Uncovering the Mechanisms Governing Childhood Infectious Disease Dynamics

by

Kevin M Bakker

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2017

Doctoral Committee:

Assistant Professor Marisa C. Eisenberg, Co-Chair
Associate Professor Timothy Y. James, Co-Chair
Associate Professor Meghan A. Duffy
Professor Edward L. Ionides
Professor Aaron A. King
Professor Mercedes Pascual, University of Chicago

Kevin M Bakker
bakkerke@umich.edu
Orcid ID: 0000-0002-7084-9291

DEDICATION

To my three dogs; Puck, Fluff, and Hermione for keeping me sane, delivering unconditional love, and always providing a welcome distraction.

ACKNOWLEDGEMENTS

To say this dissertation wouldn't be possible without the following individuals would be an understatement. As you will soon read, I knew next to nothing about anything included in this dissertation when I arrived at the University of Michigan. I consider my two greatest academic mentors to be Micaela Martinez and Pej Rohani. The two of them provided me a crash course in mathematics, modeling, and infectious diseases. They continue to mentor me, along with Mercedes Pascual, Aaron King, Ed Ionides, Marisa Eisenberg, Bob Woods, and my various co-authors, who further refined my thought processes and provded new ideas and direction on various projects. I also want to note the importance of the members of the 2011-2013 Rohani lab (notably Julie Blackwood, Sourya Shrethsa, Micaela Martinez, Vicki Brown, Eli Goldwyn, and Maria Riolo) who kindly accepted me and had the patience to teach me some mathematics and modeling.

Many thanks to my graduate school friends (alphabetical); Kevin Boehnke, Rachel Cable, Katherine Crocker, Paul Glaum, Alison Gould, Celia Miller, Jill Myers, Jeff Shi, Alex Taylor, Pascal Title, and many others, for their friendship, advice, and ability to get me away from work. I would also like to thank my parents, who worked extremely hard to create opportunities for my sister and I to succeed, their continued support has made this possible.

Finally I need to acknowledge my three best friends for whom this, and all future work, is dedicated to; Hermione, Fluff, and Puck.

PREFACE

In the fall of 2010 I was hired by Pejman Rohani (Pej) to work as his lab manager digitizing seasonal pertussis data and to help with other data-related tasks. I was hired despite having no background in mathematics, modeling, or infectious diseases. I had recently left the University of Georgia, where I studied marine sciences. After a few weeks of work I thought I had figured out why childhood infections were seasonal - children were being born seasonally and then infected with various pathogens when their maternal antibodies had waned! This was my first descent into data-mining, and I loved it. Under the mentorship of Pej, and the advice and support from the members of his lab I decided to apply for the Ph.D. program.

I experienced tremendous intellectual growth during my time at the University of Michigan. With the help of generous colleagues, administrators, co-authors, fellow graduate students, and most importantly academic mentors, I was able to grasp the basics of data analysis and modeling, as well as write a few publications. I will never be able to thank those individuals enough for the time they spent instructing and assisting me along the way.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendices

ABSTRACT

To mitigate infectious diseases via vaccination, quarantine, or sanitation, it is necessary to understand the mechanisms by which pathogens are spread from the environment or person to person. This work provides a spatiotemporal characterization of human birth seasonality across the northern hemisphere, in which births replenish the pool of individuals susceptible to disease. Additionally I investigated the role of digital epidemiology when no clinical case data were available, and scrutinized the hypothesized mechanisms behind herpesvirus transmission and reactivation. While birth rates have previously been shown to impact infectious disease dynamics, birth timing and birth amplitude had not been previously explored in this context. By digitizing 78 years of monthly birth data for every continental US state, I revealed rich intra-annual patterns in birth seasonality in the United States, which I then incorporated into theoretical disease transmission models to demonstrate that birth timing can either increase or decrease the magnitude of an outbreak. Secondly, I established that birth amplitude could alter the timing and magnitude of disease outbreaks. My second research chapter tested whether Google trends could be used to predict chickenpox outbreaks. This work determined chickenpox searches were seasonal around the world, and correlated well to clinical case reports of chickenpox, especially in countries that did not immunize. I built a forecasting model, which was able to accurately predict the timing and magnitude of chickenpox outbreaks, and further demonstrated that in countries that

immunize, chickenpox search seasonality had all but disappeared, validating the effectiveness of chickenpox immunization. My third research chapter explored herpesviruses transmission and reactivation dynamics using varicella zoster virus (chickenpox and shingles) as a study system. This work revealed seasonal patterns in shingles reactivation, which was previously unknown, and the strong association between shingles and it's likely driver of reactivation - ultraviolet irradiation. I then fit both the transmission and reactivation dynamics using a model, and examined what effect immunization might have had on disease dynamics if it had been implemented when licensed. I revealed that more than a half-million cases of chickenpox could have been prevented during our 9-year study period. The results suggest this could be accomplished with little impact on shingles dynamics, however this is difficult to determine without further study. This work has the potential to drastically reduce global disease burden by informing policy makers with: (1) an understanding of the global variation in birth seasonality which allows for locally tailored immunization campaigns, (2) model-based outbreak predictions and counterfactuals when data are lacking, providing opportunity for prevention, and (3) insights into the mechanisms driving herpesvirus transmission and reactivation.

CHAPTER I

Introduction

Seasonality exists in almost all ecological systems. The earth's tilt and rotation around the sun drives seasonal changes in temperature, day length, rainfall, and other abiotic factors. These abiotic components in turn drive the seasonality of biotic factors such as phytoplankton blooms, migration, and breeding cycles. Biotic factors such as breeding have previously been shown to be strongly tied to season [28], and many infectious diseases have been shown to be driven by seasonal mechanisms [2,149]. However, because seasonality is so widespread, it is difficult to discern correlation from causation without a dedicated effort involving information from multiple locations over many years. By examining inter-annual deviations which occur from year-to-year, we can inform models and inferences about these processes. However, even then, it remains difficult to understand, as relationships between mechanism and seasonality may vary across a spatial landscape. Thus, one must combine seasonality in a spatial context with these inter-annual deviations to understand the processes regulating seasonality. This research; (1) characterizes human birth seasonality on a global scale, examining the intra- and inter-annual heterogeneity, verifying its potential to alter the magnitude and timing of childhood infectious disease outbreaks, (2) utilizes alternative data and methodologies to identify previously unknown spatial complexities of childhood infectious disease dynamics using digital epidemiology, and (3) reveals previously unknown herpesvirus reactivation dynamics, and diligently links it to seasonal environmental variables.

In disease ecology one can use long time series with high spatial resolution to understand the mechanisms driving disease persistence and transmission. For example, the spatial variation in the timing of pertussis outbreaks in the US during the 1950s led Choisy and Rohani to discover that nearby cities had synchronized outbreaks that catalyzed outbreaks in local regions that then moved inward towards the interior of the country from both coasts [47]. Similarly, Grenfell et al. [89] demonstrated that the historical persistence of measles in England and Wales was driven by seasonal outbreaks in larger cities which re-seeded smaller populations where the virus had previously gone extinct. These source-sink dynamics have also been identified in historical US polio epidemics [133] Finally, human demographics have been shown to play an important role in early childhood infections, where regions with higher birth rates had their peak rotavirus season months prior to regions with lower birth rates [151]. These studies combined long-term, spatially replicated data with disease-specific transmission models to discern what drove the seasonality and persistence of disease transmission. Spatiotemporal studies such as those mentioned are invaluable to discern the mechanisms which drive disease transmission; however high resolution data across a long time series are rare, which has limited our ability to perform thorough and extensive research on the mechanisms driving disease dynamics.

Many studies have been performed attempting to link causative mechanisms to the outbreak dynamics of childhood pathogens. The population dynamics of infectious diseases in a variety of organisms show seasonality in their incidence [2,88]. Childhood infectious diseases such as measles, mumps, rubella, diphtheria, and chickenpox have been previously identified to occur seasonally [9,88,136]. It is understood that seasonal changes in behavior, such as a change in contact rates [127] can drive the magnitude and frequency of epidemics in directly transmitted diseases such as measles, mumps, and rubella [20, 76, 77, 180]. While pre-20th century epidemics of typhoid fever and polio may be attributed to poor water treatment, by the early 1920's all major cities had proper water treatment facilities, indicating that contaminated water reservoirs were unlikely the major mechanism of transmission in these

two pathogens [183]. A strong correlation between clinical cases of typhoid fever and temperature has been identified and proposed as a mechanism, but not yet rigorously tested using models [183, 195]. Similarly there is evidence of increased polio transmission at higher temperatures [193, 232] and higher humidity [94]. Meanwhile, in directly transmitted respiratory infections such as influenza, the seasonal mechanism has been shown to include humidity [128, 190] and dynamic changes social contacts [68]. Other winter respiratory infections such as scarlet fever and diphtheria, have been shown to be affected by temperature [115], dry summers [142], and humidity [206], but not social contacts [142]. Despite the evidence demonstrating that seasonal forcings, such as environmental and behavioral variables, have been shown to shape disease dynamics in measles, mumps, and rubella [69], mechanisms driving seasonal patterns of incidence remain poorly understood, and untested, for many notifiable childhood diseases, including typhoid fever, polio, scarlet fever, and diphtheria.

To better understand childhood infectious disease dynamics, one must first understand the role demographics play in these outbreaks. Demographics consist of many aspects, but here I will focus on birth dynamics, concentrating on birth seasonality. Birth seasonality consists of three components, the birth rate, the strength of the birth pulse, and the timing of when that birth pulse occurs. Reproduction is seasonal across all living organisms, from plants [27], and insects [201, 236], to reptiles [105, 125], birds [56, 231], and mammals [6, 30], including humans [132]. The first documentation of seasonality in human births goes back to European studies in the early 19th century [154, 220]. In developed countries located in the northern hemisphere, modern humans exhibit a strong annual rhythm in reproduction tied to latitude [132]. However, this pattern weakens as you move closer to the equator, and there remains a lack of research focused on birth seasonality in developing countries. There have been many proposed mechanisms of human birth seasonality including environmental, biological, and socioeconomic variables.

Environmental mechanisms related to latitude, such as temperature and photoperiod, are

the most extensively studied mechanisms driving birth seasonality. Temperature has a strong negative correlation with conceptions in many countries [13, 14, 45, 160, 169, 198]; however, since 1950 conception timing has become independent of ambient temperatures [163, 164], potentially due to industrialization [137, 184]. Additionally, in many countries temperature alone cannot fully explain birth seasonality, and humidity in regions where temperatures exceed 25 degrees C has been linked to reduced conceptions [130]. Roenneberg and Aschoff proposed that the seasonality of births seemed to be a function either temperature or photoperiod, where regions with extreme temperatures typically had biannual birth peaks [163, 164]. Others found that intercourse increased in cooler temperatures and that both nutrition and temperature drove ovulation cycles [14]. Meanwhile, it has been argued that in eskimos, photoperiod explains the variation of births [71]. Others have discovered that the seasonality of conception is more likely driven by socioeconomic behaviors combined with temperature rather than just photoperiod itself [130]. In a long-term study of births in 20th century Scotland, mechanisms varied depending on social status, with temperature being the main driver pre-industrialization, but declining sharply as the standard of living increased [171]. The majority of studies on birth rhythms have been based on long time series within a specific geographical region [52] or on a few short time series from selected regions across the globe [10]. Thus, while photoperiod, temperature, and latitude have all been shown to have an effect on the seasonality of births, no single environmentally related mechanism can fully explain global birth seasonality, specifically in developing countries where social factors seem to play a larger role.

Social and cultural mechanisms have also been examined for their potential role in driving the seasonality of birth. In historically Catholic countries, the spring peak in births has been linked to the traditional July/August wedding season [163]. But elsewhere, the annual distribution of marriages correlates only weakly to births occurring 9 to 15 months later [179]. Social status has been mentioned as a potential mechanism of birth seasonality, with high-income groups having more seasonality than low-income groups [22]. However, historical

comparisons within country of urban and rural populations [163], as well as differences in occupation (which included seasonal laborers, farmers, and office jobs) revealed little difference in the timing of the birth peak [198] yet large differences in the amplitude of birth seasonality, with rural regions having larger amplitudes than those in urban cores [14]. Recently, a study proposed that global populations are a mixture of photoperiodic responders and non-responders, thus providing an explanation as to why specific mechanisms seem to explain the seasonality in certain locations, but not others [31]. While the mechanisms driving birth seasonality continue to be debated, the one clear result is that industrialization, and thus the increased use of air conditioning, indoor work, and fewer seasonal jobs, has diminished the amplitude of birth seasonality [137, 162–164, 184]

Jürgen Aschoff remarked that the lack of experimental possibilities to investigate birth seasonality must be compensated by a comparative analysis of data that are widely distributed in time and area [7]. Unlike the annual rhythmicity of animal reproduction, the rhythm of human reproduction only becomes evident in large population studies [163]. To date there has been a scarcity of globally distributed, long-term, peer-reviewed studies of birth seasonality due to the lack of available data.

Due to this dearth of data, for chapter 2 I digitized 78 years of monthly birth data for each state in the continental U.S. and more than 100 other countries located in the northern hemisphere in order to examine the role of birth seasonality on childhood infectious disease dynamics. As mentioned earlier, birth rates have been shown to play a role in childhood infectious disease dynamics by altering the supply of susceptible individuals for infection. A higher birth rate allows for larger, or annual, outbreaks, while lower birth rates produces weaker, or multi-enniel outbreaks. By examining intra-annual heterogeneity in births, I will be able to reveal both the inter- and intra-annual variability of birth seasonality across a spatiotemporal landscape. Then, I will build a theoretical model to test what role birth seasonality, in terms of peak birth timing and birth amplitude, plays in childhood infectious disease dynamics This work examines birth seasonality across a spatiotemporal

scale, providing accurate demographic data, which is critical for understanding the potential mechanisms governing childhood infectious disease dynamics.

Because high resolution, long-term time series data aren't always available, and in some cases real-time data are required, other methods need to be implemented. Digital epidemiology – the application of real-time mobile or social digital data– is one such method. The increase in big data, which includes digital epidemiology, has created an enormous database of mineable data. This data can be harnessed to assist in understanding disease dynamics, when local clinical case data are unavailable. Digital epidemiology is perhaps best known for its ability to forecast influenza outbreaks [84], an ability that was quickly destroyed due to its inability to predict outbreaks once the research made national news, because of the overwhelming amount of 'influenza' searches [122]. However it has since been successfully implemented in observing the movement of individuals during an outbreak [17], indicating its potential.

Chapter 3 focuses on utilizing digital epidemiology to identify patterns of an extremely common, easily identifiable disease - chickenpox. As with many other diseases, there remains a lack of spatiotemporal data chickenpox due to its non-notifiable status. Disease surveillance systems largely focus on infectious diseases with high mortality, while benign diseases often go unreported. Using chickenpox as an example, I demonstrated that internet queries can be used as a proxy for disease incidence when reporting is lacking. This work demonstrates that alternative data can be used to understand childhood infectious disease dynamics across a broad geographic and temporal landscape when clinical case data are lacking.

Herpesviruses are defined by their latent, recurring infections. All human herpesviruses are transmitted, typically in childhood, and many will reactivate later in life, sometimes multiple times. This is the only known family of pathogens where one is symptomatic from an initial transmission and, potentially, multiple reactivations. Unfortunately, our understanding of herpesvirus dynamics is limited because chickenpox is the only notifiable disease, and those

data are also rare. Chickenpox is the symptomatic manifestation of disease caused by the varicella zoster virus (VZV), which retreats to the nervous ganglia once immune antibodies are produced, and the infection is fought off. However, as with all herpesviruses, the virus is not cleared. In approximately 20-30% of individuals, reactivation will occur, causing shingles. Due to shingles and other herpesviruses being non-notifiable, there have been no population-level studies examining the mechanisms by which they are reactivated.

Furthermore, a highly effective and safe chickenpox vaccine was licensed in 1995, however it remains contentious and most countries have not yet implemented it. It is controversial because of the potential interaction between chickenpox and shingles. It has been shown in most childhood diseases, that exposure to a pathogen after one has been previously infected will 'boost' their immunity by stimulating an antibody response. Antibody boosting, in the context of VZV, is thought to increase VZV antibodies in adults when exposed to children recently infected with chickenpox, thus providing a 'boost' to their immunity. By vaccinating children against VZV, this boosting would not occur, as fewer natural chickenpox infections would limit the boosting, thus increasing the rate at which antibodies wane. This waning, in turn, would theoretically increase shingles incidence. Thus, to introduce VZV immunization (a booster dose is also available for adults), one must carefully consider all potential dynamical impacts for both chickenpox and shingles.

In chapter 4, I examined herpesvirus dynamics with the goal of uncovering the mechanisms driving transmission and reactivation. While not spatiotemporal, it is the first work to examine potential mechanistic drivers of herpesvirus reactivation dynamics at the population-level. By understanding these mechanisms, this work will provide policy makers on the impacts of various immunization strategies.

This research investigates potential processes that drive infectious disease seasonality across space and time. First, by examining the intra- and inter-annual variation in birth seasonality, which seasonally contribute to the susceptible pool for all childhood infectious

diseases, I will expose one of the complexities driving variation of disease dynamics across geographic regions. Second, by utilizing alternative data and methods to identify the underlying seasonality of childhood infectious disease, I will demonstrate the usefulness of surrogate data when clinical case reports are lacking. Third, by analyzing an obscure dataset I will unveil the undiscovered mechanisms driving the seasonality of herpesvirus transmission and reactivation. Finally, I will assemble appropriate biologically based models to examine; (1) the effects birth seasonality has on the seasonal outbreaks of childhood infectious diseases, (2) whether alternative data can be used to forecast childhood infectious disease outbreak seasonality, and (3) what underlying mechanisms drive the seasonality of herpesvirus transmission and seasonality.

CHAPTER II

Human Birth Seasonality: Latitudinal Gradient and Interplay with Childhood Disease Dynamics

2.1 Preface

This chapter of my dissertation was published in the *Proceedings of the Royal Society: Biological Sciences* in 2014 (DOI: 10.1098/rspb.2013.2438). Authors include M. Martinez-Bakker, A. King, P. Rohani, and myself. I was responsible for the collection and analyses of the human birth data, which identified birth seasonality in the United States and other developed countries in the Northern Hemisphere. Together with my co-first author, M. Martinez-Bakker, we developed a theoretical model to explore the range of outbreak possibilities under various birth seasonality conditions. It was during this work when I received my first hands-on training with *pomp* [112], which is an R package for fitting process-based models to time series. *Pomp* is computational in nature, and can be utilized to efficiently explore parameter space. I utilized this model-fitting software in each of my additional chapters. M. Martinez and I wrote the manuscript together under the guidance of P. Rohani and A. King; additional analyses and additional modeling and analyses are included in Appendix A. *Pomp* code for this chapter, as well as Chapters 3 & 4 are included in Appendix D.

