection: Exploring the trade-offs of changes in vaccination timing among older adults. *Clinical Infectious Diseases* **70,** 1550–1559 (2020).

57. Ferdinands, J. M. *et al.* Intraseason waning of influenza vaccine protection: evidence from the US influenza vaccine effectiveness network, 2011–2012 through 2014–2015. *Clinical Infectious Diseases* **64,** 544–550 (2017).

58. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437,** 209–214 (2005).

59. Ferguson, N. M. *et al.* Benefits and risks of the Sanofi-Pasteur dengue vaccine: Modeling optimal deployment. *Science* **353,** 1033–1036 (2016).

60. Fiebelkorn, A. P. *et al. Mumps antibody response in young adults after a third dose of measles-mumps-rubella vaccine* in *Open forum infectious diseases* **1** (2014).

61. *FluView Interactive* Accessed: 2020-02-28. Feb. 2020. https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html.

62. Francis, M. E., King, M. L. & Kelvin, A. A. Back to the Future for Influenza Preimmunity—Looking Back at Influenza Virus History to Infer the Outcome of Future Infections. *Viruses* **11,** 122 (2019).

63. Funk, S. Socialmixr: social mixing matrices for infectious disease modelling (2018).

64. Gagniuc, P. A. *Markov chains: from theory to implementation and experimentation* (John Wiley & Sons, 2017).

65. Galazka, A. M., Robertson, S. E. & Kraigher, A. Mumps and mumps vaccine: a global review. *Bulletin of the World Health Organization* **77,** 3 (1999).

66. Georgioudakis, M. & Plevris, V. A comparative study of differential evolution variants in constrained structural optimization. *Frontiers in Built Environment* **6,** 102 (2020).

67. Giannakis, E. & Papadas, C. T. Spatial Connectivity and Regional Economic Resilience in Turbulent Times. *Sustainability* **13,** 11289 (2021).

68. Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics* **115,** 1716–1733 (2001).

69. Gog, J. *et al. Seven challenges in modeling pathogen dynamics within-host and across scales. Epidemics. 2015; 10: 45–48*

70. Gostic, K. M., Ambrose, M., Worobey, M. & Lloyd-Smith, J. O. Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. *Science* **354,** 722–726 (2016).

71. Gouma, S. *et al.* Two major mumps genotype G variants dominated recent mumps outbreaks in the Netherlands (2009-2012). eng. *The Journal of General Virology* **95,** 1074–1082. ISSN: 1465-2099 (May 2014).

72. Gouma, S. *et al.* Differences in antigenic sites and other functional regions between genotype A and G mumps virus surface proteins. en. *Scientific Reports* **8,** 13337. ISSN: 2045-2322. http://www.nature.com/articles/s41598-018-31630-z (2019) (Dec. 2018).

73. Grenfell, B. T., Bjørnstad, O. N. & Finkenstädt, B. F. Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. *Ecological monographs* **72,** 185–202 (2002).

74. Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *science* **303,** 327–332 (2004).

75. Gupta, R. K., Best, J. & MacMahon, E. Mumps and the UK epidemic 2005. *Bmj* **330,** 1132–1135 (2005).

76. Hacisuleyman, E. *et al.* Vaccine breakthrough infections with SARS-CoV-2 variants. *New England Journal of Medicine* **384,** 2212–2218 (2021).

77. Halloran, M. E., Longini, I. M. & Struchiner, C. J. Design and Interpretation of Vaccine Field Studies. en. *Epidemiologic Reviews* **21,** 73–88. ISSN: 0193-936X, 1478-6729. https://academic.oup.com/epirev/article-lookup/doi/10.1093/oxfordjournals.epirev.a017990 (2021) (Jan. 1999).

78. Hamborsky, J., Kroger, A., *et al. Epidemiology and prevention of vaccine-preventable diseases, E-Book: The Pink Book* (Public Health Foundation, 2015).

79. Handel, A., Brown, J., Stallknecht, D. & Rohani, P. A multi-scale analysis of influenza A virus fitness trade-offs due to temperature-dependent virus persistence. *PLoS computational biology* **9,** e1002989 (2013).

80. Hashimoto, H., Fujioka, M., Kinumaki, H., Group, K. A. P. S., *et al.* An office-based prospective study of deafness in mumps. *The Pediatric infectious disease journal* **28,** 173–175 (2009).

81. Hayden, F. G. Perspectives on antiviral use during pandemic influenza. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **356,** 1877–1884 (2001).

82. Hayden, F. G. Pandemic influenza: is an antiviral response realistic? *The Pediatric infectious disease journal* **23,** S262–S269 (2004).

83. He, D., Ionides, E. L. & King, A. A. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface* **7,** 271–283 (2010).

84. Hernandez-Vargas, E. A. *Modeling and Control of Infectious Diseases in the Host: With MATLAB and R* (Academic Press, 2019).

85. Hernandez-Vargas, E. A. & Velasco-Hernandez, J. X. In-host Mathematical Modelling of COVID-19 in Humans. en. *Annual Reviews in Control* **50,** 448–456. ISSN: 13675788. https://linkinghub.elsevier.com/retrieve/pii/S1367578820300638 (2021) (2020).

86. Hien H Nguyen, M. *Influenza* July 2021. https://emedicine.medscape.com/article/219557-overview#a6.

87. Ho, D. D. *Time to hit HIV, early and hard* 1995.

88. Homan, E. J. & Bremel, R. D. Are cases of mumps in vaccinated patients attributable to mismatches in both vaccine T-cell and B-cell epitopes? An immunoinformatic analysis. *Human vaccines & immunotherapeutics* **10,** 290–300 (2014).

89. Hu, X., Chakravarty, S. D. & Ivashkiv, L. B. Regulation of interferon and Toll-like receptor signaling during macrophage activation by opposing feedforward and feedback inhibition mechanisms. *Immunological reviews* **226,** 41–56 (2008).

90. *Immunization and vaccine-preventable communicable diseases* https://www.who.int/data/gho/data/theme

91. Ionides, E. L., Bretó, C. & King, A. A. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **103,** 18438–18443 (2006).

92. Jick, H., Chamberlin, D. P. & Hagberg, K. W. The origin and spread of a mumps epidemic: United Kingdom, 2003–2006. *Epidemiology* **20,** 656–661 (2009).

93. Jin, L. *et al.* Genomic diversity of mumps virus and global distribution of the 12 genotypes. en. *Reviews in Medical Virology* **25,** 85–101. ISSN: 1099-1654. https://onlinelibrary.wiley.com/doi/abs/10.1002/rmv.1819 (2019) (2015).

94. Kanegae, Y. *et al.* Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. *Journal of virology* **64,** 2860–2865 (1990).

95. Keeling, M. J. & Rohani, P. *Modelling Infectious Diseases: In Humans and Animals* (Princeton University Press, 2008).

96. Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals* (Princeton University Press, 2011).

97. Kermack, W. O. & McKendrick, A. G. Contributions to the mathematical theory of epidemics—I. *Bulletin of mathematical biology* **53,** 33–55 (1991).

98. Kim, K. S. *et al.* A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. en. *PLOS Biology* **19.** Publisher: Public Library of Science, e3001128. ISSN: 1545-7885. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001128 (2021) (Mar. 2021).

99. Kim, K. S. *et al.* A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. en. *PLOS Biology* **19** (ed Sugden, B.) e3001128. ISSN: 1545-7885. https://dx.plos.org/10.1371/journal.pbio.3001128 (2021) (Mar. 2021).

100. Kim, Y.-I. *et al.* Infection and rapid transmission of SARS-CoV-2 in ferrets. *Cell host & microbe* **27,** 704–709 (2020).

101. King, A. A., Domenech de Cellès, M., Magpantay, F. M. & Rohani, P. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences* **282,** 20150347 (2015).

102. King, A. A., Nguyen, D. & Ionides, E. L. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software* **69,** 1–43 (2016).

103. King, A. A. *et al. pomp: Statistical Inference for Partially Observed Markov Processes* R package, version 3.5 (2021). https://kingaa.github.io/pomp/.

104. King, A. A. & Ionides, E. L. *Introduction to partially observed Markov processes* https://kingaa.github.io/short-course/intro/intro.html.

105. King, A. A. & Ionides, E. L. *Simulation of Stochastic models* https://kingaa.github.io/short-course/stochsim/stochsim.html.

106. Kloepfer, K. M. & Gern, J. E. Ecological and individual data both indicate that influenza inhibits rhinovirus infection. *Proceedings of the National Academy of Sciences* **117,** 6987–6987 (2020).

107. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314,** 1898–1903 (2006).

108. Koelle, K. & Rasmussen, D. A. The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. en. *Elife* **4,** e07361 (Sept. 2015).

109. Kontio, M., Jokinen, S., Paunio, M., Peltola, H. & Davidkin, I. Waning antibody levels and avidity: implications for MMR vaccine-induced protection. *The Journal of infectious diseases* **206,** 1542–1548 (2012).

110. Koutsakos, M. *et al.* Human CD8+ T cell cross-reactivity across influenza A, B and C viruses. *Nature immunology* **20,** 613–625 (2019).

111. Kullback, S. & Leibler, R. A. On information and sufficiency. *The annals of mathematical statistics* **22,** 79–86 (1951).

112. Laurie, K. L. *et al.* Interval between infections and viral hierarchy are determinants of viral interference following influenza virus infection in a ferret model. *The Journal of infectious diseases* **212,** 1701–1710 (2015).

113. Lazear, H. M., Schoggins, J. W. & Diamond, M. S. Shared and distinct functions of type I and type III interferons. *Immunity* **50,** 907–923 (2019).

114. Lee, C.-Y. & Lowen, A. C. Animal models for SARS-CoV-2. *Current Opinion in Virology* (2021).

115. Lewnard, J. A. & Grad, Y. H. Vaccine waning and mumps re-emergence in the United States. en. *Science Translational Medicine* **10,** eaao5945. ISSN: 1946-6234, 1946-6242. http://stm.sciencemag.org/content/10/433/eaao5945 (2018) (Mar. 2018).

116. Li, R. *et al.* Estimated demand for US hospital inpatient and intensive care unit beds for patients with COVID-19 based on comparisons with Wuhan and Guangzhou, China. *JAMA network open* **3,** e208297–e208297 (2020).

117. Linderman, S. L. *et al.* Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013–2014 influenza season. *Proceedings of the National Academy of Sciences* **111,** 15798–15803 (2014).

118. Liu, Y. *et al.* Cross-lineage protection by human antibodies binding the influenza B hemagglutinin. *Nature communications* **10,** 1–13 (2019).

119. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438,** 355–359 (2005).

120. Lowery, S. A., Sariol, A. & Perlman, S. Innate immune and inflammatory responses to SARS-CoV-2: Implications for COVID-19. *Cell Host & Microbe* (2021).

121. Łuksza, M. & Lässig, M. A predictive fitness model for influenza. *Nature* **507,** 57–61 (Feb. 2014).

122. MacDonald, N., Hatchette, T., Elkout, L. & Sarwal, S. Mumps is back: why is mumps eradication not working? *Hot topics in infection and immunity in children VII,* 197–220 (2011).

123. Magpantay, F., Riolo, M., de Cellès, M., King, A. & Rohani, P. Epidemiological Consequences of Imperfect Vaccines for Immunizing Infections. *SIAM Journal on Applied Mathematics* **74,** 1810–1830. ISSN: 0036-1399. https://epubs.siam.org/doi/10.1137/140956695 (2019) (Jan. 2014).

124. Marino, S., Hogue, I. B., Ray, C. J. & Kirschner, D. E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. en. *Journal of Theoretical Biology* **254,** 178–196. ISSN: 0022-5193. https://www.sciencedirect.com/science/article/pii/S0022519308001896 (2021) (Sept. 2008).

125. St-Martin, G. *et al.* Mumps resurgence in Denmark. *Journal of Clinical Virology* **61,** 435–438 (2014).

126. Martin, S. J. *The biochemistry of viruses* (CUP Archive, 1978).

127. Martinez, P. P., King, A. A., Yunus, M., Faruque, A. & Pascual, M. Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world. *Proceedings of the National Academy of Sciences* **113,** 4092–4097 (2016).

128. Matthies, H. & Strang, G. The solution of nonlinear finite element equations. *International journal for numerical methods in engineering* **14,** 1613–1626 (1979).

129. McLean, A. R. & Blower, S. M. Imperfect vaccines and herd immunity to HIV. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **253,** 9–13 (1993).

130. Menzies, N. A. *et al.* Progression from latent infection to active disease in dynamic tuberculosis transmission models: a systematic review of the validity of modelling assumptions. *The Lancet Infectious Diseases* **18,** e228–e238 (2018).

131. Miao, H., Xia, X., Perelson, A. S. & Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM review* **53,** 3–39 (2011).

132. Morrison, J. C., Givens, J. R., Wiser, W. L. & Fish, S. A. Mumps oophoritis: a cause of premature menopause. *Fertility and sterility* **26,** 655–659 (1975).

133. Mortimer, P. Mumps prophylaxis in the light of a new test for antibody. *Br Med J* **2,** 1523–1524 (1978).

134. Mossong, J. & Muller, C. P. Modelling measles re-emergence as a result of waning of immunity in vaccinated populations. *Vaccine* **21,** 4597–4603 (2003).

135. Mossong, J. *et al.* Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. en. *PLOS Medicine* **5,** e74. ISSN: 1549-1676. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050074 (2018) (Mar. 2008).

136. Möst, J. & Weiss, G. Consecutive infections with influenza A and B virus in children during the 2014–2015 seasonal influenza epidemic. *The Journal of infectious diseases* **214,** 1139–1141 (2016).

137. Mullen, K., Ardia, D., Gil, D., Windover, D. & Cline, J. DEoptim: An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software* **40,** 1–26 (2011).

138. Mullen, K., Ardia, D., Gil, D. L., Windover, D. & Cline, J. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software* **40,** 1–26 (2011).

139. *Mumps health care information* https://www.cdc.gov/mumps/hcp.html. Mar. 2021.

140. *Mumps vaccination* Jan. 2021. https://www.cdc.gov/vaccines/vpd/mumps/index.html.

141. Murphy, K. & Weaver, C. *Janeway's immunobiology* (Garland science, 2016).

142. *National Notifiable Diseases Surveillance System (NNDSS).* https://wonder.cdc.gov/nnd

143. Nelder, J. A. & Mead, R. A simplex method for function minimization. *The computer journal* **7,** 308–313 (1965).

144. Nguyen, T.-Q., Rollon, R. & Choi, Y.-K. Animal Models for Influenza Research: Strengths and Weaknesses. *Viruses* **13,** 1011 (2021).

145. Nguyen, V. K. & Hernandez-Vargas, E. A. Identifiability Challenges in Mathematical Models of Viral Infectious Diseases**This work was supported by iMed - the Helmholtz Initiative on Pesonalized Medicine. en. *IFAC-PapersOnLine* **48,** 257–262. ISSN: 24058963. https://linkinghub.elsevier.com/retrieve/pii/S2405896315027597 (2021) (2015).

146. Nickbakhsh, S. *et al.* Virus–virus interactions impact the population dynamics of influenza and the common cold. *Proceedings of the National Academy of Sciences* **116,** 27142–27150 (2019).

147. O'Leary, J. J., Sheils, O., Martin, C. & Crowley, A. TaqMan Technology and Real-Time Polymerase Chain Reaction. *Molecular biology in cellular pathology,* 251 (2003).

148. Owusu, D., Hand, J., Tenforde, M. W., *et al.* Early Season Pediatric Influenza B/Victoria Virus Infections Associated with a Recently Emerged Virus Subclade—Louisiana, 2019. *MMWR. Morbidity and Mortality Weekly Report* **69,** "40–43" (2020).

149. Palese, P. & Wang, T. T. Why do influenza virus subtypes die out? A hypothesis. *MBio* **2,** e00150–11 (2011).

150. Park, A. & Iwasaki, A. Type I and type III interferons–induction, signaling, evasion, and application to combat COVID-19. *Cell host & microbe* **27,** 870–878 (2020).

151. Park, S. H. Resurgence of mumps in Korea. *Infection & chemotherapy* **47,** 1–11 (2015).

152. Pattabhi, S. *et al.* Targeting innate immunity for antiviral therapy through small molecule agonists of the RLR pathway. *Journal of virology* **90,** 2372–2387 (2015).

153. Paul Glezen, W., Schmier, J. K., Kuehn, C. M., Ryan, K. J. & Oxford, J. The burden of influenza B: a structured literature review. *American journal of public health* **103,** e43–e51 (2013).

154. Paules, C. & Subbarao, K. *Influenza. Lancet Lond Engl* 2017.

155. Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood* (Oxford University Press, 2001).

156. Paxson, V. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review* **27,** 5–18 (1997).

157. Peltola, H. *et al.* Mumps outbreaks in Canada and the United States: time for new thinking on mumps vaccines. *Clinical infectious diseases* **45,** 459–466 (2007).

158. Petrova, V. N. & Russell, C. A. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology* **16,** 47–60 (2018).

159. Plotkin, J. B., Dushoff, J. & Levin, S. A. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. en. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6263–6268 (Apr. 2002).

160. Plotkin, S., Offit, W. & KM, P. E. *Plotkin's Vaccines. ed* 2018.

161. *Project Tycho* https://www.tycho.pitt.edu/data.

162. Ramanathan, R., Voigt, E., Kennedy, R. & Poland, G. Knowledge gaps persist and hinder progress in eliminating mumps. *Vaccine* **36,** 3721–3726 (2018).

163. Ramos, I. & Fernandez-Sesma, A. Modulating the innate immune response to influenza A virus: potential therapeutic use of anti-inflammatory drugs. *Frontiers in immunology* **6,** 361 (2015).

164. Randall, R. E. & Goodbourn, S. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *Journal of general virology* **89,** 1–47 (2008).

165. Read, A. F. *et al.* Imperfect vaccination can enhance the transmission of highly virulent pathogens. *PLoS biology* **13,** e1002198 (2015).

166. Rijo-Ferreira, F. & Takahashi, J. S. *Circadian rhythms in infectious diseases and symbiosis* in *Seminars in Cell & Developmental Biology* (2021).

167. Riolo, M. A., King, A. A. & Rohani, P. Can vaccine legacy explain the British pertussis resurgence? *Vaccine* **31,** 5903–5908. ISSN: 0264-410X. http://www.sciencedirect.com/science/article/pii/S0264410X13012516 (2018) (Dec. 2013).

168. Riolo, M. A. & Rohani, P. Combating pertussis resurgence: One booster vaccination schedule does not fit all. en. *Proceedings of the National Academy of Sciences* **112,** E472–E477. ISSN: 0027-8424, 1091-6490. http://www.pnas.org/lookup/doi/10.1073/pnas.1415573112 (2020) (Feb. 2015).