2.2 Abstract

More than a century of ecological studies have demonstrated the importance of demography in shaping spatial and temporal variation in population dynamics. Surprisingly, the impact of seasonal recruitment on infectious disease systems has received much less attention. Here, we present data encompassing 78 years of monthly natality in the United States, and reveal pronounced seasonality in birth rates, with geographic and temporal variation in both the peak birth timing and amplitude. The timing of annual birth pulses followed a latitudinal gradient, with northern states exhibiting spring/summer peaks and southern states exhibiting fall peaks, a pattern we also observed throughout the northern hemisphere. Additionally, the amplitude of US birth seasonality was more than two-fold greater in southern states versus those in the north. Next, we examined the dynamical impact of birth seasonality on childhood disease incidence using a mechanistic model of measles. Birth seasonality was found to have the potential to alter the magnitude and periodicity of epidemics, with the effect dependent on both birth peak timing and amplitude. In a simulation study, we fitted an SEIR model to simulated data, and demonstrated that ignoring birth seasonality can bias the estimation of critical epidemiological parameters. Finally, we carried out statistical inference using historical measles incidence data from New York City. Our analyses did not identify the predicted systematic biases in parameter estimates. This may be due to the well-known frequency-locking between measles epidemics and seasonal transmission rates or may arise from substantial uncertainty in multiple model parameters and estimation stochasticity.

2.3 Introduction

The ubiquity of seasonal variation in the incidence of infectious diseases has driven much epidemiological research focused on understanding the responsible undelrying mechanisms [77, 127, 136, 166, 192]. Surprisingly, there remains much uncertainty regarding the drivers

of seasonal incidence for numerous infections including polio, pertussis, scarlet fever, diphtheria, rotavirus, among others [83, 136, 165, 209]. Early work on diphtheria and measles implicated elevated contact rates among children in school as the driver of pulsed transmission [192, 203–205], leading to much emphasis on school-term forcing [68, 77, 108, 127, 136]. More recently, additional mechanisms of seasonal transmission have been identified, including climatic drivers of pathogen survival [94], transmission [128, 190], and vector activity [126,129], seasonal host migration [19], and seasonal fluctuations in host immunity [36,65]. Here, we propose that seasonality in host recruitment rates may also shape epidemiology. This is a possiblity that has been appreciated in studies of wildelife diseases [3,23,67,100,108]. For instance, studying cowpox virus in voles, Begon *et al.* found that susceptible recruitment is seasonal, and higher breeding-season birth rates delayed epidemic peaks [16]. However, despite evidence demonstrating the importance of host demography in recurrent epidemics [32,69,151,199], and the ubiquitous appreciation of seasonal reproduction in broader ecology and evolution [80], we submit that a deep understanding of the dynamical impact of birth seasonality on infectious diseases of humans is currently lacking.

To explore this phenomenon, we first characterize the landscape of birth seasonality in modern human populations, and second determine if/how it can impact epidemic dynamics, particularly for immunizing childhood infections. Some precedent has been set in the field of demography, with seasonal variation in human births first documented in the early 1800s [154, 220] and currently recognized as a global characteristic of humans [15, 50, 52, 117]. Early studies of vital statistics in various US regions established a national-level seasonal pattern of births with troughs in the spring and peaks in autumn [52, 169]. Subsequent research has focused on either a few locations over long time periods, or many locations over short time periods. Collectively, these studies showed that northern and southern states have differences in their seasonal birth amplitude [29, 50, 52, 117, 169] and birth/conception minima [29, 138, 184]. Studies of births in Africa and Asia have been sparse, but seasonal

peak-trough differences in conception ranging from 11–64% have been documented in Africa and 8–58% in Asia [15]. To date there has been no long term, large scale, spatiotemporal analysis of births in either the US or worldwide.

We have compiled the most extensive spatiotemporal data set on human births to date, and explored the effect of birth seasonality on childhood disease incidence using simulated and empirical data. Measles was chosen because it is the paradigmatic example of a childhood disease, with two key features: (i) a low mean age of infection during the pre-vaccine era, with infections occurring in the youngest age group, the size of which is tightly linked to the birth rate, and (ii) seasonal transmission, which is a feature of many childhood diseases. We focus on birth seasonality in the presence of seasonal transmission to explore the interplay between these two forces. Our novel demographic data set is comprised of birth records across the globe, consisting of 7.3×10^8 births. Specifically, these data consist of monthly births spanning a 78-year period (1931–2008) for each state in the continental US along with over 200 additional time series from countries spanning the Northern Hemisphere. We have analyzed these data in combination with a transmission model and statistical inference tools to examine the dynamic implications of birth seasonality on childhood infection.

2.4 Data

Monthly state-level time series of live births from 1931–2008 were downloaded from US Vital Statistics [216] and digitized. Annual state-level population size data were collected from the US Census Bureau [215] and used to construct monthly time series of birth rates per 1000 individuals per month. Worldwide monthly births were retrieved from the United Nations database [214] and filtered for countries containing at least 5 years of consecutive data.

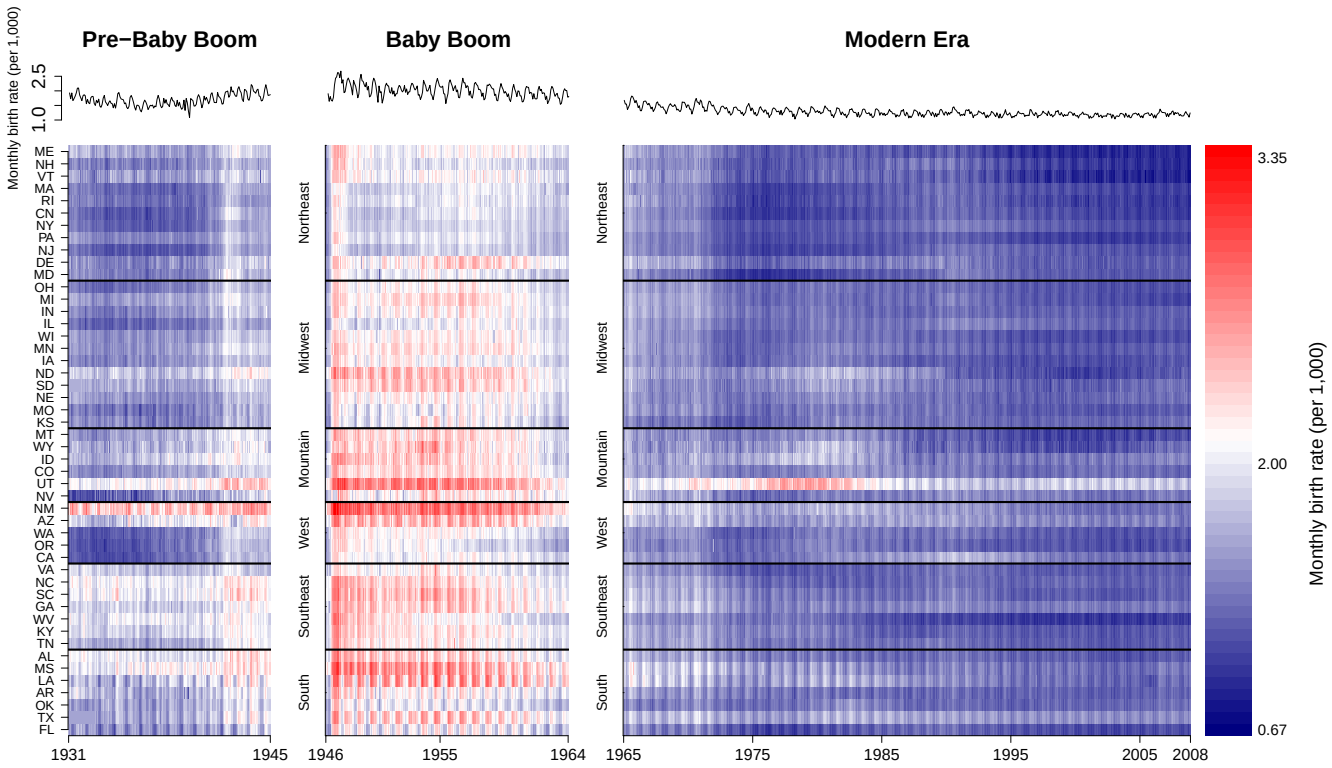


Figure 2.1: Temporal patterns of birth rates (per 1000 individuals per month) in the US organized by geographic region, separated into three eras: Pre-Baby Boom (1931–1945), Baby Boom (1946–1965), and Modern Era (1965–2008). The time series for Louisiana is plotted at the top as an example.

The US data were split into three eras, to account for the baby boom: (i) Jan 1931–Dec 1945, which we term the Pre-Baby Boom Era, (ii) Jan 1946–Dec 1964, the Baby Boom Era, and (iii) Jan 1965–Dec 2008, the Modern Era. To test for periodicity, a wavelet spectral analysis [210] was performed independently for each US state in each era and for each country in the global data set. The significance of each period was tested by comparing the power of each period against a noise background, using a lag–1 autocorrelation test. For each data series significant at a 1 year period, phase angle time series were constructed to determine the timing of birth peaks occurring at 11–13 month intervals. Independently, seasonal decomposition was run on all data series to filter out noise, and the seasonal amplitude was calculated by taking one-half the difference between the maxima and minima, measured as a percent of the annual mean (Eq A.1–A.5). Inter-annual variation was examined by analyzing

the percent change in mean birth rates from one year to the next.

2.5 Measles Models

We used a discrete-time SEIR model of measles adapted from Earn *et al.* [69], with school-term forcing based on the England & Wales school year (Table A.1; [108]). We incorporated seasonality in births using a sine function with varying amplitude and phase. The equations describing the model (Eq A.6–A.13) and parameter values (Table A.2) are provided in the Supporting Information.

We conducted statistical inference on both simulated and empirical measles incidence data to test for the effect of birth seasonality on epidemic dynamics and parameter estimation. This work aimed to answer the question: how does the omission of birth seasonality affect the precision and bias of estimated parameters? Using a Markovian analogue of our SEIR model (Eq A.6–A.14), three time series were generated assuming the following parameterizations: a birth peak day of either 162, 295, or 351 and a 28% birth amplitude (see Table A.3 for parameter values). For each simulated time series, our stochastic SEIR model was fit assuming constant births (birth amplitude set to 0%) and an unknown mean transmission rate. All other parameters were assumed known. Thus, the only free parameter was the mean transmission rate, which is directly proportional to R_0. The transmission rate was profiled and the likelihood was calculated using a particle filter (Appendix A: Materials & Methods) [112].

In order to test for bias in parameter estimation using real world data, we utilized historical measles case reports from New York City. These data are from the Baby Boom Era, when the birth amplitude was low, approximately 7% for the state of New York. To account for maternal antibodies, we fit models which lagged births 3, 6, or 9 months (see Appendix A for methods). We used both maximum likelihood via iterated filtering [112] and the TSIR

methodology of Finkenstadt & Grenfell [20, 78] to quantify the impact of seasonal births on parameter estimates (Fig 2.5).

2.6 The Seasonal Timing of Births

Figs 2.1 and A.3 provide an overview of birth rates in the US. Most states had significant seasonal (1 year) birth pulses in the Pre-Baby Boom Era, while all states showed significant birth seasonality in the Baby Boom and Modern Eras. Of the 210 time series analyzed outside the US, 132 (63%) had significant birth seasonality. Most of the locations for which seasonality was not significant were short time series (5–7 years) or countries with less than 100 births per month.

We observed a latitudinal gradient in the timing of the birth peak across the US; this same gradient was observed throughout the Northern Hemisphere (Figs 2.2 & 2.3). In general, the birth peak occurs earlier in the year in locations further from the equator. For example, in the Pre-Baby Boom Era the birth peak occurred as early as June in the northern states of Oregon and Maine, whereas the peak occurred as late as November in Florida. The variation in birth peak timing was largest during the Pre-Baby Boom Era, when the most out-of-phase states differed by more than 5 months.

During the Baby Boom Era, most states had birth peaks that occurred in August or later. The only peaks which occurred prior to August were in seven northern states and this pattern continued during the Modern Era. The earliest birth peaks always occurred in northern states, followed by mid-latitude states, and the latest peaks occurred in southern states. Across all eras, the latest peak was consistently in Florida, where the peak timing ranged from early October in the Modern Era, to early November in the Pre-Baby Boom Era.

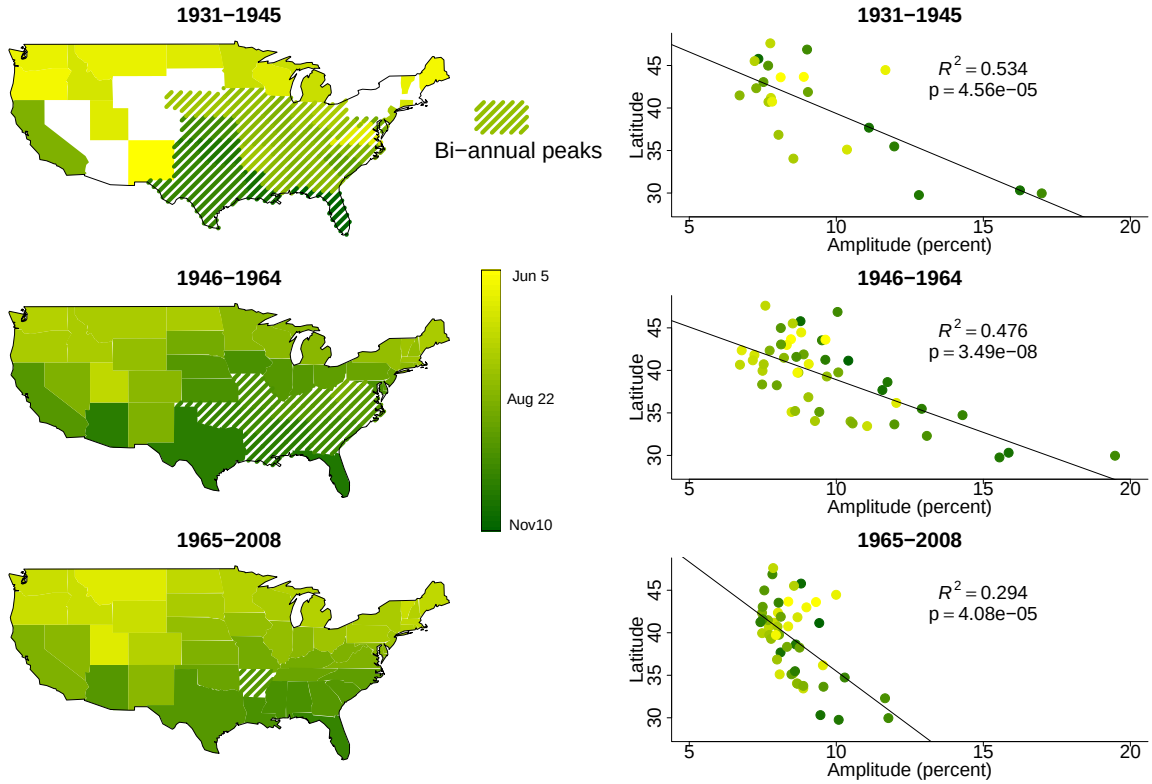


Figure 2.2: Spatiotemporal patterns of seasonal birth peak timing and amplitude in the US. (Top panels) Pre-Baby Boom (1931–1945), (Middle panels) Baby Boom (1946–1964), and (Bottom panels) Modern Era (1965–2008). Maps depict the latitudinal gradient in the timing of the birth peak. Colors indicate the mean timing of the birth peak for each state. Hatched regions represent states in which a significant bi-annual peaks were discernible and are color coded based on the timing of their primary annual birth pulse (also see Figs A.1 & A.2). States shown in white did not exhibit a significant periodicity. Regressions show the latitudinal variation in seasonal amplitude, with the colors representing the peak birth timing for the respective period.

The latitudinal gradient in peak birth timing seen in the US was reflective of a worldwide pattern. The worldwide timing also followed a latitudinal gradient with birth peaks occurring earlier at higher latitudes and later for countries closer to the equator (Fig 2.3). However, at any given latitude there was a large amount of variation in the timing of the birth peak. In the highest latitude countries ($> 50°$ N), birth peaks occurred between April and July. While there were two outlying mid-latitude countries with birth peaks in March and April (Italy, 1970–1985 and Tajikstan, 1989–1994), typical mid-latitude locations (20–50° N) had peaks between May and November. Countries in the vicinity of the equator (0–20° N) displayed the least amount of variation in timing. The equatorial countries, such as those in the

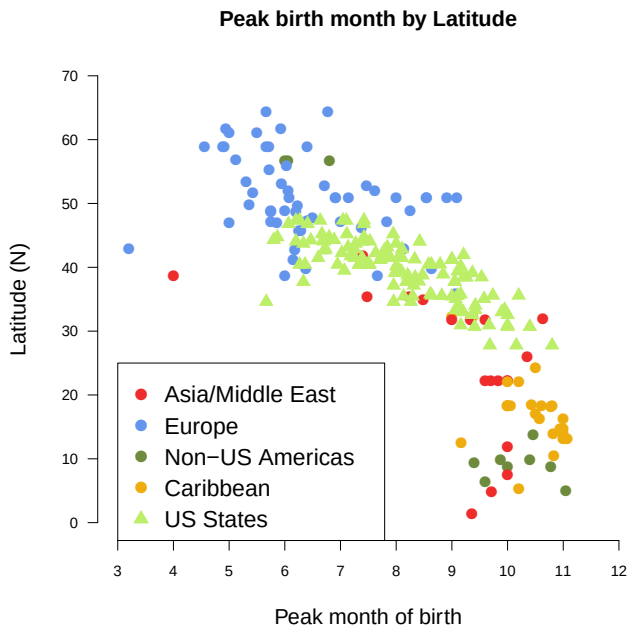


Figure 2.3: Northern Hemisphere patterns of seasonal birth pulses color coded by region. Birth pulses occurred earlier in the year at northern latitudes. Table A.5 provides the details for each country, including the time frame of the data which ranges from the 1960s to 2011.

Caribbean, consistently had birth peaks between September and November, with the latest birth peak occurring in Saint Vincent and the Grenadines during the period 1992–2005.