169. Rohani, P., Green, C., Mantilla-Beniers, N. & Grenfell, B. T. Ecological interference between fatal diseases. *Nature* **422,** 885–888 (2003).

170. Rohani, P., Earn, D. J., Finkenstädt, B. & Grenfell, B. T. Population dynamic interference among childhood diseases. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265,** 2033–2041 (1998).

171. Roosa, K. & Chowell, G. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling* **16,** 1–15 (2019).

172. Rossetto, M., Thurlby, K. A., Offord, C. A., Allen, C. B. & Weston, P. H. The impact of distance and a shifting temperature gradient on genetic connectivity across a heterogeneous landscape. *BMC Evolutionary Biology* **11,** 1–11 (2011).

173. Rota, P. A. *et al.* Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. *Virology* **175,** 59–68 (1990).

174. Rubin, S. A. *et al.* Antibody induced by immunization with the Jeryl Lynn mumps vaccine strain effectively neutralizes a heterologous wild-type mumps virus associated with a large outbreak. *The Journal of infectious diseases* **198,** 508–515 (2008).

175. Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *Elife* **7,** e32303 (2018).

176. Ryu, W.-S. *Molecular virology of human pathogenic viruses* (Academic Press, 2016).

177. Sadler, A. J. & Williams, B. R. Interferon-inducible antiviral effectors. *Nature reviews immunology* **8,** 559–568 (2008).

178. Saenz, R. A. *et al.* Dynamics of influenza virus infection and pathology. *Journal of virology* **84,** 3974–3983 (2010).

179. Sanders, C. J., Doherty, P. C. & Thomas, P. G. Respiratory epithelial cells in innate immunity to influenza virus infection. *Cell and tissue research* **343,** 13–21 (2011).

180. *SARS-CoV-2 Variant Classifications and Definitions* https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html#anchor_1632154493691.

181. Sasaki, A., Lion, S. & Boots, M. Antigenic escape selects for the evolution of higher pathogen transmission and virulence. *Nature Ecology & Evolution* **6,** 51–62 (2022).

182. Schwarz, B. & Hoag, J. Interpreting model results—examples from an energy model. *Policy Sciences* **15,** 167–181 (1982).

183. Scrucca, L. GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software, Articles* **53,** 1–37. ISSN: 1548-7660. https://www.jstatsoft.org/v053/i04 (2013).

184. Scrucca, L. On Some Extensions to GA Package: Hybrid Optimisation, Parallelisation and Islands Evolution. *The R Journal* **9,** 187–206. https://doi.org/10.32614/RJ-2017-008 (2017).

185. Shehab, Z. M., Brunell, P. A. & Cobb, E. Epidemiological standardization of a test for susceptibility to mumps. *Journal of Infectious Diseases* **149,** 810–812 (1984).

186. Shrestha, S., King, A. A. & Rohani, P. Statistical inference for multi-pathogen systems. *PLoS Comput Biol* **7,** e1002135 (2011).

187. Singh, D. & Soojin, V. Y. On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine* **53,** 537–547 (2021).

188. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *science* **305,** 371–376 (2004).

189. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. en. *Science* **305,** 371–376 (July 2004).

190. Sobol, I. M. Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics* **16,** 236–242 (1976).

191. Sobol', I. M. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **7,** 784–802 (1967).

192. Stahl, J.-P., Mailles, A., Dacheux, L. & Morand, P. Epidemiology of viral encephalitis in 2011. *Medecine et maladies infectieuses* **41,** 453–464 (2011).

193. Storn, R. & Price, K. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization* **11,** 341–359 (1997).

194. Sutton, T. C. & Subbarao, K. Development of animal models against emerging coronaviruses: from SARS to MERS coronavirus. *Virology* **479,** 247–258 (2015).

195. Tafalla, M., Buijssen, M., Geets, R. & Vonk Noordegraaf-Schouten, M. A comprehensive review of the epidemiology and disease burden of Influenza B in 9 European countries. *Human vaccines & immunotherapeutics* **12,** 993–1002 (2016).

196. Terajima, M., Babon, J. A. B., Co, M. D. T. & Ennis, F. A. Cross-reactive human B cell and T cell epitopes between influenza A and B viruses. *Virology journal* **10,** 244 (2013).

197. The MathWorks, I. *Symbolic Math Toolbox* (Natick, Massachusetts, United State, 2019). https://www.mathworks.com/help/symbolic/.

198. *U.S. Influenza Surveillance System: Purpose and Methods* Accessed: 2020-04-17. Oct. 2019. https://www.cdc.gov/flu/weekly/overview.htm.

199. Van Boven, M., Ruijs, W. L., Wallinga, J., O'Neill, P. D. & Hahne, S. Estimation of vaccine efficacy and critical vaccination coverage in partially observed outbreaks. *PLoS computational biology* **9,** e1003061 (2013).

200. VanderWaal, K. L. & Ezenwa, V. O. Heterogeneity in pathogen transmission: mechanisms and methodology. *Functional Ecology* **30,** 1606–1622 (2016).

201. Vijaykrishna, D. *et al.* The contrasting phylodynamics of human influenza B viruses. *Elife* **4,** e05055 (2015).

202. Virk, R. K. *et al.* Divergent evolutionary trajectories of influenza B viruses underlie their contemporaneous epidemic activity. *Proceedings of the National Academy of Sciences* **117,** 619–628 (2020).

203. Vygen, S. *et al.* Waning immunity against mumps in vaccinated young adults, France 2013. *Eurosurveillance* **21,** 30156 (2016).

204. Wagner, C. E., Saad-Roy, C. M. & Grenfell, B. T. Modelling vaccination strategies for COVID-19. *Nature Reviews Immunology* **22,** 139–141 (2022).

205. Wearing, H. J., Rohani, P. & Keeling, M. J. Appropriate Models for the Management of Infectious Diseases. *PLoS Medicine* **2,** e174. ISSN: 1549-1277 (2005).

206. *WHO coronavirus (COVID-19) Dashboard* https://covid19.who.int/.

207. Winfield, N. *'Not a wave, A tsunami.' Italy hospitals at Virus limit* Mar. 2020. https://www.pbs.org/newshour/health/not-a-wave-a-tsunami-italy-hospitals-at-virus-limit.

208. Wohl, S. *et al.* Combining genomics and epidemiology to track mumps virus transmission in the United States. en. *PLOS Biology* **18** (ed Del Valle, S. Y.) e3000611. ISSN: 1545-7885. https://dx.plos.org/10.1371/journal.pbio.3000611 (2020) (Feb. 2020).

209. Woolhouse, M. E. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences* **94,** 338–342 (1997).

210. Wu, H., Wang, F., Tang, D. & Han, D. Mumps orchitis: clinical aspects and mechanisms. *Frontiers in Immunology* **12,** 848 (2021).

211. Yakimovich, A. *et al.* Plaque2. 0—a high-throughput analysis framework to score virus-cell transmission and clonal cell expansion. *PloS one* **10,** e0138760 (2015).

212. Ye, L., Schnepf, D. & Staeheli, P. Interferon-λ orchestrates innate and adaptive mucosal immune responses. *Nature Reviews Immunology* **19,** 614–625 (2019).

213. Zhou, L., Ge, X. & Yang, H. Porcine reproductive and respiratory syndrome modified live virus vaccine: a "leaky" vaccine with debatable efficacy and safety. *Vaccines* **9,** 362 (2021).

APPENDIX A

SUPPLEMENTARY INFORMATION CH.2

This appendix documents the derivation for the antiviral effective reproductive number presented in Eq.3.51. We use the method of next generation operator (NGO)[47] to derive the quantity R_A. MATLAB package *symbolic toolbox* [197] was used to simplify matrix and algebraic expressions. The NGO method applied is described in the following steps:

1. From the system of differential equations documented in Eq. 3.42-3.49, compartments that contribute to the productions of novel infected cells were identified. In this case, the relevant compartments are - (1) productively infected cells (I), (2) productively infected cells bound to the antiviral molecules (I_A) and (3) free virions (V)

2. For each of these compartments, two rate vectors were defined -

$$\mathscr{F}^{\mathscr{A}} = \left[\beta_T T_0 V,\ \beta_T (1-\omega) T_0^A V,\ 0\right], \tag{A.1}$$

$$\mathscr{V}^{\mathscr{A}} = \left[(\delta_I + \beta_A) I,\ \delta_I \frac{1}{(1-\eta)} I_A - \beta_A I,\ \delta_V V - \nu\varepsilon I - (1-\alpha)\nu\varepsilon I_A\right]. \tag{A.2}$$

$\mathscr{F}^{\mathscr{A}}$ defines the vector of rates at which new infected cells are produced. $\mathscr{V}^{\mathscr{A}}$ defines the vector of the flow in and out of the infected cell compartments.