In addition to the annual birth peaks, in the US a significant bi-annual (6 month) birth pulse was detected in 24 states. In the Pre-Baby Boom Era, all states with bi-annual periodicity were clustered in the lower midwest, deep south, and southeast (Figs 2.2, A.1, & A.2). In the Baby Boom Era, only 13 states continued to exhibit a bi-annual period. Arkansas is the only state where this bi-annual birth pulse persists in the Modern Era.

2.7 The Amplitude of Seasonal Births

Birth amplitude was measured for each time series, each year, as the one-half peak-trough difference with noise removed. Amplitudes are represented as a percent of the mean annual birth rate (Appendix A: Materials & Methods). As with the seasonal timing, in the US the amplitude of birth seasonality displays a latitudinal gradient. Fig 2.2 depicts the negative relationship between birth amplitude and latitude. We found that 29–53% of the variation in

birth amplitude can be explained by latitude ($p < 4.6e\text{-}5$). However, the amount of variation in birth amplitude explained by latitude decreased through time, perhaps due to the decline in birth amplitude throughout southern states during the Modern Era (Figs 2.2 & A.4). We did not observe a latitudinal gradient in birth amplitude outside of the US (Fig A.5).

As expected, the mean combined amplitude across all states was found to be comparable to the national level amplitude reported in the literature [92] and was 9.0%, 9.8%, and 8.5% for the Pre-Baby Boom, Baby Boom, and Modern Eras, respectively. Interestingly, due to the geographical variation in birth peak timing, state-level births are out of phase. Thus, aggregated US birth data has a deceptively low amplitude that is not reflective of individual states. Birth amplitudes $> 15\%$ were observed in many southern states throughout the time series (Fig A.4).

Prior work has shown that the levels of inter-annual variation in births observed in the US can have a dynamical impact on disease incidence [69]. It would follow logically that variability of this same magnitude over a shorter time period may also be important. Thus, we sought to compare the magnitude of the seasonal variation in births with inter-annual variation. Inter-annual variation was measured for each state as the percent change in mean birth rate from one year to the next. We found that in almost every instance, seasonal variation exceeded inter-annual variation, with seasonal variation in the Modern Era being 2–3 times larger than the variation from year to year (Fig A.6).

2.8 The Effects of Birth Seasonality on Epidemic Dynamics

We investigated the impact of birth seasonality on epidemics of childhood disease by employing models of measles transmission. As shown in Fig 2.4A, birth seasonality can have the effect of amplifying or dampening incidence during epidemic years. Crucially, the impact of birth seasonality depends on the amplitude and phase relationship between susceptible

recruitment and transmission seasonality. In our simulation study we did not account for maternal antibodies, thus the peak in susceptible recruitment was equivalent to the birth peak. However, inclusion of maternal antibodies would translate into a lag between the peak in births and the peak in susceptible recruitment. We found that if the peak in susceptible recruitment occurs at the beginning of the year, when children are in school and measles transmission is elevated, the epidemic is amplified due to the availability of susceptibles. In contrast, if the peak in susceptible recruitment occurs at the end of the school year, when children are entering summer break, the epidemic is dampened (Fig 2.4A).

Independent of the timing of the birth peak, the effect of birth seasonality on measles epidemics depends on the birth amplitude (Figs 2.4B & A.7). The larger the birth amplitude the greater the change in measles incidence. Not only does the amplitude affect incidence, but birth rates with high amplitude fluctuations ($> 40\%$) can alter incidence to such an extent that they can drive dynamical transitions (Fig 2.4B).

Statistical inference on simulated data led to small biases in the estimate of R_0 for measles (Fig 2.4C). For the time series in which the birth peak occurred in mid-December, day 351, a time at which susceptible recruitment increases epidemic year incidence, omitting birth seasonality resulted in over-estimating R_0 in order to capture the elevated epidemic year incidence. In contrast, when the birth peak was set to either early June (day 162) or late October (day 295), times at which susceptible recruitment dampens epidemic year incidence and elevates skip-year incidence, we under-estimated R_0. However the bias in R_0 was small, approximately $\leq$0.4–1.3%.

We found that models with seasonal births effectively capture measles dynamics in New York City (Fig 2.5A). In contrast to our simulation study, however, when multiple unknown parameters were estimated simultaneously, the small predicted bias in R_0 was masked by uncertainty in parameters and Monte Carlo error (Fig 2.5). Hence, the maximum likelihood

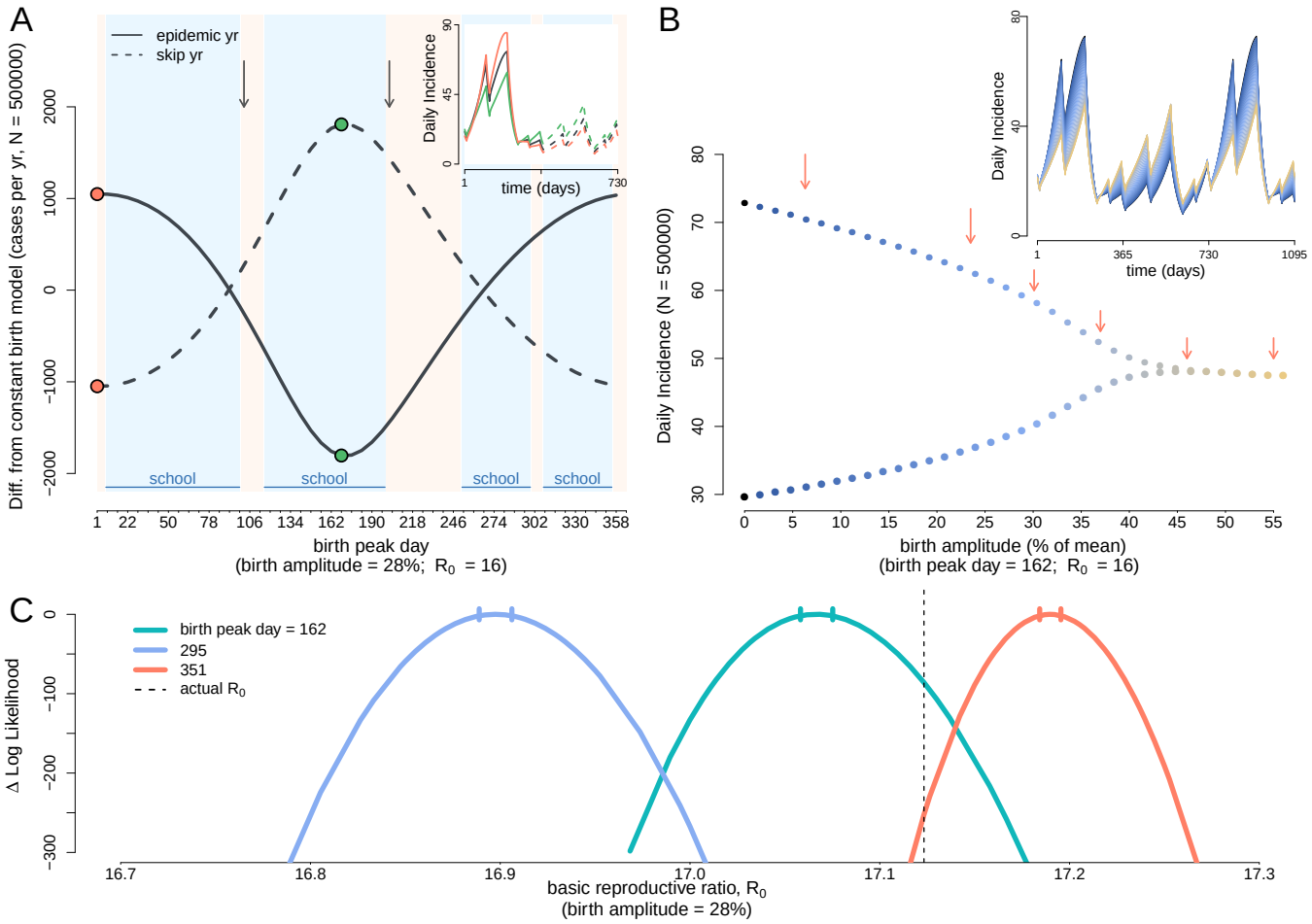


Figure 2.4: Impact of birth seasonality on childhood disease. (A) Epidemic and skip-year incidence varies with birth peak timing along the x-axis. Solid curve shows the change in epidemic year incidence when birth seasonality is added to the measles model. Dashed curve shows the change in the skip-year. The phase relationship between seasonal births and transmission determines whether birth seasonality has an effect on incidence. The greatest increase in epidemic year incidence is when the birth peak occurs after children return from winter holiday (orange points). A decrease in epidemic year incidence occurs when births peak prior to summer vacation (green points). School terms are noted and vertical arrows mark the timing of incidence peaks during the epidemic year. (Inset) Time series from the constant birth model (black), and time series corresponding to the color-matched points on the main graph. (B) Bifurcation diagram showing the change in epidemic and skip-year peak incidence with increasing birth amplitude. In the absence of birth seasonality, epidemics are biennial. As birth amplitude increases, skip-year incidence increases and epidemic year incidence decreases. When birth amplitude reaches ∼40% epidemics become annual. Time series in the inset correspond to the points in the main graph; blue time series are biennial, and golden are annual. Arrows denote the birth amplitude observed in Switzerland, Cuba, Egypt, Nigeria, Guinea, and Sierra Leone, left to right. Amplitudes for Nigeria, Guinea, and Sierra Leone are from [64]. (C) Bias in R_0 estimates due to the exclusion of birth seasonality in a SEIR model. Time series were generated using a SEIR model with a birth peak day of 162 (turquoise), 295 (blue), or 351 (orange). Each time series was fit to the SEIR model with a birth amplitude of 0%. The actual R_0 value is shown by the dashed line and the likelihood profiles show that maximum likelihood estimate of R_0 is either over- or under-estimated when birth seasonality is excluded from the model. 95% confidence intervals for MLE are indicated on profiles.

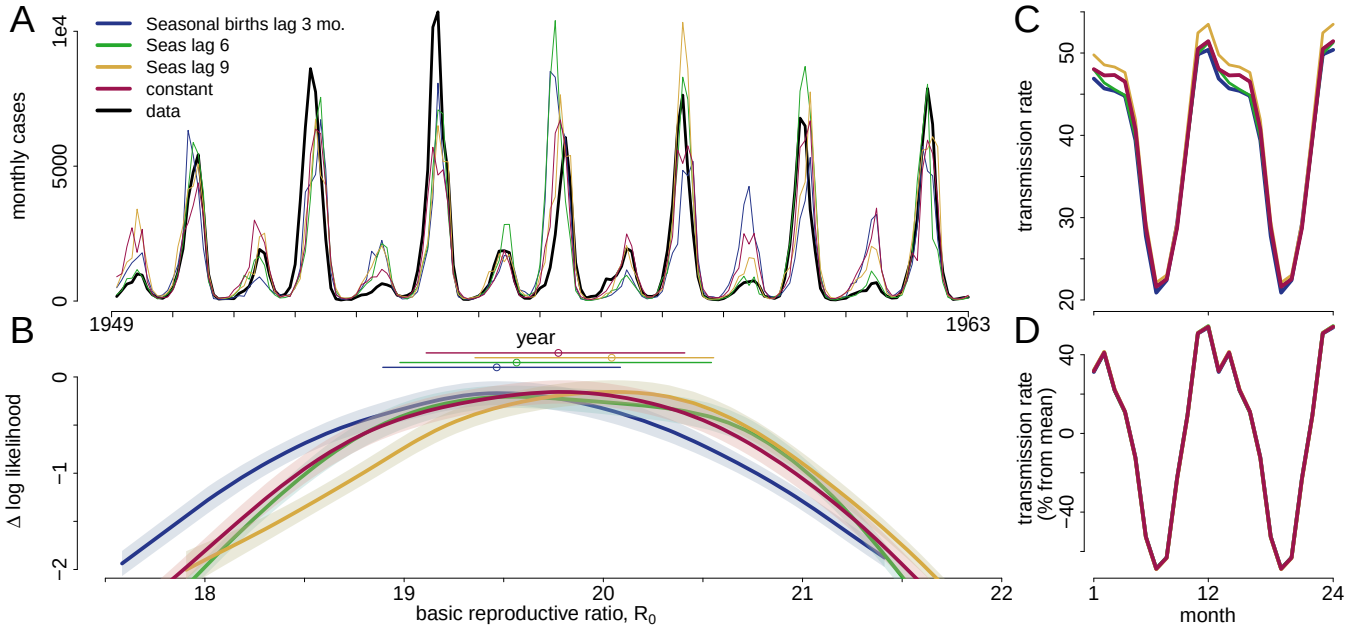


Figure 2.5: Measles cases in New York City. (A) Measles incidence (black) and a stochastic realization using the MLE for each type of birth covariate: seasonal births with a 3 month lag (blue), seasonal births with a 6 month lag (green), seasonal births with a 9 month lag (yellow), and births with no seasonality (maroon). Legend applies to all of Fig 2.5. (B) The shape of the likelihood surface with respect to R_0. The MLE R_0s are indicated by points and the standard errors are represented by horizontal lines. (C) MLE transmission splines for each model. (D) Transmission splines estimated using TSIR [20, 78] for each type of birth covariate. The MLEs differed with and without birth seasonality, but the differences in the point estimates were overwhelmed by uncertainty in parameter estimates (Figs 2,5B & 2.5C). No difference in transmission parameters was observed using the TSIR method.

parameter estimates (MLEs) for models with and without birth seasonality were nearly identical. The MLEs of the basic reproductive number, R_0, ranged from 19.3–20.3. Thus, the incorporation of birth seasonality into the model did not substantially change parameter estimates, and the dynamics of baby-boom era measles in New York City can be captured by the model without birth seasonality (Fig 2.5A).

2.9 Discussion

Seasonal fluctuations in human births are observed throughout the world. The timing of the birth peak displayed a marked latitudinal gradient throughout the Northern Hemisphere. The latitudinal gradient in peak birth timing was observed in the US for the entirety of our

data, and was reflective of a much broader geographic pattern. National level birth data from Asia, Europe, the Americas, and the Caribbean also exhibited this latitudinal gradient with birth peaks occurring months earlier at locations further from the equator.

Contemporary seasonal birth amplitudes are substantial with a range of 7–12% in the US and 6–35% in other Northern Hemispheric countries. Along with the latitudinal gradient in peak birth timing, in the US we also observed a latitudinal gradient in birth amplitude. States in the southern US have larger seasonal fluctuations in births than northern states. This negative relationship between latitude and amplitude was more pronounced in the Pre-Baby Boom and Baby Boom Eras, relative to the Modern Era. However, this pattern was not observed outside of the US, suggesting this may either be a localized phenomenon or strongly correlated with social, economic, and/or cultural factors in the US.

In addition to the striking geographical variation in timing and amplitude of the annual birth peak, these data displayed additional complexity with the occurrence of bi-annual peaks across the lower midwest, deep south, and southeastern US in the Pre-Baby Boom Era. This bi-annual pulse was lost over time, with only Arkansas exhibiting bi-annual periodicity in the Modern Era. Bi-annual fluctuations in births have been documented in previous studies [29], but our data suggests that bi-annual birth pulses in the US are a relic of the past, lost to societal changes [29,184], yet may still exist in other countries. Given the robustness of birth seasonality as a global phenomenon of contemporary human populations, it is surprising that mechanisms driving these patterns remain poorly understood. Demographers have implicated a host of social, environmental, and physiological factors that may interact to drive birth seasonality. While a consensus has yet to be reached, and mechanisms vary geographically, hypothesized drivers include income, culture, race, holidays, rainfall, cold winters, and seasonally variable sperm quality [22, 29, 118, 119, 124, 157, 168, 194]. Although we focused on characterizing the variation in birth seasonality, rather than the mechanisms underlying this variation, it is our hope that the latitudinal gradient in peak birth timing

and amplitude observed here will help elucidate the primary drivers of birth seasonality.

Despite our high resolution birth data for the Northern Hemisphere, Southern Hemispheric data proved difficult to obtain. Our analysis focused solely on the US and countries where birth data were readily available. Unfortunately this leaves out many South American and African countries where vaccine preventable childhood diseases are most prevalent. Southern Hemispheric birth data may help us understand the variation observed in the seasonality of childhood infections. For instance, historical work in Africa has shown that measles incidence peaks in April in Uganda, Kenya, and Tanzania, but earlier (November–January) in their southern neighbors Zambia, Zimbabwe, and Malawi [139]. Knowing the seasonal birth peak timing and amplitude in these locations may allow us to better understand this variation. We anticipate the latitudinal gradient in peak birth timing will also be found in the Southern Hemisphere.

The impacts of birth seasonality on epidemic dynamics were explored here in the context of childhood diseases. Our theoretical predictions indicate birth seasonality has the potential to influence the dynamics of fully immunizing infections of childhood – for which susceptible recruitment most heavily relies on births [69, 108]. We demonstrated that birth amplitude and the timing of the birth peak relative to peak transmission determine whether, and to what extent, birth seasonality affects disease incidence patterns. In our inference study using simulated data, we found that ignoring birth seasonality can bias parameter estimation. As a proof of concept study, we tested for these biases using New York City measles data from the pre-vaccine era. However, we did not detect any systematic biases. There may be a number of reasons for this finding. First, during the time span of these data, the seasonal birth amplitude was low in New York City. Second, the short infectious period of measles is known to lead to pronounced frequency-locking with forcing in transmission [11, 108, 165], which may swamp any dynamical impacts of weakly seasonal susceptible recruitment. Finally, the combination of process- and measurement-noise in the data, combined with uncertainty

in parameter estimates and Monte Carlo error may have made it impossible to detect the predicted estimation bias.

Our simulation studies demonstrated that high amplitude birth seasonality, currently observed in many African and Asian countries (Table A.5 and [15, 64]), can affect disease periodicity and epidemic magnitude. In these settings, our findings have the potential to explain some of the spatial and temporal variation observed in the periodicity of diseases such as measles, rotavirus, and polio; and present a promising avenue for future research. Indeed, a recent study of birth seasonality across developing countries found that the timing of the birth peak influences epidemic timing, and a high birth rate magnifies the effect of birth seasonality on measles epidemics [64]. Although our study—focused exclusively on measles epidemiology—suggests that high amplitude birth seasonality is required to alter disease incidence, we predict that lower birth amplitudes may have a dynamical effect when coupled with a higher mean birth rate or for childhood diseases with longer infectious periods that may exhibit less frequency-locking with seasonal transmission [165]. Ultimately, our experience with these systems indicate that the impact of seasonal births on epidemiology will likely be determined by multiple factors, including: the age-distribution of infections, age-specific pattern of contacts, differences in R_0, and the demographic context.