3. Jacobian matrices were then defined using vectors in Eq. A.1-A.2. To linearize the system in Eq. A.1-A.2, it was assumed that the susceptible target population and anti-viral bound target population were are at constant values T_0 and T_0^A respectively i.e. the two matrices evaluated at the infection free equilibrium. F^A is such that every element $\mathrm{F}^A_{i,j} = \frac{\partial \mathscr{F}^A_j}{\partial Z_i}$, and V^A is such that every element $\mathrm{V}^A_{i,j} = \frac{\partial \mathscr{V}^{\mathscr{A}}{}_j}{\partial Z_i}$, where $Z = [I_A,\ I,\ V]$ is the vector of relevant state variables that generate novel infections arise. We express the Jacobian matrices to be -

$$\mathrm{F}^A = \begin{bmatrix} 0 & 0 & \beta_T T_0 \\ 0 & 0 & \beta_T(1-\omega)T_0^A \\ 0 & 0 & 0 \end{bmatrix}, \tag{A.3}$$

$$\mathrm{V}^A = \begin{bmatrix} (\delta_I + \beta_A) & 0 & 0 \\ -\beta_A & \delta_I \frac{1}{(1-\eta)} & 0 \\ -\nu\varepsilon & -\nu\varepsilon(1-\alpha) & \delta_V \end{bmatrix}, \tag{A.4}$$

$$\mathrm{V}^{A-1} = \begin{bmatrix} \frac{1}{(\delta_I+\beta_A)} & 0 & 0 \\ \frac{\beta_A}{\delta_I(\delta_I+\beta_A)}(1-\eta) & \frac{1}{\delta_I}(1-\eta) & 0 \\ \frac{(1-\eta)}{(\delta_I+\beta_A)\delta_I\delta_V}\left(\beta_A\nu\varepsilon(1-\alpha)+\delta_I\nu\varepsilon\frac{1}{(1-\eta)}\right) & \frac{\nu\varepsilon}{\delta_I\delta_V}(1-\alpha)(1-\eta) & \frac{1}{\delta_V} \end{bmatrix}. \tag{A.5}$$

Here, V^{A-1} is the inverse of the V^A matrix which was used in the calculation of the next generation matrix.

4. We then proceeded to calculate the next generation operator as -

$$\mathrm{K}^A = \mathrm{F}^A\mathrm{V}^{A-1} \tag{A.6}$$

5. The largest eigen value of matrix K^A, calculated by solving the characteristic equation, $|\mathrm{K}^A - \Lambda\mathrm{I}| = 0$, represents R_A. Here, I is the identity matrix.

Following the NGO method and simplifying the largest cubic root of the characteristic equation,

$$R_A = R_{eff}\Big(1 - p_A(\alpha(1-\eta)+\eta) + f_0^A(1-\alpha)(1-\eta)(1-\omega)\Big). \quad \text{(A.7)}$$

APPENDIX B

SUPPLEMENTARY INFORMATION CH.4

B.1 Estimating annual migration rates

We considered a simple model to account for the demographic dynamics at a national-level for the US. We formulated a system of difference equations containing five state-variables each corresponding to a distinct age cohort in the population, $[0,5)$, $[5,15)$, $[15,25)$, $[25,40)$ and ≥ 40, which were denoted by the subscript $i = 1,\ldots,5$ respectively. The population size of the i^{th} age cohort in year t is given by $N_{i,t}$, and $N_t = \sum_i N_{i,t}$ is the total population. We denote the per-capita annual birth rate by ν_t and the per-captia age-specific migration rate (accounting for the combined effects of immigration, emmigration and death) by μ_i. The width in years of the i-th age cohort is denoted by $1/\omega_i$. Using demographic data on the annual birth rate and the population age profile, we inferred the age-specific, annual migration rates ($\mu_{i,t}$) by back solving the difference equations

The flows with this model are described as

$$\Delta N_{i,t} = \nu_t N_{i,t}\delta_{i,1} + N_{(i-1),t}\omega_{(i-1)}\delta_{i,1} - N_{i,t}(\omega_i - \mu_{i,t}), \tag{B.1}$$

where $\Delta N_{i,t}$ is the annual change in population size of age group i and $\delta_{i,j}$ is the Kronecker delta, $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise.

B.2 Transforming the POLYMOD contact matrix

The R package "socialmixr" [63] was used to create a contact matrix matching the five age classes in our data (see above). A matrix of contacts is generated by bootstrapping from the survey diary data for the UK derived from the POLYMOD study [135].

We compute the contact matrix used in the analysis from this raw contact matrix, with elements $C^{\text{raw}}_{i,j}$ corresponding the mean the daily contacts (physical and non-physical) between an individual of age class i making a contact with age class j. We rescaled the elements so that they correspond to US population sizes and corrected for reciprocity to ensure final contact matrix satisfies the condition $N_i C_{i,j} = N_j C_{j,i}$. Explicitly, the following steps were used to calculate the resulting contact matrix:

- Calculate $E_{ij} = N_{i,t=2005} C^{\text{raw}}_{ij}$. Here E_{ij} is the matrix of total daily contacts scaled to the US population for the year 2005. US population of 2005 was selected as it was the year the original POLYMOD

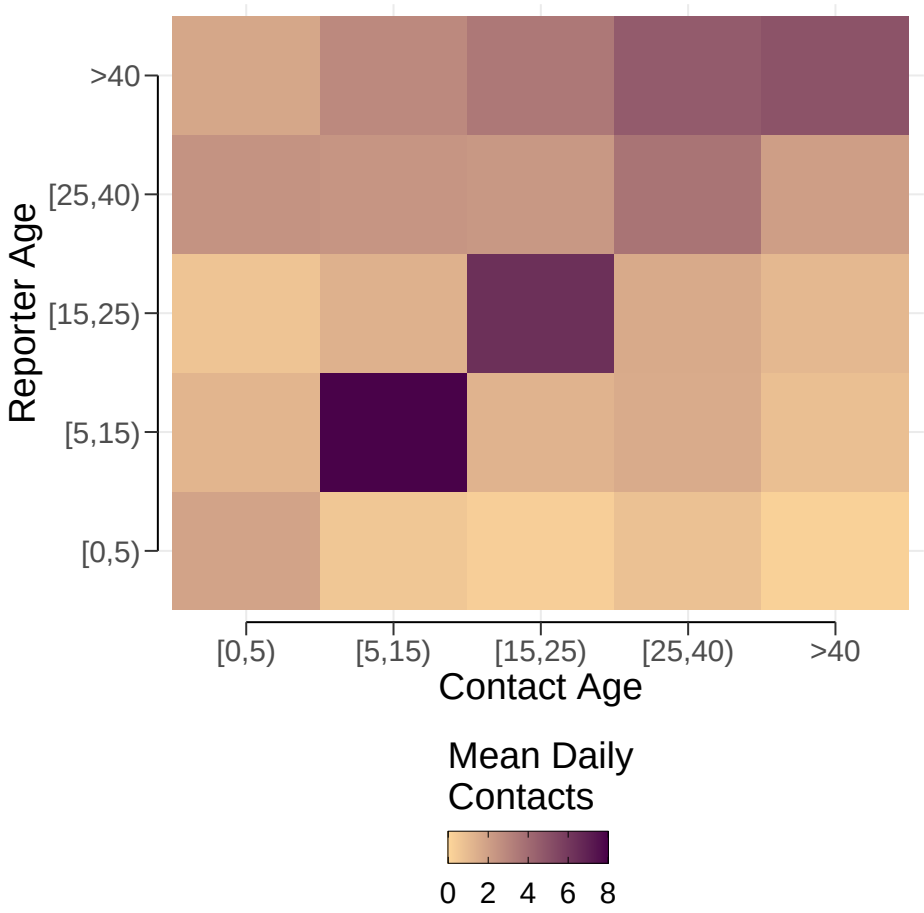

Figure B.1: **Matrix of daily contact rates.** Daily diary records of who (y-axis) comes in contact with whom (x-axis) for the UK in 2005. Contact rates were caliberated for the US using age-specific national population sizes and then corrected for reciprocity.

survey was conducted. We assumed that average contact network remains unchanged over time.

- Correct for reciprocity by calculating $E_{ij}^{\mathrm{S}} = \frac{(E_{ij}+E_{ji})}{2}$, the symmetric matrix of total scaled daily contacts between age group i and ange group j.

- Calculate the final contact matrix $C_{i,j} = \frac{E^S}{N_{i,t=2005}}$, giving the daily contacts of an individual in age class i with individuals in age class j scaled for the US population structure and corrected for reciprocity.

B.3 Reconstructing missing vaccine coverage

Vaccine coverage data for both the neonatal and booster vaccine doses were incomplete. To reconstruct the missing data we used differing approaches for each dose.

Neonatal Dose. To impute missing vaccine coverage data for the first dose (Fig. B.2A), we used a Brownian bridge (BB) with a flexible curvature function. This was achieved by defining a system stochastic differential equations with five state variables – Susceptible (S), Exposed (E), Infectious (I), Vaccinated (V) and Bridge (B). The BB is a stochastic process defined between two points $[a,b]$ over the time interval $[t_0,t_n]$ [156]. For the vaccine coverage of the neonatal dose, $t_0 = 1968$, $a = 0$, $t_n = 1985$, and $b = 0.86$; the first known value in the coverage data for the neonatal dose. The BB is given by

$$B(t) = \sigma_B W(t) - \frac{t - t_0}{t_n - t_0}\sigma_B W(t_n), \tag{B.2}$$

$$\hat{p}(t) = \left(\frac{t - t_0}{t_n - t_0}\right)^K + B(t), \tag{B.3}$$

where $W(t)$ is the Wiener process with infinitesimal variance σ_B^2 and K controls the curvature of the bridge. al end-point of the bridge which was fixed at 1985. A realized vaccine uptake trajectory, $\hat{p}(t)$, was fed into disease transmission model,

$$\frac{dS}{dt} = \nu(1-(1-\alpha)p(\hat{t}))N - (\lambda+\mu)S, \tag{B.4}$$

$$\frac{dE}{dt} = \lambda S - (\sigma+\mu)E, \tag{B.5}$$

$$\frac{dI}{dt} = \sigma E - (\gamma+\mu)I, \tag{B.6}$$

$$\frac{dV}{dt} = \nu(1-\alpha)p(\hat{t})N - (\lambda+\mu)V. \tag{B.7}$$

The BB was calibrated to generate vaccine coverage consistent with the initial decline of cases in after roll-out during 1968-1985 (Fig. B.2 B).