Dynamical consequences of birth seasonality aside, we emphasize that the spatial variation in birth seasonality documented here is pertinent when developing time-specific vaccination campaigns. For example, the World Health Organization implements time-specific vaccination campaigns to supplement routine immunization for the control of measles and polio in Africa, the Eastern Mediterranean, and South-East Asia. Clearly, if these infant immunization campaigns occur prior to the birth pulse, they will be inefficient. Thus, it is our hope that future studies aimed at mitigating childhood diseases will utilize birth seasonality to reduce the burden of disease and tackle some of the unanswered questions in disease ecology.

2.10 Acknowledgements

We thank two anonymous reviewers, the associate editor Dr Vincent Jansen, R. McDaniel, & the Rohani/King labs for their helpful comments on this work. MMB is supported by the NSF Graduate Research Fellowship Program. PR and AAK are supported by the Research and Policy in Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security, the Fogarty International Center, National Institutes of Health and by a research grant from the National Institutes of Health (R01AI101155).

CHAPTER III

Digital Epidemiology Reveals Global Childhood Disease Seasonality and the Effects of Immunization

3.1 Preface

This chapter of my dissertation was published in the *Proceedings of the National Academy of Sciences (PNAS)* in 2016 (DOI: 10.1073/pnas.1523941113). The authors include M. Martinez-Bakker, B. Helm, T. Stevenson, and myself. In March 2015 I was forwarded a set of figures showing the seasonality of Google searches for four childhood infectious diseases in the United Kingdom and Australia from co-author T. Stevenson (e.g. Fig 5.1). From my previous work on the TYCHO dataset [218], I quickly realized that the seasonal Google search patterns were similar to the seasonal incidence of chickenpox. Together with Micaela, we brainstormed on the formulation of a model, and acquired location-specific Google searches for the word 'chickenpox', for more than 30 countries across the globe. I performed the analyses and fitted the models with much cooperation from Micaela. Additional analyses and models, which were done for a time series analysis course taught by Ed Ionides (a committee member) in which I participated, are included in Appendix B. All *pomp* code for this chapter, as well as Chapters 2 & 4 are included in Appendix D.

3.2 Abstract

Public health surveillance systems are important for tracking disease dynamics. In recent years, social and real-time digital data sources have provided new means of studying disease transmission. Such affordable and accessible data have the potential to offer new insights into disease epidemiology at national and international scales. We used the extensive information repository Google Trends to examine the digital epidemiology of a common childhood disease, chicken pox, caused by varicella zoster virus (VZV), over an eleven-year period. We (1) report robust seasonal information seeking behavior for chicken pox using Google data from 36 countries, (2) validate Google data using clinical chicken pox cases, (3) demonstrate that Google data can be used to identify recurrent seasonal outbreaks and forecast their magnitude and seasonal timing, and (4) reveal that VZV immunization significantly dampened seasonal cycles in information seeking behavior. Our findings provide strong evidence that VZV transmission is seasonal and that seasonal peaks show remarkable latitudinal variation. We attribute the dampened seasonal cycles in chicken pox information seeking behavior to VZV vaccine-induced reduction of seasonal transmission. These data and the methodological approaches provide a novel way to track the global burden of childhood disease, and illustrate population-level effects of immunization. The global latitudinal patterns in outbreak seasonality could direct future studies of environmental and physiological drivers of disease transmission.

3.3 Introduction

Childhood infectious diseases continue to be a major global problem, and surveillance is needed to inform strategies for the prevention and mitigation of disease transmission. Our ability to characterize the global picture of childhood diseases is limited, as detailed epidemiological data are generally nonexistent or inaccessible across much of the world. Available data suggest that recurrent outbreaks of acute infectious diseases peak within a relatively consis-

tent, but disease-specific seasonal window, which differs geographically [2, 88, 127, 136, 218]. Geographic variation in disease transmission is poorly understood, suggesting substantial knowledge gains from methods that can expand global epidemiological surveillance. Seasonal variations in host-pathogen interactions are common in nature [134]. In humans, the immune system undergoes substantial seasonal changes in gene expression, which is inverted between European locations and Oceana [63]. The regulation of seasonal changes in both disease incidence and immune defense is known to interact with environmental factors such as annual changes in day length, humidity and ambient temperature [197]. Accordingly, quantification of global spatio-temporal patterns of disease incidence can help to disentangle environmental, demographic, and physiological drivers of infectious disease transmission. Furthermore, the recognition of the regional timing of outbreaks would establish the groundwork for anticipating clinical cases, and when applicable, initiating public health interventions.

Since childhood disease outbreaks are often explosive and short-lived [108], temporally rich (i.e., weekly or monthly) data are needed to understand their dynamics. Similarly, in order to establish the recurrent nature of outbreaks that occur at annual or multi-annual frequencies, long-term data are needed. Thus, ideal disease incidence data have both high temporal resolution and breadth (i.e., frequent observations over many years). Over the past decade, the internet has become a significant health resource for the general public and health professionals [33,96]. Internet query platforms, such as Google Trends, have provided powerful and accessible resources for identifying outbreaks and for implementing intervention strategies [34, 103, 175]. Research on infectious disease information seeking behaviour has demonstrated that internet queries can complement traditional surveillance by providing a rapid and efficient means of obtaining large epidemiological datasets [58, 59, 84, 175, 189]. For example, epidemiological information contained within Google Trends has been used in the study of rotavirus, norovirus, and influenza [58, 59, 103, 189]. These tools offer substantial promise for the global monitoring of diseases in countries that lack clinical surveillance but have sufficient internet coverage to allow for surveillance via digital epidemiology.

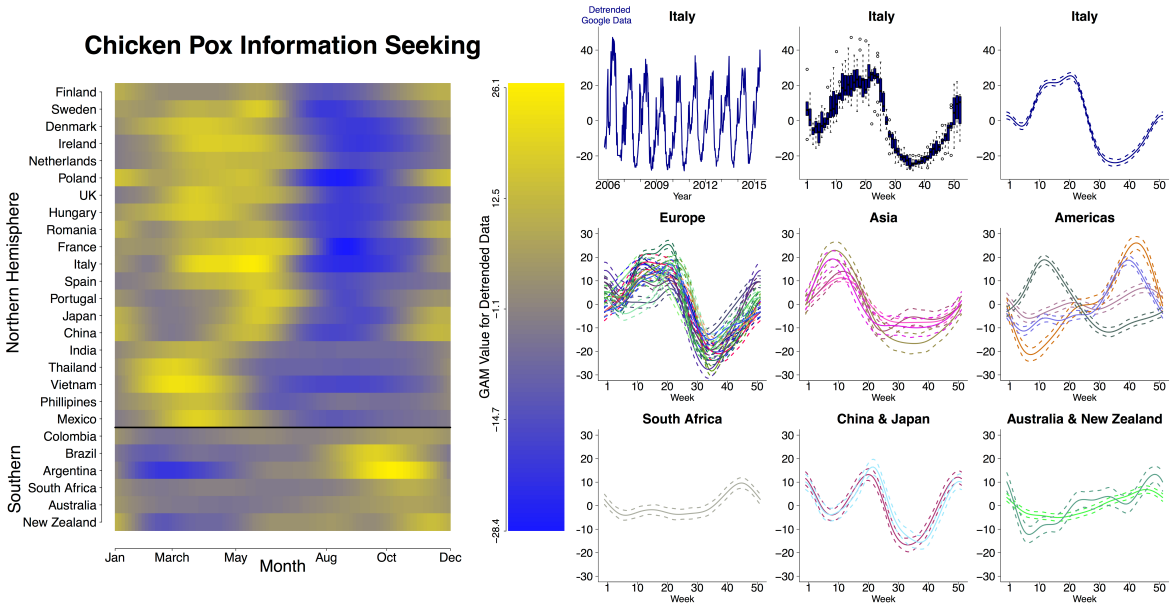


Figure 3.1: Left Figure: Global seasonality of chicken pox outbreaks measured using Google Trends as a proxy for chicken pox dynamics. Countries are organized by geographic region and latitude. Latitudinal variation in seasonal chicken pox information seeking behavior was observed for countries with wavelet confirmed significant seasonality. The seasonality was estimated by fitting a General Additive Model (GAM) to the detrended Google data from each country. GAM values using week number as the predictive variable for Google data are shown in the heatmap and correspond to the GAM curves to the right. **Right Figure:** Data processing and regional GAM values (Top row) Data processing steps. (Top row, left panel) Detrended Google data for Italy. (Top row, middle panel) Box-and-whisker plot of Google data for Italy: 1st-to-3rd quartiles in solid color with whiskers representing 95% confidence intervals. All other panels represent GAM values using week number as the predictive variable for Google data in each country. European countries include Finland, Sweden, Denmark, Ireland, Netherlands, Poland, UK, Hungary, France, Romania, Italy, Spain, and Portugal. Asian countries include Vietnam, India, Thailand, and the Philippines. Americas include Mexico (with peak in week 10), Colombia, Brazil, and Argentina.

Here we focused on one common childhood disease, chicken pox, as a study system because it would allow us to validate internet query data using clinical data from the small number of geographically distinct countries that report cases, and to address the impact of VZV vaccination on outbreaks. Chicken pox—a highly contagious disease caused by VZV—has low mortality but exceptionally high morbidity, with most unvaccinated children infected by age 15 in developed countries [185, 225]. The burden of VZV extends beyond chicken pox, because a VZV infection causes fluid filled blisters, which eventually burst, creating the opportunity for infection from various invasive bacterial pathogens (e.g., Group

A streptococcal infections) [221]. Chicken pox is not included in the World Health Organizations (WHO) global monitoring system for vaccine preventable diseases [227], meaning there are few countries which report clinical cases. In the United States (US), a country that immunizes against VZV; chicken pox was historically a notifiable disease. A lapse in national surveillance 1981–2001 compromised the ability of researchers to examine the long-term disease dynamics and the impact of immunization. [39, 41, 218]. Although the clear symptomatology of chicken pox makes the disease readily observable at the individual level, the lack of reporting makes VZV transmission dynamics cryptic at the population-level and obscures its spatio-temporal patterns.

The VZV vaccine is on the WHO list of essential medicines, which specifies the most important medicines needed for basic health systems [224], and is available as either stand-alone VZV or as the measles, mumps, rubella, and varicella vaccine (MMRV). However, the US, Germany, Canada, Uruguay, Australia and regions of Spain and Italy are among the few locations that have included VZV vaccination in their childhood immunization schedules for multiple years [70,113,150,155,191]. Short term surveillance studies in select locations of the US have demonstrated that moderate levels of vaccine coverage were able to dramatically reduce chicken pox incidence [12, 188], partially through the effect of herd immunity [48]. However, the effects of VZV vaccination on morbidity and mortality remain poorly understood [70] because global chicken pox report rates are low (e.g., the US rate is estimated to range from $< 0.1\%$ to 20% [38]). Therefore, the public health community is faced with a scarcity of chicken pox data to inform VZV vaccination policy. In certain locations (e.g., Madrid), VZV immunization has ceased, possibly due to the lack of information regarding the effects of immunization that can be used to assess health gains and economic feasibility. Clearly, there remains a lack of research, especially in countries that have recently introduced the VZV vaccine into the childhood immunization schedule [42, 70].

In this study, we took advantage of the extensive data available in Google Trends to study the global seasonal transmission of chicken pox. We (1) data mined chicken pox

information seeking behavior using language-specific Google queries of "chicken pox" from 36 countries spanning 5 continents over an eleven-year period (Table B.3), and characterized the seasonality of outbreaks. We then (2) validated Google Trends data using detailed time series of clinical cases from 5 countries within 4 continents. We (3) confirmed that statistical models have a profound ability to forecast outbreaks in unvaccinated and vaccinated populations. Finally, we (4) verified the impact of VZV vaccination on the seasonality of chicken pox outbreaks.

3.4 Seasonality of Chicken Pox Information Seeking & Validation

We detected significant seasonality of the Google Trends data [86] in 27 out of the 34 countries for which weekly data were available. Each geographic region in our study displayed distinct seasonal patterns of information seeking behaviour (Fig 3.1). A strong latitudinal pattern was clearly discernible (Fig 3.1), with peaks early in the year in the Northern Hemisphere and later in the Southern Hemisphere, corresponding to springtime outbreaks worldwide. These spring peaks agree with the seasonal timing found in historical datasets and previous studies of chicken pox [93,127,187,218,219] (Fig B.8). European countries were mostly unimodal, with a peak in March–May, but some had an additional, smaller, peak in late December. Several countries in the Southern hemisphere, including Australia, New Zealand, South Africa, and those located in South America, had a single discernible peak at various times of the year. China and Japan had bimodal peaks, which occurred March–May and December–January, balanced by deep summer troughs occurring July–August. Other Asian countries had a single peak that occurred February–March punctuated by a relatively shallow trough.

To validate Google Trends as a reasonable proxy for chicken pox dynamics, we compared clinical data to Google Trends data from five countries. We found that in the three countries lacking VZV vaccination—Mexico, Thailand, and Estonia—chicken pox information seeking was significantly correlated with reported cases of chicken pox, with $R^2 = 0.70$, 0.81, and

0.65 respectively (Fig 3.2). The correlation was reduced, but still significant, in Australia ($R^2 = 0.26$), which implemented nationwide immunization in 2005. In the US, which has actively vaccinated since 1995, the association between information seeking and reported chicken pox cases was low ($R^2 = 0.018$). To better understand the patterns in these countries, we investigated the context of the language-specific Google searches of "chicken pox". We compared the context of top searches containing "chicken pox" to determine whether they were searches for disease, symptoms, or treatment (i.e., indicative of disease in the household/community) or vaccination (i.e., not necessarily related to disease incidence).

The proportion of searches related to disease—rather than chicken pox vaccination or other search contexts—differed between focal countries. In Mexico and Thailand, which do not vaccinate, we classified the relative frequency of the top searches indicating disease to be 0.82 and 0.80 (Fig B.7). In both Australia and the US, where the VZV vaccine is required, vaccination and other search contexts had a higher relative frequency, while disease indicators had a frequency of 0.71 and 0.66, respectively. Interestingly, in the US, the second highest query involved chicken pox vaccination (Table B.2). Estonia had low search volume and the top searches were not available.

These data indicate the value of adding further layers to search terms. By doing so, we identified two patterns in information seeking: i) the data reflected seasonal dynamics in most countries (Mexico, Thailand, Estonia, and Australia); and ii) in the US, where vaccination has long been introduced, the data reflected a shift in the motivation of information seeking. These findings suggest that Google Trends reflect chicken pox dynamics more closely in countries that do not vaccinate versus countries that do vaccinate, and highlight the dynamic nature of information seeking. This likely explains why chicken pox information seeking in Mexico, Thailand, Estonia, and Australia, but not in the US, strongly reflected reported cases.

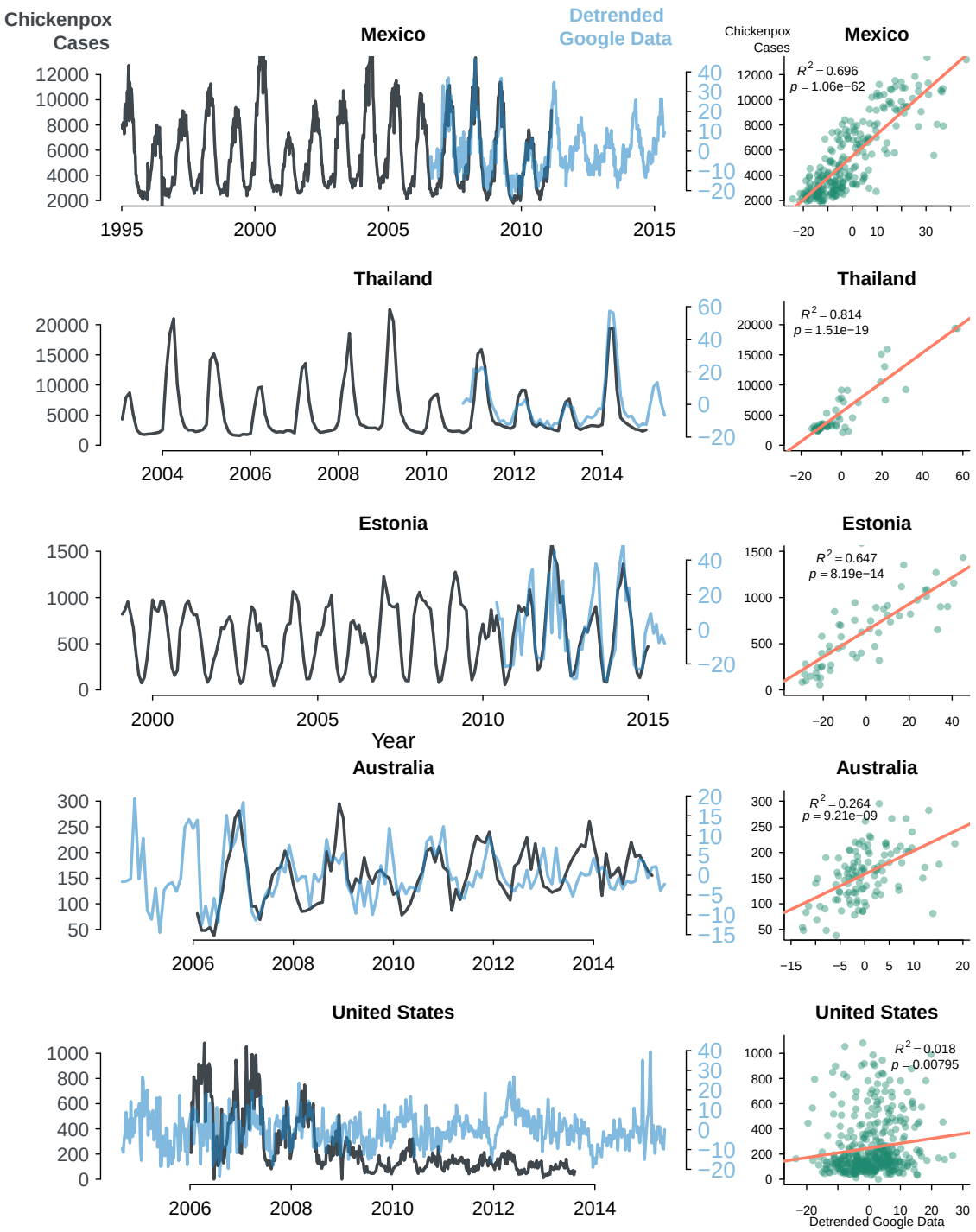


Figure 3.2: Relationship between chicken pox cases and information seeking. (Left column) Time series of reported chicken pox cases and information seeking behaviour for chicken pox (i.e., Google Trends data) in Mexico, the US, Thailand, Australia, and Estonia. Google data were detrended to remove long-term trends and focus on seasonal variation in information seeking. (Right column) Relationship between reported cases of chicken pox and chicken pox information seeking when both were available, with applicable R^2 and p values. Chicken pox case data from Mexico and the US were weekly, whereas chicken pox case data from Thailand, Australia, and Estonia were monthly.

3.5 Forecasting Outbreaks using Google Trends

To determine whether the information seeking behaviour observed in Google data, T, was able to forecast chicken pox outbreak magnitude and timing, we built and fitted eight different statistical models to forecast chicken pox case data. We evaluated the epidemiological information contained in Google Trends by comparing these Google Trends models with a seasonal null model that did not incorporate Google data (Table B.1, Figs B.1, B.2). Three models that included the Google data fit better than the null model, however we focused on the model with best fit, and refer to it as the Google model hereafter. The null model lacked information seeking in the force of infection parameter, which we defined as the monthly per capita rate at which children age 0–14 years are infected. In order to estimate the number of symptomatic VZV infections per month, we multiplied the force of infection with an estimate of the population aged 0–14 years (Eq B.3). Both the Google and null models were fitted to the case data from a VZV-unvaccinated population (Thailand), which showed robust seasonality, and a VZV-vaccinated population (Australia), which exhibited reduced seasonality. To estimate the number of symptomatic VZV infections each month, I_t, we used Google Trends data from the previous two months, T_{t-1} and T_{t-2}, where t is time in monthly time steps. The chicken pox process model tracked the force of infection, λ_t,

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right)\mathrm{T}_{t-1} + \beta_2|\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right]\epsilon_t. \tag{3.1}$$

The model also contained environmental stochasticity, ϵ_t, which was drawn from a gamma distribution with a mean of 1 and variance θ. We estimated 6 parameters for the model: the mean and the phase of the seasonality (β_1 and ω), a parameter scaling the Google Trends data (β_2), the baseline force of infection (β_3), the process noise dispersion parameter (θ), and the reporting dispersion parameter (τ) of a normal distribution, with a mean of 1, from which case reports were drawn. The parameters were estimated using maximum likelihood by iterated particle filtering (MIF) in the R-package pomp [111, 112].

The Google Trends model fit the case data (Fig B.1) and performed better than the null model (which estimated 5 parameters) for both Australia and Thailand; null model AICs were > 28 units above Google Trends model AICs in both locations. Since each model was seasonally forced, all models captured the seasonal timing of outbreaks. However, the Google Trends model was able to predict the interannual variation in outbreak size (Fig 3.3), while the null model could not (Appendix B, Fig B.1).

3.6 The Signature of Varicella Zoster Immunization

We investigated the signature of VZV immunization by examining Google Trends data in countries that actively immunize and those that do not. Seasonality of information seeking behaviour was much stronger in countries lacking active immunization programs than in countries that include the VZV vaccine as part of the childhood immunization schedule (Fig 3.3). Germany, which made the VZV vaccine mandatory in July 2004 [113, 158], had weakening seasonality in information seeking until 2009, when a second VZV booster dose was added to the immunization schedule, drastically reducing information seeking seasonality (Figs B.5, B.6). In Australia, where the VZV vaccine was publicly funded in November 2005 [37], the amplitude of information seeking was severely dampened by the end of 2007. In the US, immunization began in 1995 [188] while Canada required the vaccination starting in 2000 [200]. In these two countries, where the VZV vaccine introduction predated Google Trends data, little seasonality was observed in the Google Trends data (Figs 3.3, B.9). In Spain and Italy, where only a few regions or municipalities require VZV immunization [70], minimal change was observed in the Google Trends dynamics following immunization implementation. However, in Spain there was a reduction in search amplitude when immunization efforts were at their maximum, likely indicating a reduction in searches following immunization, similar to Germany (Fig B.6).

Since information seeking behavior strongly correlates with seasonal outbreaks of chicken pox (Fig 3.2), the loss of information seeking seasonality in countries that immunize can

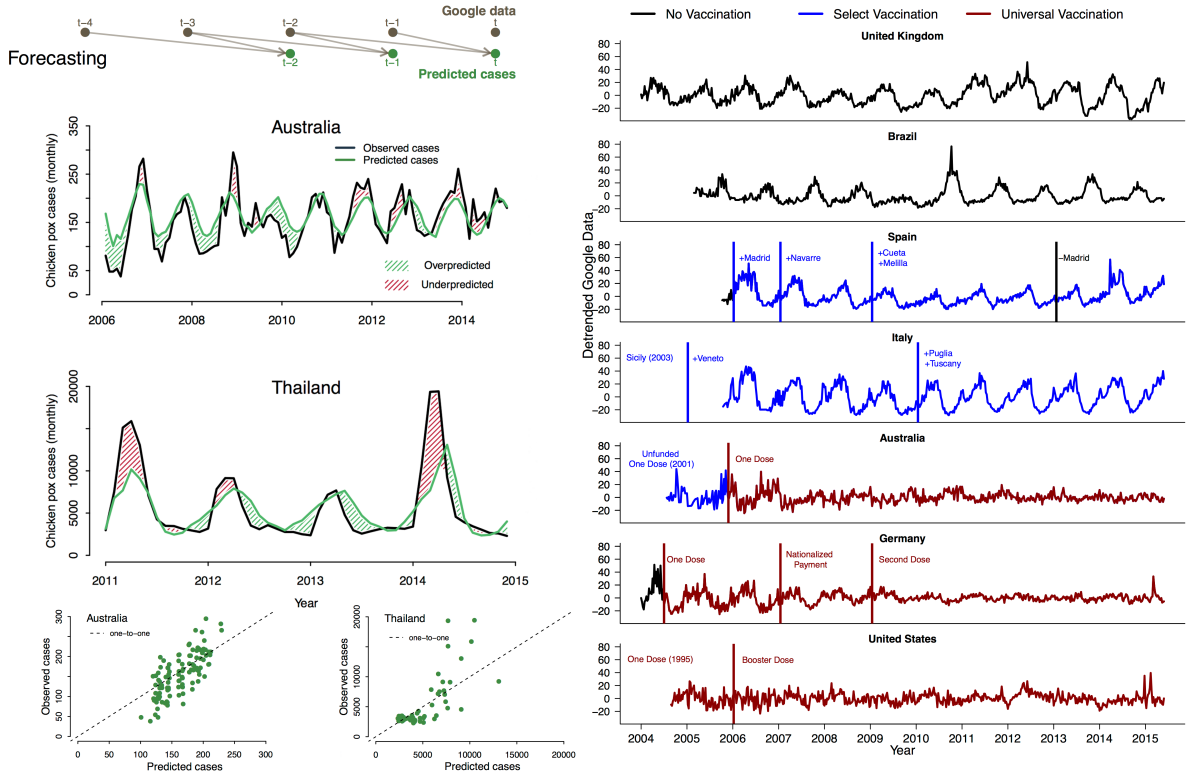


Figure 3.3: Left figure Forecasting chicken pox cases using Google Trends. (Top) Forecasting model schematic, Google Trends data from months $t-2$ and $t-1$ are used to predict chicken pox cases in month t. (Bottom left) Observed and predicted chicken pox cases in Australia (active immunization) and Thailand (no immunization) from 10000 simulations of the fitted models parameterized with the maximum likelihood estimates; overpredicted (green hash marks) and underpredicted (red hash marks) regions are indicated. (Bottom right) Model predicted cases versus observed chicken pox cases along the dotted 1-to-1 line. **Right figure** Detrended chicken pox information seeking in relation to immunization. Data are weekly; x-axes indicate time, and y-axes are the detrended Google data (same scale for all panels). Countries with universal (national) immunization in red, countries with select (regional or municipal) immunization in blue, and countries lacking any mandatory immunization in black. Panels 1-2: the UK and Brazil, two countries with no immunization. Panels 3-4: Spain and Italy, two countries with no universal (national) immunization, but with select regional or municipal immunization. Vertical lines identify the implementation (blue for select, red for national) or termination (black) of immunization efforts. Cities and regions on these panels indicate where these efforts were focused. Panels 5-6: Australia and Germany, two countries that implemented national immunization since 2004. Australia had the vaccine since 2001, but nationwide immunization was not funded by the government until November 2005. Germany required a single dose for every child in July 2004, provided nationalized payment in 2007, and required a second dose in 2009. Panel 7: the US, which has had national immunization since 1995, required a booster dose in 2006.

signal the loss of recurrent seasonal chicken pox outbreaks, indicating outbreak mitigation driven by VZV immunization. We suggest that if (i) disease transmission is seasonal—as it is for chicken pox and other childhood diseases— and (ii) vaccination reduces disease transmission, then the impact of immunization can be measured as the reduction in seasonal outbreak amplitude. This is because vaccination will strongly diminish the transmission rate during the high-transmission season. In the case of chicken pox, the reduction of seasonality in information seeking is likely due to diminished outbreak seasonality seen in clinical data (Figs 3.2, B.8) and the subsequent shift of information seeking from disease queries to vaccination queries.

3.7 Discussion

In this study, we used digital epidemiology of chicken pox to (1) reveal previously unreported seasonal outbreaks on a global scale, which displayed robust latitudinal dependence, (2) confirm the reliability of the Google data against known clinical cases, (3) forecast the size of annual outbreaks, and (4) uncover the population-level effects of routine VZV immunization. The lack of contemporary reporting, due to the benign nature of infections and the increased use of immunization, have made it difficult to decipher modern VZV global epidemiology. Here, we established that information seeking behaviour can be applied to reveal the underlying epidemiology of a childhood disease, chicken pox.

Our analyses reveal profound, global patterns of seasonality in chicken pox transmission dynamics. These seasonal patterns are spatially structured: we have demonstrated a latitudinal pattern in the timing of outbreaks, with inverted phases between the southern and northern hemisphere and an apex in the spring. Evidence of the underlying biological basis for seasonality in chicken pox transmission remains an open question. There is a significant latitudinal shift (i.e., near 6 months) in chicken pox outbreak timing from the northern to southern hemisphere, which suggests an influence of environmental, biological, and/or behavioral drivers that vary with latitude, such as seasonal immunity, environmental factors,

and/or school terms. Interestingly, Google Trends also revealed seasonal variation in croup, fifth disease, and hand, foot, and mouth disease; each of these childhood diseases exhibited a unique annual peak with little overlap in their seasonal window of outbreak occurrence. The lack of synchrony among childhood diseases likely indicates that school terms and holidays are not the primary drivers determining outbreak timing (Fig B.3). We speculate that seasonal information seeking behaviour linked to childhood illnesses reflects biologically based seasonality of host-pathogen interactions [63, 134]. Our present data open new possibilities for extensive global analyses, which could disentangle contributions of different seasonal drivers to a broad range of infectious diseases. There is a pressing need for such knowledge as global seasonality is becoming rapidly modified and disrupted through human action, with potentially far-reaching implications for infectious disease transmission [196].

By taking advantage of freely available, real-time, internet search query data, we were able to validate information seeking behaviour as an appropriate proxy for otherwise cryptic chicken pox outbreaks and use those data to forecast outbreaks one month in advance. Our modeling approach, which incorporated Google Trends and the knowledge of spring peaks, was able to better forecast outbreaks than models that ignored Google Trends. While this was particularly clear for Thailand, which does not immunize against VZV, it also held for Australia, a country that vaccinates. These results suggest that information seeking can be used for rapid forecasting, when the reporting of clinical cases is unavailable or too slow.

Comparisons of Google Trends data with the reported cases in countries that lacked VZV immunization revealed a significant positive relationship (70%, 81%, and 65% of variation in Google Trends explained by variation in reported cases). However, the relationship significantly decreased in countries that included VZV vaccination in their childhood immunization schedule and displayed either no seasonality or low-amplitude seasonal cycles (e.g., the US, 1.8% and Australia, 26%). Interestingly, in Italy and Spain, where the VZV vaccine was only required in specific regions or municipalities of the country, no change in seasonal information seeking behaviour was detected in the face of vaccination, implying that widespread

immunization is necessary to mitigate seasonal cycles of disease and information seeking. These findings, particularly from the highly vaccinated countries in our data—the US and Australia—indicate that immunization programs diminish seasonal information seeking behaviour and likely represent decreased seasonality of outbreaks.

Studies of disease transmission at the global level, and the success of interventions, are limited by data availability. Disease surveillance is a major obstacle in the global effort to improve public health, and is made difficult by under–reporting, language barriers, the logistics of data acquisition, and the time required for data curation. We demonstrated that seasonal variation in information seeking reflected disease dynamics, and as such, we were able to reveal global patterns of outbreaks and their mitigation via immunization efforts. Thus, digital epidemiology is an easily accessible tool that can be used to complement traditional disease surveillance, and in certain instances, may be the only readily available data source for studying seasonal transmission of non-notifiable diseases. We focused on chicken pox and its dynamics to demonstrate the strength of digital epidemiology for studying childhood diseases at the population level, because VZV is endemic worldwide and the global landscape of VZV vaccination is rapidly changing. Unfortunately, there is still a geographic imbalance of data sources: the vast majority of digital epidemiology data are derived from temperate regions with high internet coverage. However, because many childhood diseases remain non-notifiable throughout the developing world, digital epidemiology provides a valuable approach for identifying recurrent outbreaks when clinical data are lacking. It remains an open challenge to extend the reach of digital epidemiology to study other benign and malignant diseases with under-reported outbreaks and to identify spatio-temporal patterns, where knowledge about the drivers of disease dynamics are most urgently needed.

3.8 Materials and Methods

Google Trends data for 'chicken pox' were downloaded, and examined for seasonality. These data were then compared against reported cases of chicken pox in countries where

case reports were available. We then constructed and tested multiple statistical models to determine whether Google Trends data could forecast chicken pox seasonality. Finally, we examined the effect of national immunization campaigns on the seasonality and amplitude of Google searches. Further methodological descriptions are included in the supporting information. This study was done with freely available, de-indentified, pre-existing data, thus no consent was required.

3.9 Acknowledgements

We would like to thank Fernando Gonzalez-Dominguez and Gilberto Vaughan for providing the chicken pox case reports from Mexico, and the Estonia Health Board, Department of Communicable Disease Surveillance and Control, for Estonian chicken pox case reports. KB would like to thank Mercedes Pascual, the members of her laboratory, and Marisa Eisenberg for helpful comments. Jesus Cantu (research assistant, Princeton University) translated and categorized chicken pox searches from Mexico, Thailand, Australia, and the US.

CHAPTER IV

The Underpinnings of Herpesvirus Dynamics: Transmission & Reactivation of Varicella Zoster Virus

4.1 Preface

This work is co-authored by M. Pascual, M. Eisenberg, and M. Martinez. This chapter also materialized from the information seeking figure I received (Fig 5.1), noting that shingles displayed a seasonal peak in incidence. Since shingles is non-transmissible, and all previous studies have demonstrated that it lacks seasonal reactivation, we decided to investigate the spatial complexities of shingles reactivation and the potential processes underlying these dynamics. My co-author, M. Martinez and I conceived the study and formulated the models. I was responsible for all data analyses, modeling efforts, and wrote the first few drafts of the manuscript. The other co-authors provided modeling input, including specifics on how to incorporate and test antibody boosting, and helped with later drafts of the manuscript. All *pomp* code for this chapter, as well as Chapters 2 & 3 are included in Appendix D.

4.2 Abstract

Herpesviruses are among the most widespread pathogens in humans; nearly every person will have been infected with at least one herpesvirus in their lifetime. Herpesviruses, though they vary in many aspects of their biology, have a commonality in that they cause

recurrent disease due to viral latency in the host and subsequent reactivation. Of the nine herpesviruses that infect humans, varicella zoster virus (VZV), which causes chickenpox and reactivates as shingles, has the richest legacy of research into its transmission dynamics, primarily because chickenpox is a notifiable, vaccine-preventable childhood disease. Data availability, as well as pressing policy questions surrounding chickenpox vaccination, make it a favorable study system for understanding the integrated transmission and reactivation dynamics of herpesvirus more generally. In order to reveal the underpinnings of VZV's natural dynamics in the absence of intervention, we utilized data from Thailand, which does not vaccinate for chickenpox or shingles. We developed generalizable mechanistic mathematical models of herpesvirus transmission and reactivation to examine hypotheses on: (1) drivers of transmission and reactivation of VZV, and (2) immune-modulated interactions between chickenpox transmission and shingles reactivation. In particular, our models were used to elucidate the cause of the observed seasonal patterns in both chickenpox and shingles, along with the association between ultraviolet (UV) irradiation and shingles. The fitted models suggest independent seasonal drivers of chickenpox transmission and shingles reactivation, with the level of UV exposure being a candidate mechanistic driver of elevated risk of shingles reactivation. Moreover, model simulations suggested that Thailand could have avoided 500,000 chickenpox cases during our 9-year study period with the inclusion of the chickenpox vaccine in their childhood immunization schedule. The best-fit model also suggested the introduction of chickenpox vaccination would not result in an increase in shingles incidence, although the results are preliminary, and further work to answer this question is warranted. With contention surrounding VZV vaccination worldwide, along with vaccines against four other members of the herpesvirus family in clinical trials, a better understanding of the biology driving herpesvirus dynamics is crucial for vaccination and policy, especially the unknowns surrounding reactivation and virus population ecology.

4.3 Introduction

Viral infections in humans are typically identified when they become symptomatic; however, most viruses cause minimal to no clinical symptoms. The herpesvirus family, which infect nearly all humans at some point in their life, is unique in that they are able to cause recurring disease due to cycles of lytic (transmission) and latent (reactivation) life stages Nine herpesviruses infect humans, including herpes simplex 1 & 2 (HSV-1 & HSV-2), which cause oral and genital herpes, Epstein-Barr virus (EBV), which may cause mononucleosis, cytomegalovirus (CMV), which may result in congenital CMV, and varicella zoster (VZV) — which can present as both chickenpox and shingles. Initial infection with herpesviruses requires close-contact. Typically, infection occurs via saliva/respiratory or sexual transmission modes. Subsequent reactivation of herpesviruses is highly complex at the molecular level, and the conditions favoring reactivation remain relatively unknown. In animal models, trauma and stress (e.g., elevated body temperature) can induce reactivation [178]. Human HSV cold sores on the other hand, have been shown to correlate with fatigue and ultraviolet (UV) irradiation [229, 233]; and reactivation of EBV has occurred *in vitro* after B-cell stimulation, suggesting that a response to another infection may trigger EBV reactivation [144]. While VZV transmission, in the form of chickenpox, has been well-studied, immunity from VZV is not well understood, and the mechanisms causing shingles reactivation remain unknown.