Calibration was conducted by matching simulated cases from the model to the observed cases during this era using

$$C_t = \int_{t-1}^{t} \gamma I(s)ds. \tag{B.8}$$

We carried out one-step forward simulation of Eq. B.8. Each step simulation consisted of 1000 particles using uniformly sampled parameter values of $K \sim U[0,5]$ and $\sigma_B \sim U[1,50]$. Following the simulation step, a subset of filtered particles was produced assuming an over-dispersed normal model $\left(\mathcal{N}(\rho C_t, \sqrt{\rho C_t(1-\rho+\psi^2 C_t)})\right)$ of observation errors. The resulting weekly vaccine coverage values were aggregated to match the temporal resolution the reported cases of mumps (Fig. B.2 C). Details on the particle filtering algorithm can be found here [10].

Booster Dose. We assumed a time-dependent flexible function to construct four shapes of early booster vaccine uptake after roll-out during (1989–2000). This function is given as

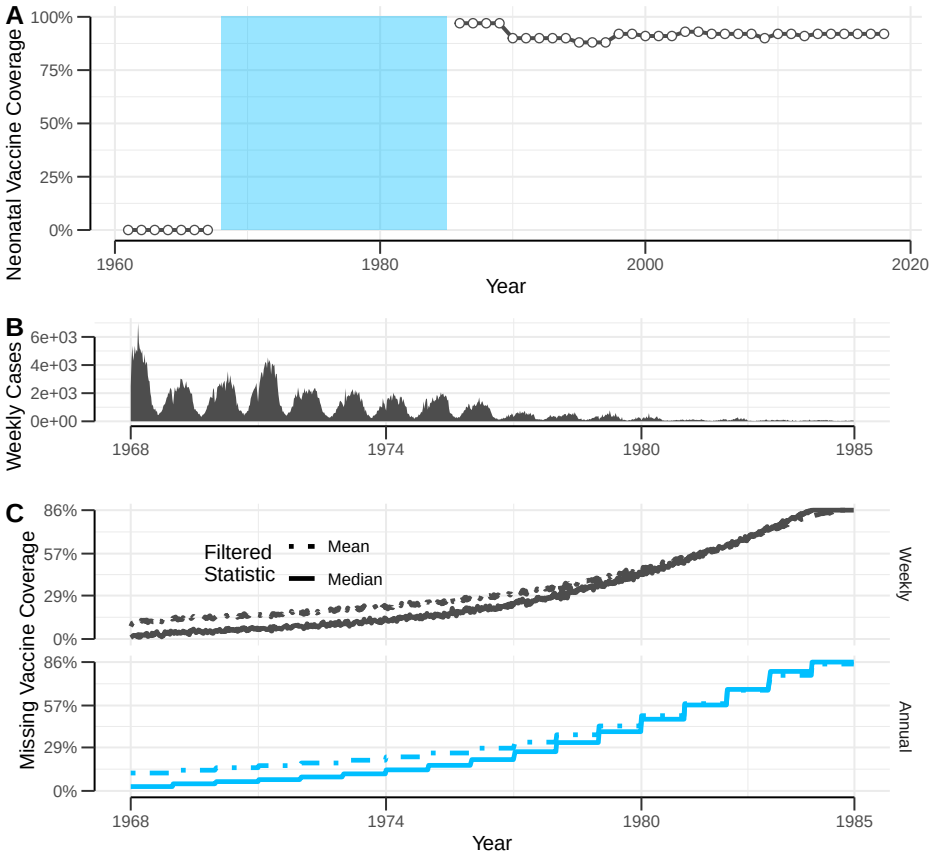

Figure B.2: **Reconstructed neonatal vaccine rate.** (A) Time series for the available mumps annual vaccine coverage rates for the neonatal dose. The blue region highlights the period of missing vaccine data. (B) Weekly mumps incidence during the period where vaccine data is missing. (C) Imputed, weekly vaccine uptake (upper panel) by recursively filtering from 1000 stochastically generated trajectories conditioned on the initial decline of mumps in (B). Filtering distribution is assumed to be an over-dispersed normal ($\psi = 0.8$, $\rho = 0.04$). Averaged annual coverage (lower panel) is used in the testing of hypothesis.

$$f(t) = \frac{a}{1 + e^{-b(t-t_0)}}. \tag{B.9}$$

In Eq. B.9, a describes the point of saturation of the logistic function, b controls the rate of saturation to the maximum, where as t_0 is the the mid-point on the time axis. Using different parametric configurations of $f(t)$, we constructed three shapes of vaccine coverage– "concave" (with parameters $a = 0.86$, $b = 0.8$ and $t_0 = 1999.7$), "sigmoid" ($a = 0.86$, $b = 0.8$ and $t_0 = 1994$) and "convex" ($a = 0.86$, $b = 1.5$ and $t_0 = 1989.8$). We also assumed a right constant interpolated booster vaccine uptake in the early years (Fig. B.3D),

B.4 Full model of vaccine traits and mumps transmission

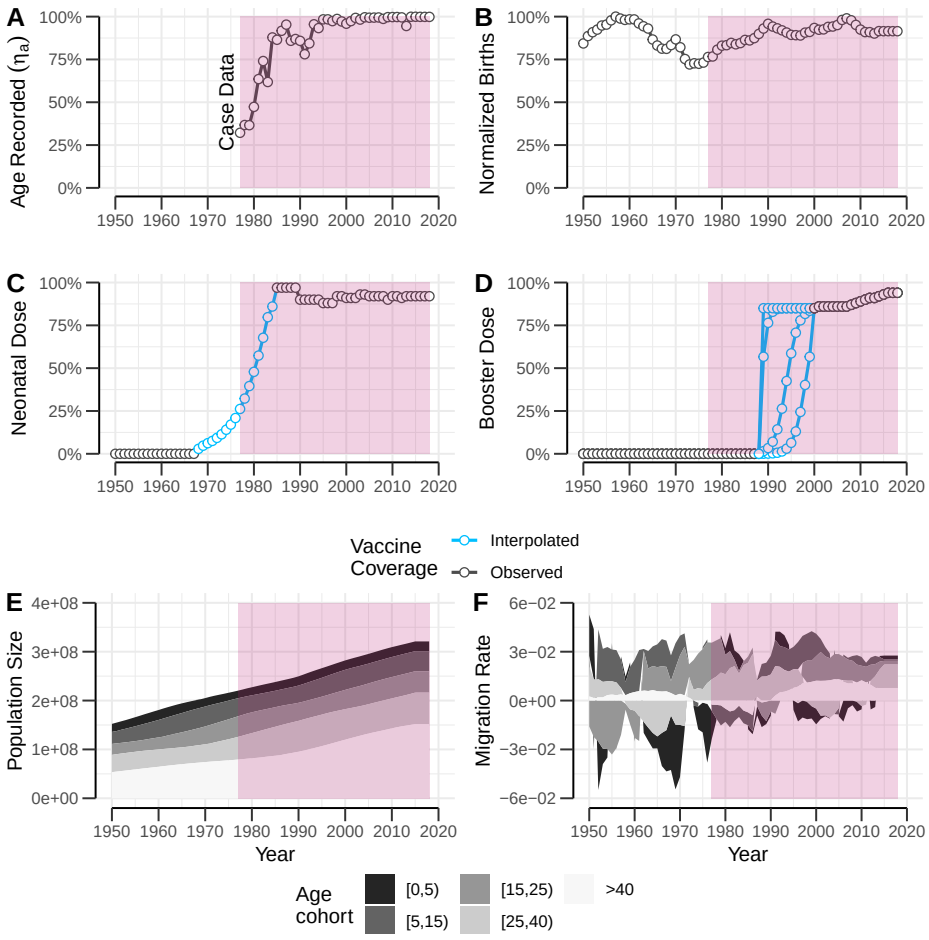

Figure B.3: **Model covariates.** (A-F) Demographic and epidemiological variables used in introducing realizm to the simulated dynamics; (A) Probability that for a given case age was reported; (B) Total normalized annual births; Annual mumps vaccination rates for the neonatal dose (C) and booster dose (D), Missing early uptake values were interpolated (blue); (E) Age-specific population size; And (F) migration rates broken down by age cohorts.

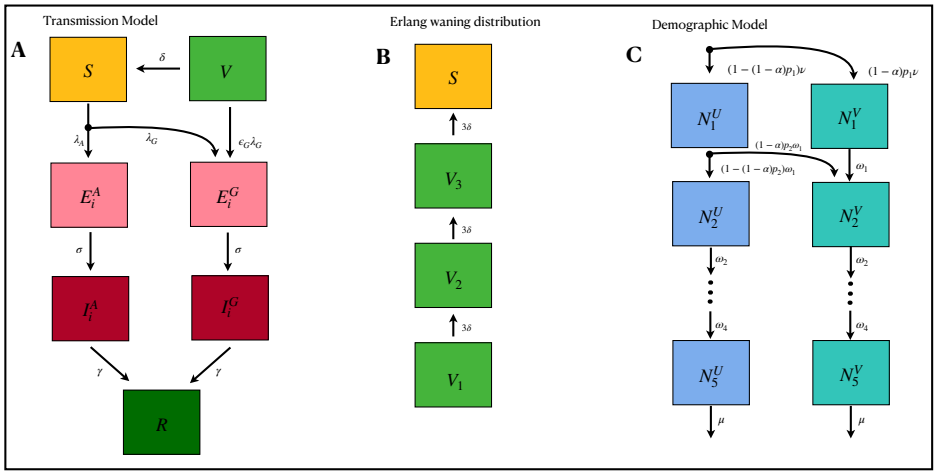

Figure B.4: **Transmission model schematic diagram** (A) Basic skeleton of model with vaccine imperfection. (B) Erlang distributed waning distribution. (C) Demographic model to account for vaccination, and aging.