Varicella zoster virus is transmitted when it presents as chickenpox, but later reactivates from latency as shingles. By the age of 15 years, 99% of individuals across the world have antibodies to VZV [185, 225], but only approximately 30% of individuals will ever experience a symptomatic reactivation of VZV expressed as shingles [61]. In general, chickenpox is the only notifiable herpesvirus disease. Since chickenpox dynamics are known to be impacted by transmission during school-terms [104, 127], VZV is the ideal study system for understanding herpesvirus reactivation using population-level approaches. This is because (1) data are available for study and (2) since the transmission process is well understood, it reduces the dimensionality of the unknowns to be focused on the reactivation process. Chickenpox

is characterized by explosive springtime outbreaks [9], while shingles has been noted to lack seasonal patterns in incidence [24, 40, 46, 79, 97, 156, 182]. VZV reactivation has been hypothesized to be modulated by UV irradiation, which has been shown to reactivate HSV [233] by suppressing immunity [57, 81, 185, 238], but to date there have been no definitive studies linking UV as a mechanism driving shingles dynamics.

Despite the VZV vaccine being first approved in the US in 1995, vaccination policy for VZV is a topic of much debate. It is a live-attenuated vaccine administered in two doses during childhood, while a booster dose (i.e., the shingles vaccine) later in life suppresses potential VZV reactivation in adults [185]. Because it is attenuated, those immunized as children have a greatly reduced potential for later reactivation. Childhood VZV vaccination is only included in the vaccine schedule in a limited number of countries. The vaccine is contentious because childhood immunization against VZV has been hypothesized to prevent VZV circulation, and adult exposure could result in natural boosting of immunity, reducing the risk of VZV reactivation later-in-life as shingles [26,97,145,146,207]. Complications from chickenpox are rare, with less than 1% of infected individuals ever experiencing complications [221], though with an estimated 4 million cases prevented by vaccination each year in the United States [41] this is not a trivial number.

The hypothetical reduction of VZV antibodies in adults due to a decrease in antibody boosting, is the primary reason VZV vaccination has not been implemented worldwide. We refer to antibody boosting here as a proxy for cellular immunity, which is primary immune mechanism that protects from shingles reactivation [146]. Theoretical models examining chickenpox and shingles interactions that utilize antibody boosting have all predicted a sharp increase in shingles with the inclusion of the chickenpox vaccine [25, 26, 90, 107, 152, 181, 217]. Empirical evidence from surveillance programs in locations that immunize are less conclusive, observing both an increase [85, 123, 235] and no change [106, 131] in shingles incidence. Other countries, such as Canada [24, 172], the United Kingdom [24], and pre-vaccine United States [156] have previously experienced an increase in shingles cases without

vaccination. In Spain, which vaccinates in some regions (e.g. Madrid [9]), shingles incidence has been steadily increasing, partially due to demographic changes via a sharp drop in birth rates [135]. With the availability of the VZV vaccine, and EBV [72], CMV [239], and HSV [49] vaccines currently in clinical trials, a better understanding of the biology driving herpesvirus dynamics, specifically reactivation, is needed.

In this study, we took advantage of population-level data on chickenpox and shingles to examine herpesvirus transmission and reactivation, using the following procedures:

I) We obtained spatiotemporal chickenpox and shingles monthly clinical case reports, and annual age-specific incidence from Thailand, spanning 2003-2011 to characterize the population dynamics at the national and regional scale.

II) Since herpesviruses have been identified to reactivate from latency by UV irradiation, we acquired national and regional UV data for Thailand, to examine the relationship between UV and shingles incidence.

III) We built eight mechanistic models to test various drivers of herpesvirus transmission and reactivation.

IV) Finally, we simulated the best-fit model under various conditions of immunization in Thailand which suggested that more than 500,000 chickenpox cases could have been avoided during this time period.

4.4 Data

Here we took advantage of clinical case notifications of chickenpox and shingles made publicly available by the Ministry of Health in Thailand [202] to examine herpesvirus transmission and reactivation. We obtained regionally-resolved monthly chickenpox and shingles case reports, and annual age-specific incidence from 2003-2011. To test for potential effects of UV on shingles, we also collated complementary UV covariate data from the National Center for Atmospheric Research (NCAR) [1]. Case data and covariates were coupled with mechanistic transmission-reactivation models to test hypotheses regarding seasonality of VZV

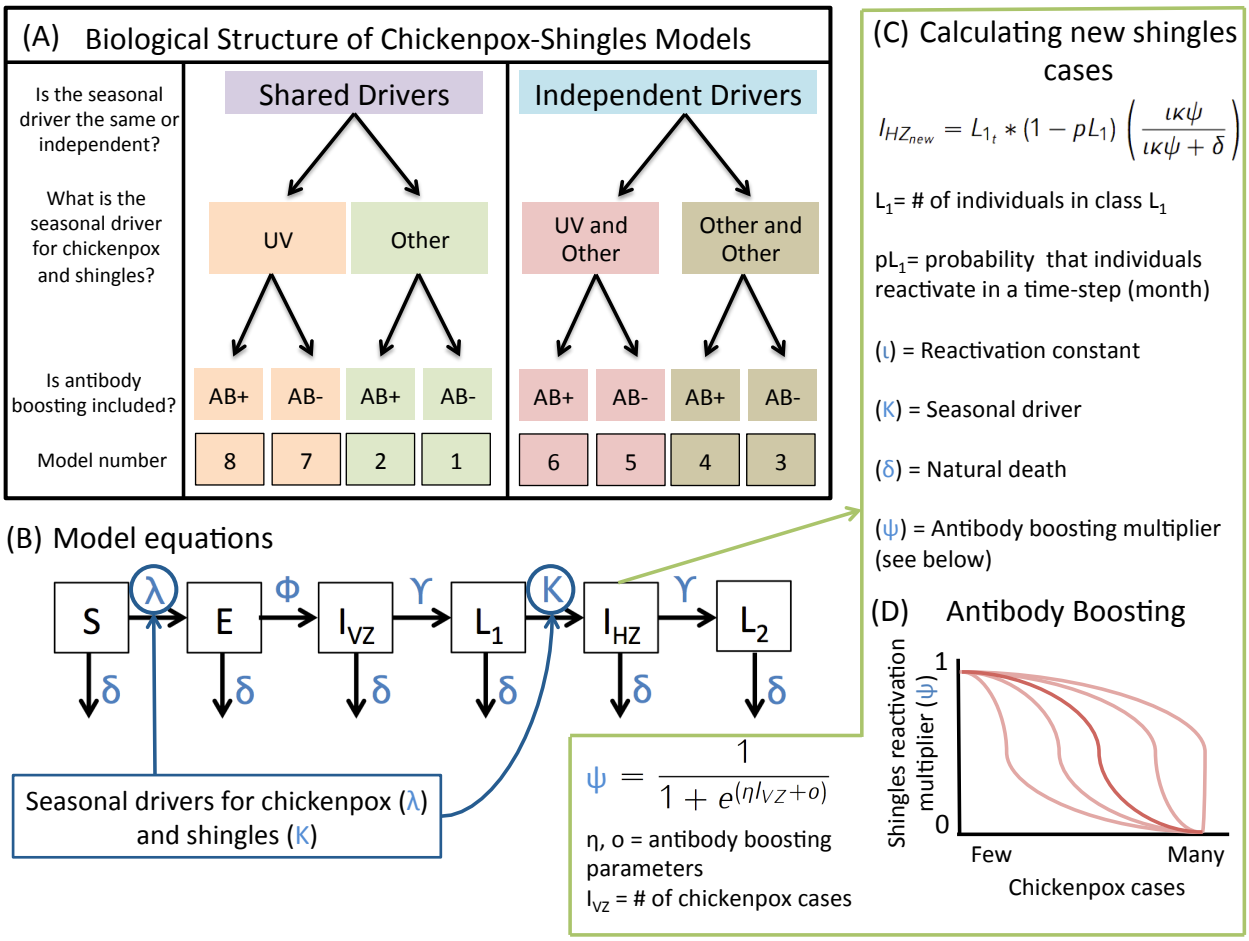


Figure 4.1: (A) Biological structure of the eight model variants fit to the chickenpox and shingles data from Thailand. AB is antibody boosting, where it is either present (AB+) or absent (AB-) in the model. (B) Model schematic, model variants differed in the λ and κ parameters which can be seen in Eqs C.16-C.19. State transitions and the force of infection are described in Eqs C.1-C.15. (C) The equation for new shingles infections ($I_{HZ_{new}}$), Eqn C.15, is described in detail. ϕ is the antibody boosting multiplier, γ is a fixed death rate parameter, and ι is a fixed rate of reactivation. (D) Functional form of antibody boosting tested in half of the model variants. Antibody boosting assumes high chickenpox incidence boosts VZV antibodies in adults through exposure, examples curves are shown in red. The equation for antibody boosting, Eqn C.20, is shown here.

transmission and reactivation, as well as immunological interactions between chickenpox and shingles. All analyses were done in R.

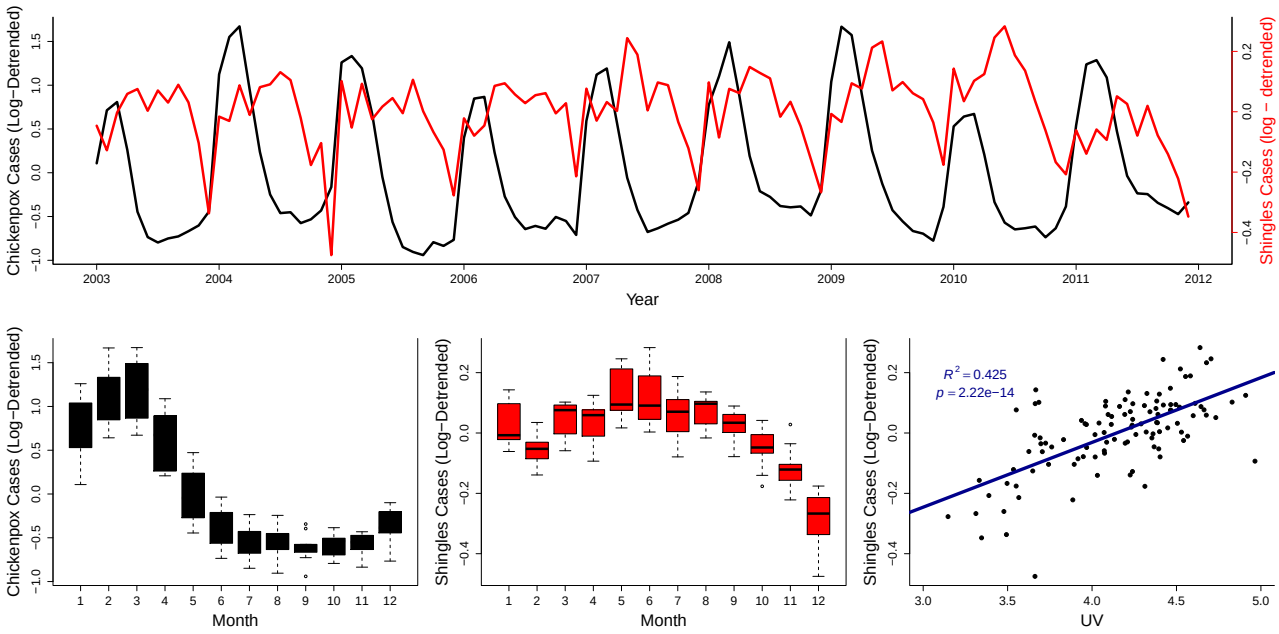


Figure 4.2: Chickenpox and shingles cases in Thailand with complementary UV covariates. (A) Monthly log-detrended cases of chickenpox (black) and shingles (red). (B) Boxplot of log-detrended chickenpox cases in Thailand. (C) Boxplot of log-detrended shingles cases in Thailand. (D) The correlation between (log-detrended) shingles cases and UV, monthly.

4.5 Models

Since the mechanisms driving the transmission and reactivation of herpesviruses are not well understood, multiple biological hypotheses were considered to interpret VZV dynamics at the population level (Fig 4.1A). A modular compartmental model, a redeveloped classical S-I-R [109], was utilized for its simplicity, appropriateness, and potential to be easily transformed and applied to other herpesvirus systems. The model included seasonal variation in the transmission rate (i.e., transmission resulting in chickenpox) and the reactivation rate (i.e., reactivation resulting in shingles). We used model variants that allowed us to estimate the shape of seasonal variation in transmission and reactivation such that the model could capture (i) effects of elevated transmission when children were in school, (ii) variation in reactivation due to hypothesized climate effects, such as UV radiation, or (iii) other unidentified seasonal drivers (Fig 4.1). Model variants include those that assumed chickenpox transmission and shingles reactivation have (1) identical or (2) independent seasonal variation

in rate. We also tested models that explicitly included UV irradiation data as a covariate driving seasonality. Importantly, all model variants were formulated in combination with the inclusion of antibody boosting to test whether the data support immunological interactions between VZV transmission and reactivation (Fig 4.1). Lastly, we simulated our best-fit model under three hypothetical conditions of VZV vaccination. We incorporated uptake data from the measles and hepatitis vaccine rollouts for Thailand [228], and a hypothetical ideal vaccine rollout where 99% coverage would be achieved at the end of the first year and maintained. In these simulations, we only immunized newborns, as we assumed there would be no national immunization days, or catch-up campaigns, for VZV vaccinations. For this simulation, we accessed chickenpox data going back to 1995 (shingles data were unavailable), when the VZV vaccine was first licensed in the US. We then estimated the number of chickenpox cases that would have been averted had VZV immunization been realized. All models were implemented in the R package *pomp* and fit using Maximum Likelihood by Iterated Particle Filtering (mif) [111]. Each model variation searched a minimum of 500,000 unique parameter combinations. Additional modeling details, including all equations, can be seen in the Appendix C & D information.

4.6 VZV Dynamics

As anticipated, strong seasonal cycles were detected in chickenpox dynamics. Surprisingly, shingles cases also displayed seasonal variation of incidence (Fig 4.2A,C). Chickenpox dynamics were characterized by explosive outbreaks which began at the end of the year, where the epidemic take-off was Nov–Dec, culminating with seasonal peaks from Feb–Mar, followed by deep troughs that lasted Jun–Oct. The seasonal peak in shingles cases occurred May–Jun, with an annual trough in December. Interestingly, January consistently had a relatively high number of shingles cases, but the reactivation take-off was then stunted by a three-month trough from Feb–Apr, before an eventual two month long annual reactivation peak from May–Jun (Fig 4.2A,C). This peak was followed by a slow decline in cases through

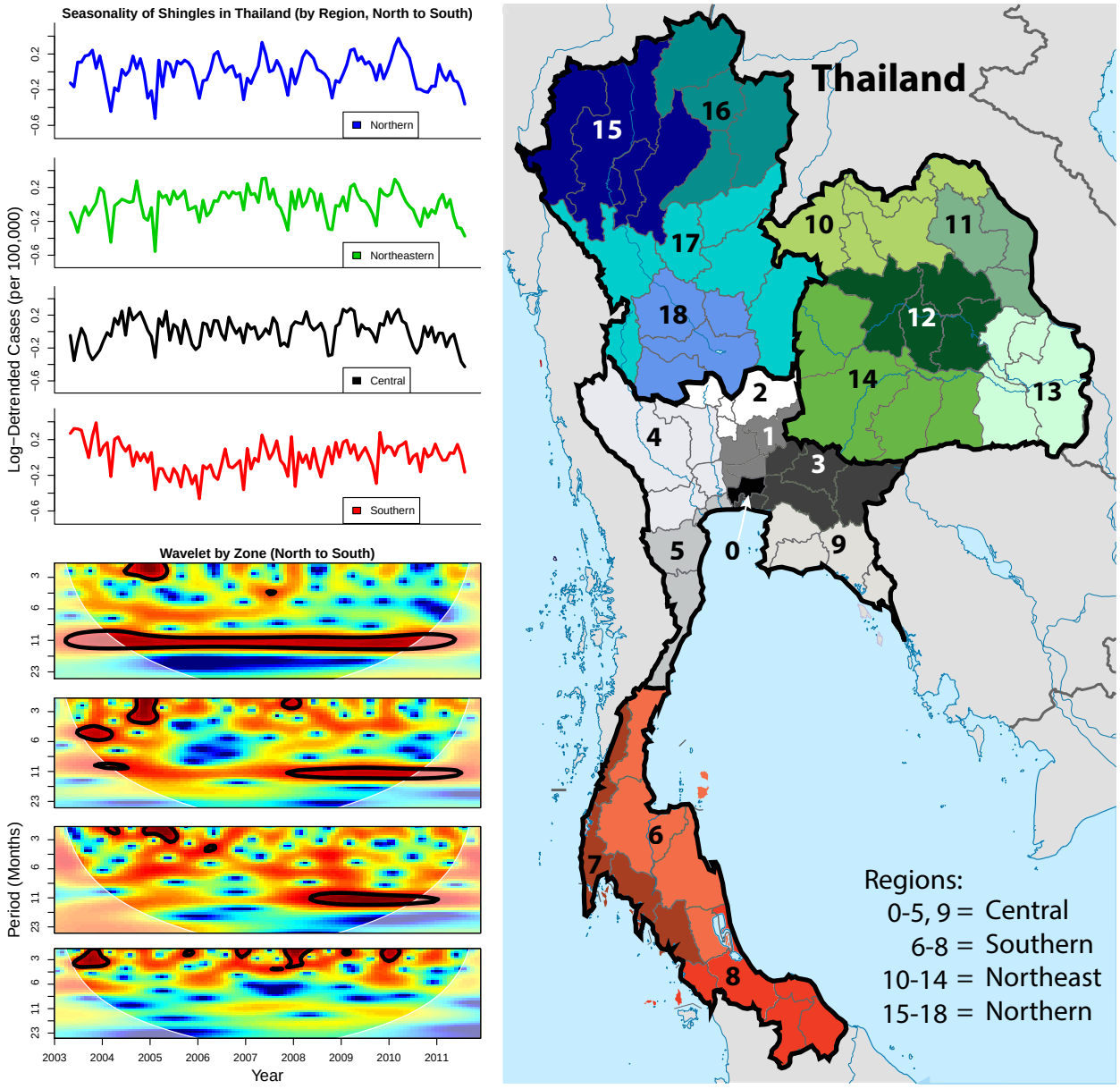


Figure 4.3: (A) (top four panels) Log-detrended monthly incidence of shingles by region demonstrating the seasonality - arranged North to South from top. (bottom four panels) Wavelet analyses of log-detrended monthly shingles incidence by region - arranged North to South from top. Areas within the semi-circle indicate the wavelet cone of influence. (B) Map of Thailand with regions outlined in thick black lines and zones outlined in light grey. credit: Nathan Warriner

December. When examined using a cross-wavelet analysis, both diseases displayed significant annual (12-month) periodicity, with a quarter-year lag (3 months) between chickenpox and shingles incidence (Fig C.8).

Since country-level shingles case reports indicated a novel seasonal pattern, we inves-

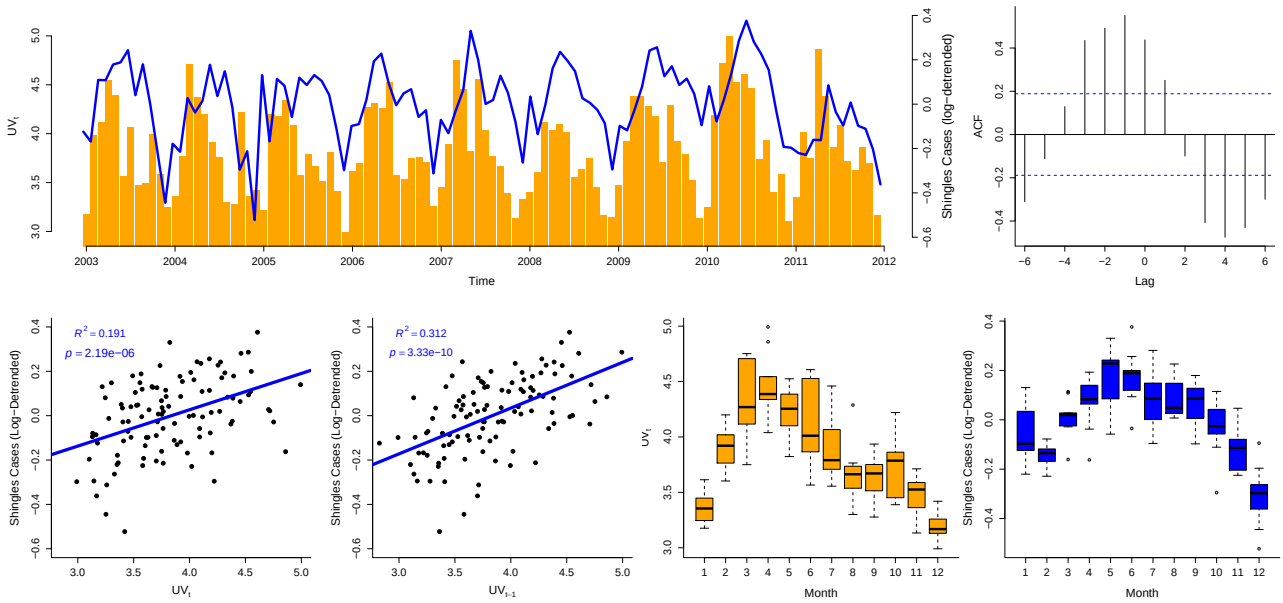


Figure 4.4: The relationship between shingles and UV in Northern Thailand. (A) Monthly log-detrended shingles cases (blue) and monthly UV index (orange). (B) Cross-correlation between shingles and UV. (C) Correlation between UV and log-detrended shingles cases. (D) Correlation between log-detrended shingles cases and UV the previous month. (E-F) Boxplots showing the seasonal distribution of UV index and log-detrended shingles cases.

tigated the possibility of spatial variation in seasonal incidence by region. A latitudinal gradient was observed in shingles seasonality, with the highest latitude region having the strongest consistent seasonality (Fig 4.3A), a pattern which diminished in the southern regions. The northern region, where the strongest seasonal signal of both shingles and UV was located, also had the highest shingles incidence rates in Thailand (Fig C.3).

We observed a significant correlation between the monthly UV index and monthly number of shingles cases. Country level patterns were striking, with increased shingles cases ($R^2 = 0.425$) in months with higher UV exposure (Fig 4.2D). On a regional scale, the UV-shingles relationship was stronger in the northern, northeastern, and central regions than southern (Figs C.4, C.5). Thailand, as a country, spans a latitudinal gradient from 5-21 degrees North, with the northern region encompassing roughly 15-21 degrees north, and the southern region including 5-12 degrees north. Throughout the study-period, 12 month periodicities in both UV and shingles were significant, with a 1/8 year (approximately 1.5 month) lag between

the initial increase of UV and the increased number of shingles cases (Fig C.7). If the UV-shingles relationship is due to the hypothesized UV effects on the immune system, then this lag could be due to the time it takes for (1) UV to affect the immune system, (2) VZV to reactivate and cause shingles, (3) health-care seeking, and (4) subsequent reporting of clinical cases.

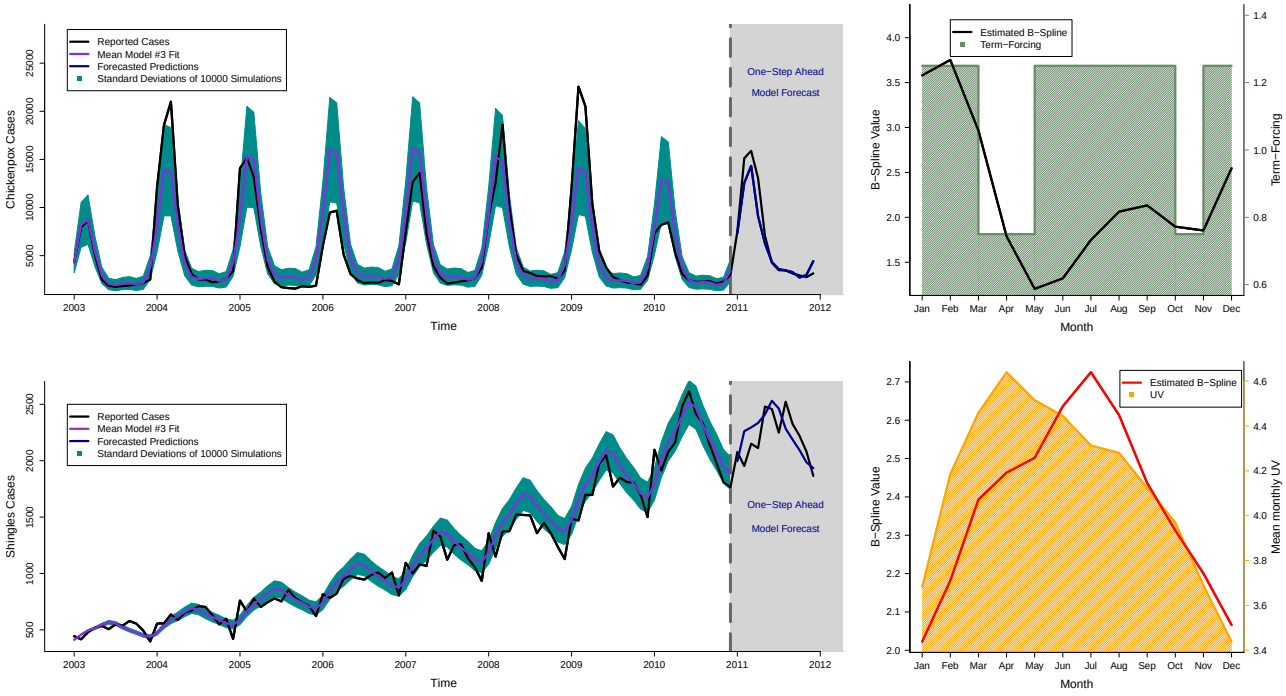


Figure 4.5: Comparison of simulated monthly cases of chickenpox and shingles from the best fit model (#3) with reported clinical cases in Thailand 2003-2011. (A) Reported clinical cases of chickenpox (black) with the mean simulated fit from 10000 simulations (purple) with standard deviation (aqua) through 2010, with forecasted (not fitted) values shown for 2011 (dark blue). (B) Maximum likelihood (B-spline) estimate of the monthly seasonal forcing for chickenpox from the best fit model (black), and school terms for Thailand (green). (C) Reported clinical cases of shingles (black) with the mean simulated fit from 10000 simulations (purple) with standard deviation (aqua) through 2010, with forecasted (not fitted) values shown for 2011 (dark blue). (D) Maximum likelihood (B-spline) estimate of the the monthly seasonal forcing for shingles from the best fit model (red), and mean monthly UV values for Thailand (orange).

4.7 Model Fitting

The current best fit model, model #3, utilized separate seasonal drivers to fit the chickenpox and shingles clinical case reports (Table C.1) (Fig 4.5). This is unsurprising because

the method we used to estimate seasonality, B-splines, are able to encompass all seasonal processes, including any regular seasonal patterns, such as our other proposed mechanism, UV irradiation. This B-spline seasonal driver is a non-parametric representation of seasonal forcing that can account for flexible seasonal patterns, however it is unable to account for any inter-annual variation (deviations) in seasonality. The best-fit mechanistic model, model #3, was simulated 10,000 times and compared with observed cases of chickenpox and shingles (Fig 4.5A,C). The seasonal transmission and reactivation parameters estimated from this model were highly seasonal and closely matched known school terms in Thailand for chickenpox (Fig 4.5B), and national level UV for shingles (Fig 4.5D.) Interestingly, it was unable to fit the early (Feb-Apr) drop in shingles cases (Fig 4.2A & Fig 4.5C)

Furthermore, while model #3 mathematically fits best, the model variation (#5) that included a seasonal forcing for chickenpox and a UV irradiation function for shingles reactivation also fit well, as did the associated models that contained antibody-boosting (#4 & #6) (Table C.1). Further model testing will be required to tease apart these two models, especially since the inclusion of antibody boosting in models (all even numbered models) did not improve AIC (Table C.1). Because all models that do not include antibody-boosting (models 1, 3, 5, & 7) are nested within those that include antibody-boosting (models 2, 4, 6, & 8), the full models (that include antibody-boosting) should always have a better log-likelihood, but they do not. The fact they do not fit as well is likely due to the fact that the models have not yet converged at their maximum likelihoods. However, given the difficulty these models have with meeting, much less exceeding, their model counterparts lacking antibody-boosting, it seems that antibody-boosting is unable to explain a significant amount of additional variance in the data, which we could further examine with F-tests. Thus, even though model #3 fit both chickenpox and shingles dynamics well, with out-of-fit predictions correctly estimating the seasonality and predicting a similar amount of cases for both diseases, it is not yet possible to rule out alternatives, such as the more mechanistic model #6. For now, model #3 does demonstrate that, though simple, it was able to capture

the dynamics of both diseases. (Fig 4.5).

Model Simulations

Model simulations that included immunization demonstrated a surprisingly large drop in the number of chickenpox cases (Fig 4.6). Utilizing the conservative uptake values from the Thailand measles vaccine rollout, where coverage of newborns didn't eclipse 50% until the 4th year, 75% until the 9th year, and 90% until the 13th year (Fig 4.6E), approximately 500,000 cases would have been prevented during our study period 2003-2011 (Fig 4.6D). Alternatively, a more efficient rollout, such as the one for hepatitis would have prevented 550,000 cases; while an ideal scenario of 99% coverage would have prevented 570,000 cases (Fig 4.6D). Since we were intentionally conservative in these simulations by only immunizing newborns, any VZV catch-up immunization campaigns would further reduce the number of simulated cases. Due to the lack of interaction between chickenpox transmission and shingles reactivation in model #3, the simulations had minimal effect on shingles incidence.

4.8 Discussion

In this study we utilized population level herpesvirus data in Thailand to (1) reveal seasonal patterns of chickenpox transmission and shingles reactivation, (2) confirm the strong relationship between ultraviolet irradiation and shingles incidence, (3) identify that chickenpox transmission and shingles reactivation are driven by separate mechanisms and (4) simulate the best fit model under immunization conditions to suggest that 500,000 cases of chickenpox could have been prevented in Thailand had the vaccine been implemented when first licensed.

Our analyses indicated a strong springtime chickenpox transmission pulse, and surprisingly, a 3-month lagged seasonal shingles reactivation peak (Fig 4.2). Previous studies of shingles reactivation at smaller scales (e.g. hospitals or towns) indicated a lack of observable

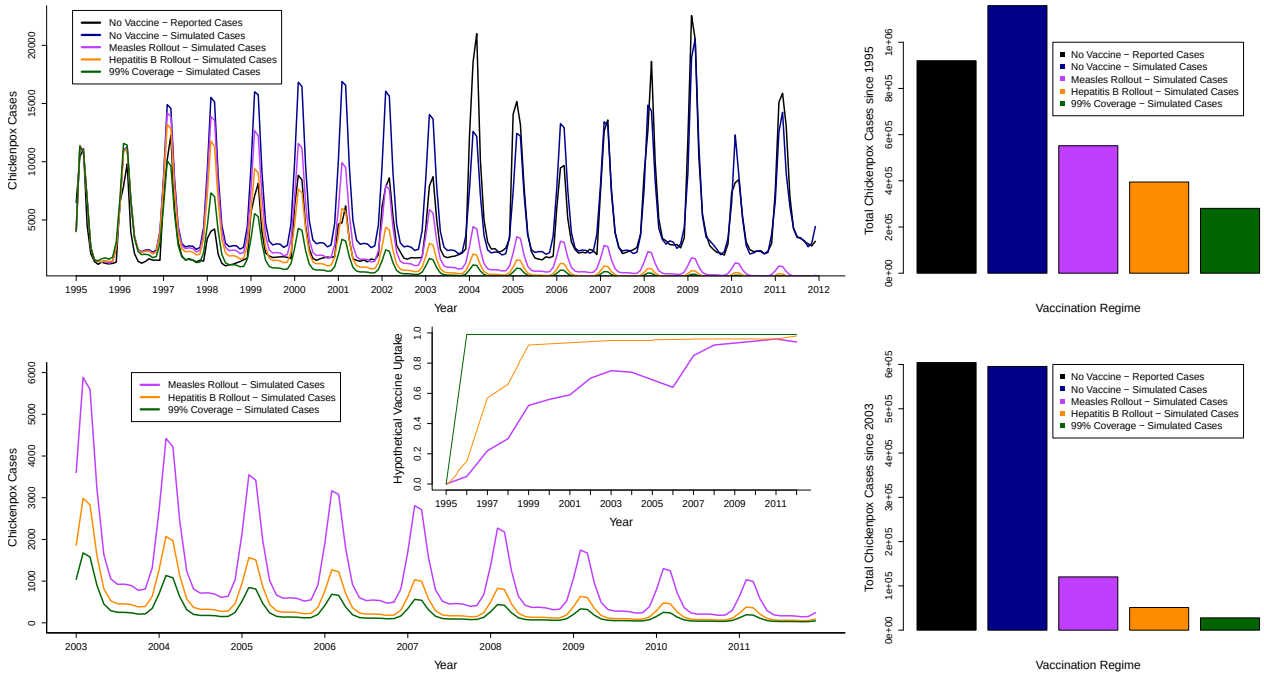


Figure 4.6: Best fit model simulations under three immunization regimes. For all panels; reported chickenpox cases (black), simulated chickenpox cases (blue), simulated chickenpox cases utilizing the data from the 1984 measles vaccine rollout in Thailand (magenta), simulated chickenpox cases utilizing the data from the 1992 hepatitis B, 3rd dose vaccine rollout in Thailand (orange), and simulated chickenpox cases utilizing and ideal rollout where near-perfect (99%) coverage is achieved, and maintained. (A) Time series of monthly cases since 1995. (B) Total number of chickenpox cases over the 17 year period. (C) Time series of simulated models under immunization regimes, during our study period (2003-2011). (D) Total number of chickenpox cases over our 9 year study period. (E) Vaccine uptake used for the three immunization simulations - uptake data for measles and hepatitis are from the actual Thailand rollouts for these vaccines.

seasonality in reactivation [24, 40, 46, 79, 97, 156, 182]. Two possible explanations for the lack of shingles seasonality in previous studies, and in the southern region of our study, include: There could be regional differences in seasonality, with some regions lacking seasonal variation in reactivation risk. For example, the mechanism driving reactivation may not vary seasonally or may not affect a large enough part of the population (e.g. if UV drove shingles reactivation, but everyone worked inside). Alternatively, it could be that seasonal variation in reactivation risk exists across regions, but due to the low amplitude of the variation, it is hard to distinguish from noise in locations with low incidence/fewer cases; thus, having less power to detect seasonality. In our study it became more difficult to identify seasonality as

we broke the data into regions and zones (Figs C.2, C.4, & C.5). Shingles incidence displayed a latitudinal gradient of seasonality (Fig S2 & S4), similar to transmissible diseases such as; polio [133], influenza [98], rotavirus [51], chickenpox [9], bacterial meningitis [147], and respiratory syncytial virus [21]. We identified a strong correlation between shingles reactivation and UV irradiation, providing evidence that UV could have a biological impact on shingles reactivation, similar to the effect it has on HSV.

The best-fit models (#3 — #6) revealed the drivers of chickenpox transmission and shingles reactivation are distinct (Fig 4.5). The fitted seasonal force of infection function for VZV transmission contained increased seasonality in Jan-Feb and a smaller increase in Aug-Sep, both which lagged Thailand school terms. Decreased seasonal forcing for chickenpox was seen in May-Jun and Oct-Nov, both of which were preceded by weeks when students were on vacation (Fig 4.5B). Model #3 identified that the shingles peak reactivation forcing was lagged from the UV-peak, which could be due to the time between the start of the cellular reactivation pathway and the appearance of symptoms [237] (Fig 4.5D). Alternatively, this reactivation delay may be from the time it takes for an accumulation of immune-related stress from extended UV exposure. It is also possible for a reporting lag to exist between symptom appearance and a clinic/doctor visit. UV may, therefore, play a role in the reactivation of shingles, albeit delayed. Molecular investigation of this potential interaction could shed light on the biology underlying herpesvirus reactivation in general. Additional testing of models, especially those examining model #6, will be done, potentially including a lag-function for the relationship between UV and shingles incidence.

The role of antibody-boosting remains unresolved. While model #3 demonstrated that chickenpox exposure need not have any effect on immune boosting and shingles reactivation risk to explain the data, this contradicts prevailing concerns that the VZV vaccine will result in higher shingles incidence by reducing antibody-boosting. This model was also unable to fit the Feb-Apr trough in shingles cases or the inter-annual variation in chickenpox cases, specifically in 2004 and 2009. Additionally, the next best fit model with different mecha-

nisms driving seasonality, model #6, identified antibody-boosting as a critical component for model-fit. Moreover, the models have not all yet converged to their maximum likelihood estimates, as evidenced by the fact that nested models such as #3 & #4 do not show the expected likelihoods (e.g. #4 should be able to do at least as well as #3 but does not). Thus, further work needs to be done to delineate between these two models and, specifically, what effect antibody boosting has on model fit.