Table B.1: Model parameters with ranges were estimated by fitting model to a time series of Annual Mumps Incidence data for the US (1977-2018)

Parameter	Value/Range	Definition
	Process model	
q_i	$[0,1]$	Age-specific Susceptibility of the i^{th} age-cohort
$1/\gamma$	5 Days	Duration of infection
$1/\sigma$	$[12,25]$ Days [139]	Duration of latency
α	0.054 [45, 199]	Primary vaccine failure probability
$1/\omega_i$	5, 10, 10, 15, 40 Years	Age-cohort widths
ι	1 Year^{-1}	Importation rate of infectious indviduals
$1/\delta$	≥ 0 Years	Duration of vaccine induced immunity
β_1	$[0,1]$	Amplitude of seasonality
ε	$[0,1]$	Vaccine leakiness
p_{intro}	$\{[0,5),[5,15),[15,25),[25,40)\}$	Age-group of the imported cases
t_{intro}	[2000, 2005]	Year of the imported cases
	Observation model	
ρ_i	$[0,1]$	Probability of reporting for the i^{th} age-cohort
ρ_u	$[0,1]$	Probability of reporting for the unstructured cases
ψ_i	≥ 0	Dispersion parameter for the i^{th} age-cohort
ψ_u	≥ 0	Dispersion parameter for the unstructured cases

Table B.2: Model Compartments for the i^{th} age-cohort

Compartment	Definition
S_i	Susceptible individuals
V_i	Vaccinated individuals
E_i^n	Exposed individuals to Genotype $n \in \{A, G\}$
I_i^n	Infectious individuals to Genotype $n \in \{A, G\}$
R_i	Recovered individuals

We developed an age-structured vaccinated, susceptible, exposed, infectious recovered (VSEIR) compartmental model. The full model consisted of three modes of vaccine imperfection: primary vaccine failure, waning immunity and vaccine leakiness against the imported mumps genotype G. We first outline the model assuming the duration of vaccine derived immunity wanes exponentially (Fig. B.4A). To account for demographic effects, we included processes in the model corresponding to births, aging and deaths, in addition to vaccination and pathogen transmission. As outlined above, our full age-structured model has five age-cohorts $[0,5)$, $[5,15)$, $[15,25)$, $[25,40)$, and > 40 which we indicate using the latin subscript i which ranges from 1 to 5. Furthermore, the model allows for the circulating of two genotypes, A and G, which we label with the latin superscript n. Vaccination is implemented in two doses: (1) a neonatal dose, with vaccination probability $p_1(t)$, in age group $i = 1$ (the $[0,5)$ cohort), and (2) a booster dose, with probability $p_2(t)$, in age group $i = 2$ (the $[5,15)$ age cohort). Tables B.1 & B.2 provide definition of model parameters, and compartments used in the full model.

The force of infection experienced by age group i due to genotype n for this model is given by

$$\lambda_i^n(t) = q_i \sum_{j=1}^{5} C_{i,j} \frac{I_j^n}{N_j} \phi_{i,j}(t), \tag{B.10}$$

where q_i is the age-specific susceptiblity, $C_{i,j}$ is the contact rate between i and j (see Fig. B.1), I_j^n is the number of individuals infectious with genotype n in age group j and finally $\phi_{i,j}(t)$ is a sinusoidal function with

amplitude β_1 and period $T = 1$ year accounting for term time forcing associated with school attendance by individuals in the $[5, 15)$ age group,

$$\phi_{i,j}(t) = \begin{cases} 1 - \beta_1 sin\left(\frac{2\pi t}{T}\right) & \text{if } i, j = 2, \\ 1 & \text{otherwise,} \end{cases} \tag{B.11}$$

and has the effect of shinking the force of infection for vaccinated individuals, with $\varepsilon = 0$ corresponding to no leakiness and $\varepsilon = 1$ corresponding to no vaccine protection against infection.

The vaccine leakiness depends on the strain the vaccinated individual is exposed to,

$$\varepsilon^n = \begin{cases} \varepsilon \text{ for } n = G, \\ 0 \text{ for } n = A \end{cases} \tag{B.12}$$

In addition to secondary transmission via the force of infection, we also include importation of the virus to the population in our model. The importation term $\iota_i^n(t)$ gives the rate at which infections of genotype n are imported into age group i at time t. For our purposes, we only model the first introduction of genotype G the population which occurs at time specified by the parameter t_{intro} into age group p_{intro} , i.e. $\iota_i^n(t) = \iota \delta_{n,G} \delta_{t,t_{intro}} \delta_{i,p_{intro}}$.

The following system differential equations was used to describe dynamics of the age structured model. For detailed definitions of model parameters and compartments, refer to Table B.1–B.2 The dynamics of the first age class, $i = 1$, are given by

$$\frac{dS_1}{dt} = \big(1-(1-\alpha)p_1(t)\big)\nu(t)N(t) + \delta V_1 - \left(\sum_n \lambda_1^n(t) + \mu_1(t) + \omega_1\right)S_1, \tag{B.13}$$

$$\frac{dV_1}{dt} = (1-\alpha)p_1(t)\nu(t)N(t) - \left(\delta + \sum_n \varepsilon^n\lambda_1^n + \mu_1(t) + \omega_1\right)V_1, \tag{B.14}$$

$$\frac{dE_1^n}{dt} = \lambda_1^n(t)(S_1 + \varepsilon^n V_1) - (\sigma + \mu_1(t) + \omega_1)E_1^n + \iota_1^n(t), \tag{B.15}$$

$$\frac{dI_1^n}{dt} = \sigma E_1^n - (\gamma + \mu_1(t) + \omega_1)I_1^n, \tag{B.16}$$

$$\frac{dR_1}{dt} = \gamma\sum_n I_1^n - (\mu_1(t) + \omega_1)R_1. \tag{B.17}$$

where we introduce the neonatal vaccine dose $p_1(t)$. The differential equations for the older age groups, $i = 2,\ldots,5$ have a similar structure, however instead of new births they have terms corresponding to the aging of individuals from the previous age group,

$$\frac{dS_i}{dt} = (1-v_i(t))\omega_{i-1}S_{i-1} + \delta V_i - \left(\sum_n \lambda_i^n(t) + \mu_i(t) + \omega_i\right)S_i, \tag{B.18}$$

$$\frac{dV_i}{dt} = v_i(t)\omega_{i-1}S_{i-1} + \omega_{i-1}V_{i-1} - \left(\delta + \sum_n \varepsilon^n\lambda_i^n(t) + \mu_i(t) + \omega_i\right)V_i, \tag{B.19}$$

$$\frac{dE_i^n}{dt} = \omega_{i-1}E_{i-1}^n + \lambda_i^n(t)(S_i + \varepsilon^n V_i) - (\sigma + \mu_i(t) + \omega_i)E_i^n + \iota_i^n(t), \tag{B.20}$$

$$\frac{dI_i^n}{dt} = \omega_{i-1}I_{i-1}^n + \sigma E_i^n - (\gamma + \mu_i(t) + \omega_i)I_i^n, \tag{B.21}$$

$$\frac{dR_i}{dt} = \omega_{i-1}R_{i-1} + \gamma\sum_n I_i^n - (\mu_i(t) + \omega_i)R_i. \tag{B.22}$$

The term $v_i(t)$ captures vaccination of individuals with the booster dose as they enter the second age class, $v_i(t) = (1-\alpha)p_2(t)$ for $i = 2$ and 0 otherwise, where $p_2(t)$ is the annual booster vaccine uptake defined above.

In order to fit to the data, we calculated the number of new cases in age group i in the time interval $[t-1,t)$. This is given by a piece-wise continuous integral,

$$C_i(t) = \int_{t-1}^{t} \gamma \sum_n I_i^n(s)ds. \tag{B.23}$$

As we fit the model to annual data, $t = 1$ year.

B.5 Full model of Erlang distributed waning immunity

To realize generate a peaked distribution of vacine derived immune duration, we utilized the method of stages (see [96] for details) and broke the vaccine compartment into three sub-classes (V^1, V^2, and V^3). The combined time spent transitioning between each of these sub-compartments corresponds to an Erlang distubted immune duration. Individuals stay within a vaccinated sub-class for an average of $\frac{1}{x\delta}$ years, such that the parameter $1/\delta$ retains its definition as average duration of vaccine derived immunity. As the Erlang distributed waning hypothesis only requires one genotype (genotype A) we have dropped the superscript n in the following equations. Aside from these two changes, the model equations have the same structure as above. For the first age cohort the equations are,

$$\frac{dS_1}{dt} = \big(1-(1-\alpha)p_1(t)\big)\nu(t)N(t) + 3\delta V_1^3 - \Big(\lambda_1(t)+\mu_1(t)+\omega_1\Big)S_1, \tag{B.24}$$

$$\frac{dV_1^1}{dt} = (1-\alpha)p_1(t)\nu(t)N(t) - \Big(3\delta+\mu_1(t)+\omega_1\Big)V_1^1, \tag{B.25}$$

$$\frac{dV_1^2}{dt} = 3\delta V_1^1 - \Big(3\delta+\mu_1(t)+\omega_1\Big)V_1^2, \tag{B.26}$$

$$\frac{dV_1^3}{dt} = 3\delta V_1^2 - \Big(3\delta+\mu_1(t)+\omega_1\Big)V_1^3, \tag{B.27}$$

$$\frac{dE_1}{dt} = \lambda_1(t)S_1 - (\sigma+\mu_1(t)+\omega_1)E_1, \tag{B.28}$$

$$\frac{dI_1}{dt} = \sigma E_1 - (\gamma+\mu_1(t)+\omega_1)I_1, \tag{B.29}$$

$$\frac{dR_1}{dt} = \gamma I_1 - (\mu_1(t)+\omega_1)R_1. \tag{B.30}$$

For age groups $i = 2,\ldots 5$, we have

$$\frac{dS_i}{dt} = (1-v_i(t))\omega_{i-1}S_{i-1} + 3\delta V_i^3 - \Big(\lambda_i(t)+\mu_i(t)+\omega_i\Big)S_i, \tag{B.31}$$

$$\frac{dV_i^1}{dt} = v_i(t)\omega_{i-1}S_{i-1} + \omega_{i-1}V_{i-1}^1 - \Big(3\delta+\mu_i(t)+\omega_i\Big)V_i^1, \tag{B.32}$$

$$\frac{dV_i^2}{dt} = \omega_{i-1}V_{i-1}^2 + 3\delta V_i^1 - \Big(3\delta+\mu_i(t)+\omega_i\Big)V_i^2, \tag{B.33}$$

$$\frac{dV_i^3}{dt} = \omega_{i-1}V_{i-1}^3 + 3\delta V_i^2 - \Big(3\delta+\mu_i(t)+\omega_i\Big)V_i^3, \tag{B.34}$$

$$\frac{dE_i}{dt} = \omega_{i-1}E_{i-1} + \lambda_i(t)S_i - (\sigma+\mu_i(t)+\omega_i)E_i, \tag{B.35}$$

$$\frac{dI_i}{dt} = \omega_{i-1}I_{i-1} + \sigma E_i - (\gamma+\mu_i(t)+\omega_i)I_i, \tag{B.36}$$

$$\frac{dR_i}{dt} = \omega_{i-1}R_{i-1} + \gamma I_i - (\mu_i(t)+\omega_i)R_i. \tag{B.37}$$

B.6 Derivation of R_0 and R_p

For the derivation of the reproductive numbers, we simplified the model described in Eqs.B.13–B.22. We assume: (1) There is a fixed flat demographic profile, such that $1/\nu = 1/\sum_i^5 \omega_i = 80$ years and $\mu_i(t) = 0$; (2) The model is evaluated at a constant vaccine coverage p_1 for the neonatal administration, and p_2 for the booster dose; (3) No seasonal forcing in transmission process. Dynamically, fixing these parameters removes all time-dependence in the model parameters, ensuring that our reproductive number calculations are not specific to a particular time point, but instead represent general quantities. Making these approximations to the VSEIR model equations with exponential waning (Eqs.B.13–B.22), for the first age class we have

$$\frac{dS_1}{dt} = (1-(1-\alpha)p_1)\nu N - \Big(\sum_n \lambda_1^n(t) + \omega_1\Big)S_1, \tag{B.38}$$

$$\frac{dS_1^W}{dt} = \delta V_1 - \Big(\sum_n \lambda_1^n(t) + \omega_1\Big)S_1^W, \tag{B.39}$$

$$\frac{dV_1}{dt} = (1-\alpha)p_1\nu N - \Big(\delta + \varepsilon_n \sum_n \lambda_1^n(t) + \omega_1\Big)V_1, \tag{B.40}$$

$$\frac{dE_1^n}{dt} = \lambda_1^n(t)(S_1 + S_1^W + \varepsilon_n V_1) - (\sigma + \omega_1)E_1^n, \tag{B.41}$$

$$\frac{dI_1^n}{dt} = \sigma E_1^n - (\gamma + \omega_1)I_1^n, \tag{B.42}$$

$$\frac{dR_1}{dt} = \gamma \sum_n I_1^n - \omega_1 R_1, \tag{B.43}$$

and for age classes $i = 2, \ldots 5$ we have

$$\frac{dS_i}{dt} = (1-v_i)\omega_{i-1}S_{i-1} - \big(\sum_n \lambda_i^n(t) + \omega_i\big)S_i, \tag{B.44}$$

$$\frac{dS_i^W}{dt} = \omega_{i-1}S_{i-1}^W + \delta V_i - \big(\sum_n \lambda_i^n(t) + \omega_i\big)S_i^W, \tag{B.45}$$

$$\frac{dV_i}{dt} = \omega_{i-1}V_{i-1} - \big(\delta + \sum_n \varepsilon_n\lambda_i^n(t) + \omega_i\big)V_i, \tag{B.46}$$

$$\frac{dE_i^n}{dt} = \omega_{i-1}E_{i-1}^n + \lambda_i^n(S_i + S_i^W + \varepsilon_n V_i) - (\sigma + \omega_i)E_i^n, \tag{B.47}$$

$$\frac{dI_i^n}{dt} = \omega_{i-1}I_{i-1}^n + \sigma E_i^n - (\gamma + \omega_i)I_i^n, \tag{B.48}$$

$$\frac{dR_i}{dt} = \omega_{i-1}R_{i-1} + \gamma\sum_n I_i^n - \omega_i R_i. \tag{B.49}$$

At an average vaccine uptake p_1 an p_2 respectively, the endemic equilibria for the system states for age class i = 1, ... 5 is given as

$$S_i^* = \begin{cases} \left(1-(1-\alpha)p_1\right)\frac{\nu N}{\omega_1} & \text{if } i=1, \\ \frac{\omega_1}{\omega_2}\left(1-(1-\alpha)p_2\right)S_1^* & \text{if } i=2, \\ \frac{\omega_{i-1}S_{i-1}^*}{\omega_i} & \text{otherwise,} \end{cases} \tag{B.50}$$

$$V_i^* = \begin{cases} \frac{(1-\alpha)p_1\nu N}{(\delta+\omega_1)} & \text{if } i=1, \\ \frac{\omega_1}{(\delta+\omega_2)}\left[(1-\alpha)p_2(S_1^*+S_1^{W*})+V_1^*\right] & \text{if } i=2, \\ \frac{\omega_{i-1}V_{i-1}^*}{(\delta+\omega_i)} & \text{otherwise,} \end{cases} \tag{B.51}$$

$$S_i^{W*} = \begin{cases} \frac{\delta V_1^*}{\omega_1} & \text{if } i=1, \\ \frac{\delta V_2^*}{\omega_2}+\frac{\omega_1}{\omega_2}\left[\left(1-(1-\alpha)p_2\right)S_1^{W*}\right] & \text{if } i=2, \\ \frac{\delta V_i^*}{\omega_i}+\frac{\omega_{i-1}S_{i-1}^{W*}}{\omega_i} & \text{otherwise,} \end{cases} \tag{B.52}$$

$$E_i^* = I_i^* = R_i^* = 0. \tag{B.53}$$

In the special case $p_1 = p_2 = 0$, i.e. without vaccination, we have $S_i^* = \frac{\nu N}{\omega_i}$ for each i with all other compartments equal to zero.

The basic reproductive number R_0 is defined as the average number of secondary infections that an index case can generate when introduced in a completely susceptible population. As our population is age stratified, we calculated R_0 using the next generation matrix method [47]. In the NGO analysis we focus on the disease-free equilibrium, therefore there are no competition effects between genotypes. For this reason we consider each genotype separately. To simplify the notation, we will therefore suppress the genotype index n in what follows.

First, we construct four vectors from the above system of differential equations. The first two, with elements $\mathscr{F}_i^1 = \lambda_i^n(S_i + S_i^W + \varepsilon^n V_i)$ and $\mathscr{F}_i^2 = 0$, correspond to the rate at which newly infected individuals enter the the compartment E_i and I_i respectively. Secondly we have two vectors, with elements $\mathscr{V}_i^1 = (\sigma + \omega_i - \omega_{i-1})E_i$ and $\mathscr{V}_i^2 = -\sigma E_i + (\gamma + \omega_i - \omega_{i-1})I_i$, that correspond to the rates at which infected individuals leave the compartments E_i and I_i respectively.

We then construct the matrices F^{11} , F^{12}, F^{21} and F^{22} by differentiating the first pair of vectors, $\mathscr{F}_i^1$ and $\mathscr{F}_i^2$, with respect to E_i and I_i,

$$F_{ij}^{11} = \frac{\partial \mathscr{F}_i^1}{\partial E_i} = 0, \tag{B.54}$$

$$F_{ij}^{12} = \frac{\partial \mathscr{F}_i^1}{\partial I_i} = q_i \frac{C_{i,j}}{N_j}(S_i + S_i^W + \varepsilon^n V_i), \tag{B.55}$$

$$F_{ij}^{21} = \frac{\partial \mathscr{F}_i^2}{\partial E_i} = 0, \tag{B.56}$$

$$F_{ij}^{22} = \frac{\partial \mathscr{F}_i^2}{\partial I_i} = 0. \tag{B.57}$$

A second set of matrices, V^{11} , V^{12}, V^{21} and V^{22}, are found similarly by differentiating the pair of vectors, $\mathscr{V}_i^1$ and $\mathscr{V}_i^2$,

$$V_{i,j}^{1,1} = \frac{\partial \mathscr{V}_i^1}{\partial E_i} = (\sigma + \omega_i - \omega_{i-1})\delta_{i,j}, \tag{B.58}$$

$$V_{i,j}^{1,2} = \frac{\partial \mathscr{V}_i^1}{\partial I_i} = 0, \tag{B.59}$$

$$V_{i,j}^{2,1} = \frac{\partial \mathscr{V}_i^2}{\partial E_i} = -\sigma\delta_{i,j}, \tag{B.60}$$

$$V_{i,j}^{2,2} = \frac{\partial \mathscr{V}_i^2}{\partial E_i} = (\gamma + \omega_i - \omega_{i-1})\delta_{i,j}, \tag{B.61}$$

where $\delta_{i,j}$ is the Kronecker's delta function. The two block matrices $\mathbf{F} = \begin{bmatrix} F^{11} & F^{12} \\ F^{21} & F^{22} \end{bmatrix}$ and $\mathbf{V} = \begin{bmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{bmatrix}$ form the NGO, $\mathbf{K} = (\mathbf{F}\mathbf{V}^{-1})$. Finally, the basic reproductive number, R_0, is given by the spectral radius of the NGO evaluated at the disease-free equilibrium in a fully susceptible population [47]. In addition, we calculate the vaccine reproductive number (R_p) which we define to be the spectral radius of the NGO evaluated at the disease-free equilibrium in a population with a given vaccine uptake.

In general, this eigenvalue problem lacks an analytical solution. We therefore use numerical methods to solve it and find R_0 and R_p.

B.7 Sensitivity analysis

The model described in Eqs.B.13-B.22 was modified and we considered a system with the circulation of a single pathogen. For analysing the effects of waning intensity in vaccine derived immunity we introduced additional state variables S^W and I^W, to explicitly keep track of contribution of individuals in whom the vaccine immunity has waned. Table B.3 documents the state-variable definitions for this model.

Table B.3: Model Compartments for the i^{th} age-cohort

Compartment	Definition
S_i^S	Unvaccinated Susceptible individuals
S_i^W	Vaccinated Susceptible individuals
E_i^S	Unvaccinated Exposed individuals
E_i^W	Vaccinated Exposed individuals
I_i^S	Unvaccinated Infectious individuals
I_i^W	Vaccinated Infectious individuals
V_i	Vaccinated individuals
R_i	Recovered individuals

The force of infection experienced by age group i for the modified model is given by

$$\lambda_i(t) = q_i \sum_{j=1}^{5} C_{i,j} \frac{I_j^S + I_j^W}{N_j} \phi_{i,j}(t), \tag{B.62}$$

where $I_j = I_j^S + I_j^W$ are the total number of infectious individuals in age group j. Other terms in Eq.B.62 are identical to the ones described in Eq.B.10. For the first age class the equations describing the transmission dynamics are given as

$$\frac{dS_1^S}{dt} = \big(1-(1-\alpha)p_1(t)\big)\nu(t)N(t) - \Big(\lambda_1(t)+\mu_1(t)+\omega_1\Big)S_1^S, \quad \text{(B.63)}$$

$$\frac{dS_1^W}{dt} = \delta V_1 - \Big(\lambda_1(t)+\mu_1(t)+\omega_1\Big)S_1^W, \quad \text{(B.64)}$$

$$\frac{dV_1}{dt} = (1-\alpha)p_1(t)\nu(t)N(t) - \Big(\delta+\mu_1(t)+\omega_1\Big)V_1, \quad \text{(B.65)}$$

$$\frac{dE_1^S}{dt} = \lambda_1(t)S_1^S - \Big(\sigma+\mu_1(t)+\omega_1\Big)E_1^S \quad \text{(B.66)}$$

$$\frac{dE_1^W}{dt} = \lambda_1(t)S_1^W - \Big(\sigma+\mu_1(t)+\omega_1\Big)E_1^W \quad \text{(B.67)}$$

$$\frac{dI_1^S}{dt} = \sigma E_1^S - \Big(\gamma+\mu_1(t)+\omega_1\Big)I_1^S, \quad \text{(B.68)}$$

$$\frac{dI_1^W}{dt} = \sigma E_1^W - \Big(\gamma+\mu_1(t)+\omega_1\Big)I_1^W, \quad \text{(B.69)}$$

$$\frac{dR_1}{dt} = \gamma\Big(I_1^S+I_1^W\Big) - \Big(\mu_1(t)+\omega_1\Big)R_1. \quad \text{(B.70)}$$

For the age groups i = 2, ..., 5 we have

$$\frac{dS_i^S}{dt} = (1-v_i)\omega_{i-1}S_{i-1}^S - \Big(\lambda_i(t)+\mu_i(t)+\omega_i\Big)S_i^S, \tag{B.71}$$

$$\frac{dS_i^W}{dt} = (1-v_i)\omega_{i-1}S_{i-1}^W + \delta V_i - \Big(\lambda_i(t)+\mu_i(t)+\omega_i\Big)S_i^W, \tag{B.72}$$

$$\frac{dV_i}{dt} = v_i\omega_{i-1}(S_{i-1}+S_{i-1}^W) - \Big(\delta+\mu_i(t)+\omega_i\Big)V_i, \tag{B.73}$$

$$\frac{dE_i^S}{dt} = \omega_{i-1}E_{i-1}^S + \lambda_i(t)S_i^S - \Big(\sigma+\mu_i(t)+\omega_i\Big)E_i^S, \tag{B.74}$$

$$\frac{dE_i^W}{dt} = \omega_{i-1}E_{i-1}^W + \lambda_i(t)S_i^W - \Big(\sigma+\mu_i(t)+\omega_i\Big)E_i^W, \tag{B.75}$$

$$\frac{dI_i^S}{dt} = \omega_{i-1}I_{i-1}^S + \sigma E_i^S - \Big(\gamma+\mu_i(t)+\omega_i\Big)I_i^S, \tag{B.76}$$

$$\frac{dI_i^W}{dt} = \omega_{i-1}I_{i-1}^W + \sigma E_i^W - \Big(\gamma+\mu_i(t)+\omega_i\Big)I_i^W, \tag{B.77}$$

$$\frac{dR_i}{dt} = \omega_{i-1}R_{i-1} + \gamma\Big(I_i^S+I_i^W\Big) - \Big(\mu_i(t)+\omega_i\Big)R_i. \tag{B.78}$$

B.8 Differential evolutionary algorithm

The basic principle of DE algorithms revolves around evolution of fitness of a solution by reproducing constituents of a population (g). These evolved agents are called trial mutants. For all hypotheses, a large population of parametric configurations ($g = 500$) was used to initialize the optimizer. The initial population was generated using a Sobol design to ensure efficient sampling of the parametric space [190, 191]. The parameter $F = 0.6$ is the moderate scaling factor, which controls how fast mutants mutate.

In this study, we implemented the standard differential evolution algorithm [193]. Owing to the complexity of the estimation problem, a very

strict convergence criterion was set to ensure convergence to a global maximum (*steptol* = 750 iterations, $reltol = 10^{-4}$, $MAXFES = 1 \times 10^5$, $CR = 0.9$) [84]. Here $MAXFES$ denotes maximum number of objective function evaluations, *reltol* is the relative tolerance among function evaluations, and *CR* denotes the cross over rate. Fig. B.5 describes the details of the implemented algorithm.

Algorithm 1: Differential Evolution

Input: g: Population size; F: Mutation factor; CR: Cross Over Probability; $MAXFES$: Maximum number of objective function evaluations, D: Total number of parameters.

INITIALIZE $w = 0$; $FES = 0$; *Initialize a Sobol sample of size g from the search space.*
while $FES < MAXFES$ **do**
  for $i \leftarrow 1$ *to* g **do**
    for $j \leftarrow 1$ *to* D **do**
      DRAW *3 random samples* $x_{s_1,w,j}$, $x_{s_2,w,j}$, $x_{s_3,w,j}$, *from the target population g, such that* $s_1, s_2, s_3, i \in \{1, 2, ..., g\}$ *&* $s_1 \neq s_2 \neq s_3 \neq i$
      MUTATE *Using the 3 samples generate the donor as* $v_{i,w+1,j} = x_{s_1,w,j} + F \times (x_{s_2,w,j} - x_{s_3,w,j})$
      CROSSOVER *Trial sample* $u_{i,w+1,j}$ *is generated as either from the target* $x_{i,w,j}$ *or the donor vector* $v_{i,w+1,j}$ *such that*
      $$u_{i,w+1,j} = \begin{cases} v_{i,w+1,j} & r_{i,j} \leq CR \text{ or } j = j_{rand} \\ x_{i,w,j} & otherwise. \end{cases}$$
      where $r_{i,j} \sim U(0,1)$ *is a uniformly distributed random number and* $j_{rand} \in \{1, 2, ..., D\}$ *is a random integer to ensure that* $u_i \neq x_i$ *for all cases*
    end
    EVALUATE *For the D-dimensional objective function f:*
    if $f(\vec{u}_{i,w+1}) > f(\vec{x}_{i,w})$ **then**
      Replace the D-dimensional target vector $\vec{x}_{i,w}$ *with trial vector* $\vec{u}_{i,w+1}$
    end
  end
  FES = FES + g
  w = w + 1
end

Figure B.5: Pseudo code for the DE implementation