In addition to the convergence and estimation issues making distinguishing a single best-fit model difficult, the issue is further complicated by the flexibility of the spline forms. Because the chickenpox data is highly seasonal (and the model fit even more so), much of the effects of antibody-boosting can likely be instead captured by the B-splines, similarly to how the B-splines appear to be capturing term-forcing and UV effects. This suggests that using model fit alone to uncover the antibody-boosting and the other proposed mechanisms (such as UV) will depend strongly on the inter-annual variation, which the splines do not capture (which is also at least partly why the current best-fit model does not capture the inter-annual variation well). Thus, further fitting of the near-best fit models (#4-6) is warranted. However, since UV does not appear to be a driver for chickenpox, we do not have any inter-annually-varying driver that can affect the model-fitted chickenpox (and thus provide inter-annual variation in antibody-boosting), which may limit our ability to determine the role of antibody-boosting without additional time series data on chickenpox drivers (e.g. demographics or contact pattern data). One approach to addressing the lack of time series data on chickenpox drivers might be to fit a reduced model of shingles-only, which uses the inter-annually varying chickenpox data itself as a driver for shingles, via antibody boosting. Another option would be to use the reported monthly chickenpox cases as a covariate in our antibody-boosting function, rather than the model-estimated chickenpox cases (which again, has little inter-annual variation due to the B-spline).

Given that much of the effects of antibody-boosting can be captured with the B-splines, this makes simulation of counterfactuals regarding shingles patterns difficult. Antibody-

boosting effects that are captured in the spline patterns will not respond to changes in chickenpox dynamics as a mechanistic antibody boosting model would, and so won't be affected by simulated vaccination interventions. Thus, we must be careful before interpreting the shingles predictions. By contrast, chickenpox predictions will be less affected by the mechanistic uncertainty in our models, and thus likely to be more robust to the model uncertainty shown here. In future work, we plan to use the set of best-fit candidate models for our intervention simulations, to better capture the range of uncertainty in the antibody-boosting effects.

Because the current best-fit model does not include interaction between chickenpox transmission and shingles reactivation—via the boosting of VZV antibodies in adults—there was minimal effect of vaccination on the number of shingles cases when childhood vaccination was included in the fitted our model. These simulated results are in-line with the US experience with VZV vaccination, which had a 67–84% reduction in overall cases after only 5 years of immunization [188]. While long-term VZV antibody boosting may still be affected by chickenpox incidence, a decrease in boosting is likely unavoidable, as previous epidemiological studies have shown a drop in boosting due to demographic shifts in the population structure, via a decreased birth rate [135]. In Thailand, a massive demographic shift has been occurring over the last four decades, driven by a decreasing birth rate, which has been shown to impact other diseases [54]. Regardless, any drop in antibody-boosting due to chickenpox immunization could be countered by dropping the recommended age for shingles vaccination.

The model was intentionally modular and simple, allowing us to test competing hypotheses for the biological drivers of herpesvirus transmission and reactivation. Our generalized model could be utilized for other herpesviruses, allowing for multiple movements between the infected and latent classes for HSV-1 and HSV-2. For CMV, the model could be altered to include congenital infection in newborns instead of (shingles) reactivation. Finally, a model for EBV would include only one diseased class (mononucleosis), but reactivation would allow the possibility of transmission to susceptible individuals. Simulations which

included realistic immunization uptake rates demonstrated the potential for a substantial drop of nearly a half million chickenpox cases during our 9-year study period (Fig 4.6). As expected, increased vaccine coverage reduced the total number of cases; however, we found that even low vaccine uptake would drastically reduce chickenpox morbidity if higher targets (e.g. $\geq 95\%$) were not feasible.

Herpesviruses are incredibly contagious and the nine viruses within the family possess unique lytic (transmission) and latent (reactivation) life stages, both of which need to be better understood. With almost 100% infection risk globally, herpesviruses silently persist within the human body during latent periods. Likewise, since herpesvirus diseases are non-notifiable, they silently persist within human populations, contributing to the hidden burden of disease. Though uncommon, complications from herpesviruses are severe. With a proven safe, and effective VZV vaccine that protects against chickenpox and shingles, and clinical trials for four other herpesvirus vaccines currently underway, knowledge of the underpinnings of herpesvirus dynamics can unlock the potential for prevention and control of these diseases

4.9 Acknowledgements

I would like to thank Robert Woods, Pej Rohani, and John Drake for multiple discussions and feedback regarding this manuscript.

CHAPTER V

Conclusions

This work examined human birth and disease dynamics across both space and time to better understand the mechanisms that may be responsible for the observed inter- and intra-annual variation seen in infectious disease outbreak dynamics. In addition to these three chapters, I have other research projects in progress; both in terms of understanding birth seasonality and its role in childhood disease dynamics, but also an investigation into the infectious disease dynamics across a broad spatiotemporal landscape.

Chapter 2 [132], focused on characterizing the birth seasonality of each state in the US over a period of 78 years. This research included analyses of birth data from developed countries in the Northern Hemisphere via the World Health Organization (WHO) [226]. I discovered a strong latitudinal relationship to the timing of the birth peak, with countries further from the equator having an earlier birth peak, and those closer to the equator having a later birth peak, indicating a similar mechanism acting on developed countries in the northern hemisphere [132]. Northern US states had birth peaks which occurred $2-4$ months prior to states located in the Southern US, a pattern which also epitomized developed Northern Hemispheric countries. Additionally, in the US a strong latitudinal gradient was observed in the strength of the birth pulse, with states in the south exhibiting much stronger pulses, despite that pulse occurring later in the year. Finally, I implemented this knowledge into a Susceptible-Infected-Recovered (SIR) model to examine whether the seasonality in human

births had the potential to play a role in childhood disease dynamics. I discovered that the timing and strength of the birth pulse in the US was sufficient to alter the magnitude of measles epidemics in 20th century America. Additionally, I found that a stronger seasonal pulse, such as those reported in developing countries [64, 132] would have the potential to significantly alter childhood disease dynamics, especially in diseases which infect younger children (such as rotavirus or respiratory syncytial virus).

While Chapter 2 focused on characterizing birth seasonality in developed countries in the northern hemisphere [132], one of the main critiques of that work was the lack of data from either developing countries or those located in the southern hemisphere. These types of records do not exist for more than a dozen or so countries, so I looked elsewhere, eventually being granted access to a demographic survey database from 102 countries. These survey data were taken over the course of 25 years in developing countries in Africa, Southeast Asia, and South America [60]. This research will take advantage of these long term surveys from developing countries and explore the multivariate facets of birth seasonality (timing and amplitude), and incorporate annual birth rates from the World Health Organization [226], allowing for a full reconstruction of historical monthly birth data in these countries. From these DHS survey data, I have discovered that (I) birth seasonality exists in developing countries and (II) that across developing countries, birth seasonality does not have a latitudinal gradient similar to the developed world [132], but rather a heterogeneous pattern.

In my second chapter, I found that the amplitude of birth seasonality has been decreasing in the United States since the baby-boom [132] A similar pattern was identified in a study focused on births in Spain from 1941-2000 [35]. This led me to ask: how long has this decline in birth amplitude been occurring? I subsequently found, and digitized, 300 years of monthly birth and death data, ranging from the years 1541-1836, from England and Wales, to analyze the season of birth. Upon examining the 60 million births from this period, I found that births were bi-annual from the start of the time series (1541) until the end of the 17th century, around the same time deaths started to decrease and the overall population of

England began to increase rapidly. Birth amplitude was large at the start of the study, due to the small population size, and steadily decreased through the time series. This pattern is surprising, because the decreasing amplitude in birth seasonality has previously been attributed to modern medicine, changes in the workplace environment (farm to office), and the invention of air-conditioning, all of which are 20th century phenomena. Death had an annual periodicity that persisted across the entire time series (47M deaths), and I was able to link periods of increased death to historical pathogen outbreaks (i.e. the Great Plague of London, various influenza epidemics, and repeated smallpox outbreaks.) Combined, these analyses have revealed that the loss of amplitude in birth seasonality has been occurring for much longer than previously identified, that the switch from bi-annual to annual birth peaks in the U.K. transitioned hundreds of years before it occurred in the U.S., and allowed for the linking of increased patterns of deaths to historical disease epidemics.

With all the work I have done characterizing historical birth seasonality around the world, and the less-than-expected impact it has on measles epidemics in the U.S. [132], under what conditions can birth seasonality play a role in childhood disease dynamics? Multiple studies have demonstrated the importance of demographics for understanding childhood infectious disease transmission, particularly the importance of (i) birth seasonality [64, 132] and (ii) age-stratified contact rates [167]. Chapter 2 demonstrated that birth seasonality, in terms of the timing and amplitude of seasonal birth pulses, can independently impact disease dynamics [132]. A recent study that quantified age-stratified contacts across multiple countries in Europe [140] has since been integrated into transmission models of childhood infection to explain demographic shifts in disease incidence [167]. However, these two mechanisms, birth seasonality and an age-stratified network, have never been considered in combination. This work is unique from my previous research, because in an age-stratified model, births enter the first age class rather than a single age class representing the entire population, and term-time forcing can be applied to the appropriate age classes rather than the entire population. This work will allow me to explore how the global range of birth seasonality

(timing, amplitude, rate) parameters affect disease dynamics. Additionally, this work will disentangle the dynamical effects of birth seasonality, independent of term-time forcing on childhood infectious diseases. This research will combine an age-stratified contact network with birth seasonality parameters from across the world to determine how they interact to alter disease epidemic magnitude, timing, and frequency.

While Chapter 2 examined the role of birth seasonality in the dynamics of childhood infectious disease outbreaks, there remained a lack of accessible population-level clinical case reports for many diseases. These are mandatory to begin understanding which mechanisms may be driving seasonal outbreaks of disease. Chapters 3 & 4 were motivated by Fig 5.1, which I received from Tyler Stevenson (a co-author on [9] - Chapter 3).

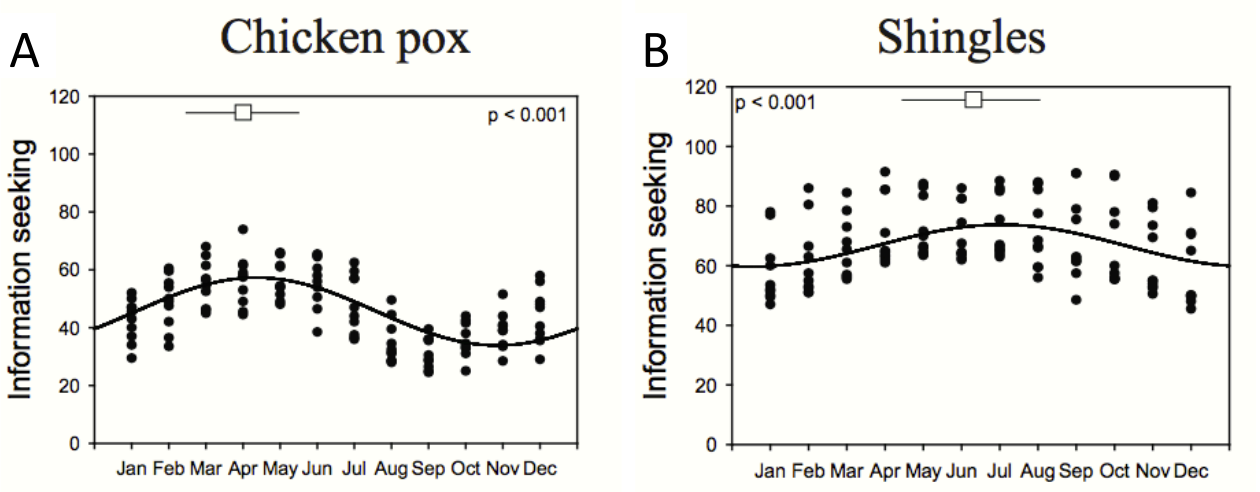


Figure 5.1: Information seeking behaviour data from Google Trends, for (A) chicken pox and (B) shingles searches in the United Kingdom, monthly averages from 2004-2013.

The first of these two chapters [9], Chapter 3, examined digital epidemiology, and the alternate avenues to examine disease dynamics when clinical case data are unavailable. Disease surveillance systems largely focus on infectious diseases with high mortality, while benign diseases often go unreported. Using chicken pox as an example, I demonstrated that internet queries can be used as a proxy for disease incidence when reporting is lacking. I established that Google Trends accurately reflected clinical cases in countries with surveillance, and

thus, population-level dynamics of chicken pox incidence. Then I discovered robust seasonal variation in query behavior, with a striking latitudinal gradient on a global scale. Third, I showed that real-time data-mining of queries could forecast the timing and magnitude of outbreaks. Finally, the analyses revealed that countries with government-mandated vaccination programs have significantly reduced seasonality of queries, indicating vaccination efforts mitigated chicken pox outbreaks.

Other research of mine has focused on understanding the spatial complexities in the timing of historical childhood infectious disease outbreaks in the US and across the world, with a goal of identifying the environmental or demographic mechanisms driving these patterns. Understanding the spatial and temporal variation of infectious disease incidence is pertinent to understanding the ecology governing disease transmission. To date, the testing of hypotheses regarding the mechanisms that generate patterns of disease across a landscape have been limited by the lack of longitudinal disease data. Here, I take advantage of a recently digitized historical dataset of clinical case reports spanning the 20th century in the United States [218] and from over 90 countries traversing the years 1943-1975 [226]. Using these data, I ask (I) do these 17 notifiable infectious diseases occur seasonally, (II) does heterogeneity exist in the epidemic timing of these diseases, and if so, (III) is the heterogeneity driven by seasonally forced environmental or demographic mechanisms? Initial analyses have revealed that all seventeen of these diseases have significant annual outbreaks, with measles, mumps, and rubella displaying multienniel periodicities. A south (early)-to-north (late) gradient in the timing of the outbreak peak for both polio and typhoid fever was detected. A strong geographical pattern in the timing of the outbreak peak for diphtheria and scarlet fever were also identified, with states in the southeast had outbreak peaks 2 months earlier than those states in the midwest, northeast and western US. Analyses are ongoing for country level data, but all examined thus far display annual periodicities. With this knowledge, I will examine the association between three environmental variables (rainfall, temperature, and humidity) and four demographic variables (birth rate, timing, amplitude,

population size) and the incidence of any diseases (so far; typhoid fever, scarlet fever and diphtheria) that display a geographic pattern with a mechanistic model. Finally, I will integrate gravity-coupling into my mechanistic models to capture the spatiotemporal properties of these epidemics.

One of these diseases is polio, which is the next human disease slated for eradication. The first recorded outbreaks of Wild Poliovirus (WPV) are over 100 years old, when an almost simultaneous appearance of epidemic WPV appeared across Scandinavia, Western Europe, and the United States [121]. Prior to the introduction of the polio vaccine in the 1950s [176], WPV outbreaks were strongly seasonal in temperate regions, with increased incidence in the summer months [102], while seasonality was virtually absent near the equator [73,186]. The seasonal patterns of WPV incidence have led to hypotheses that link seasonal environmental factors, such as temperature, humidity, or rainfall to the seasonality of WPV transmission [5, 94, 102, 193, 232]. However, to date there has been no definitive study to identify which factors are most strongly correlated to the seasonal patterns observed in WPV outbreaks [141]. With the 2016 re-emergence of WPV in Nigeria, there are currently three countries experiencing endemic transmission; Afghanistan, Nigeria, and Pakistan [87]. Polio is a viral infection transmitted fecal-orally. Symptomatic infections last for approximately 10 days and include fever, sore throat, headache, and stiffness. Paralytic polio infections are rare, comprising less than 1% of all infections. A typical infection lasts 5-45 days, which includes the infectious period where an infected individual is actively transmitting WPV to other individuals [141, 141, 176].

WPV seasonality has been noted for its irregular and non-synchronized dynamics across space in temperate regions [102,133]. In regions near the equator, little to no seasonality has been observed in WPV dynamics [73, 186]. Recent work has illustrated that in order to reduce or eradicate any infectious pathogen, it is essential to understand the local climatic [114] and demographic [75] conditions. This research will explore environmental, demographic, and social factors that may be associated with WPV seasonality across a spatially resolved

landscape in both temperate and tropical regions. By examining WPV seasonality across latitudinal and longitudinal gradients, I will gain an understanding of the factors most commonly associated with WPV incidence in each location. In order to infer which of these factors are most important, I will combine process-based dynamical models with long-term time series spatially replicated across populations, to understand the environmental, demographic, and social factors associated with WPV persistence and transmission. Similar methods have been used in the study of other seasonal childhood infectious diseases; for example, the spatial variation in the timing of pertussis outbreaks in the US during the 1950s led Choisy and Rohani to discover that nearby cities had synchronized outbreaks that catalyzed outbreaks in local regions that then moved inward towards the interior of the country from both coasts [47]. Similarly, Grenfell et al. [89] demonstrated that the historical persistence of measles in England and Wales was driven by seasonal outbreaks in larger cities which re-seeded smaller populations where the virus had previously gone extinct. These studies combined long-term, spatially replicated data with disease-specific transmission models to discern what drove the seasonality and persistence of disease transmission.

Despite more than 100 years of polio outbreaks, there remains a lack of understanding as to why polio transmission and incidence has such strong seasonality. While the link between environmental variables and transmission has been suggested, there are no data-driven modeling efforts to link a specific factor with WPV seasonality. This work will use extensive historical records to examine which environmental, demographic, or social factors are most strongly associated with historical polio outbreaks, by constructing models that test these hypotheses.

The Global Polio Eradication Initiative (GPEI) has set and missed four polio eradication deadlines. Current immunization strategies include national and subnational/supplementary immunization activities wherein eradication campaigners maximize vaccination efforts for certain days during the year [87]. Despite the GPEIs efforts in distributing the attenuated Oral Polio Vaccine (OPV), WPV remains endemic in Pakistan, Afghanistan, and Nigeria.

In 2014, these countries re-seeded WPV in Central Africa, the Horn of Africa, war-torn Syria, and the Middle East [62]. In Nigeria in 2016, where WPV was thought extinct, new cases emerged in the northern regions of war-torn (Boko Haram) Borno state [87]. With global unrest in Syria and multiple African nations unlikely to end soon, a new strategy must be considered to maximize vaccine coverage in the global push towards final eradication. Historical work on polio has established that individuals vaccinated with OPV shed this attenuated poliovirus seasonally, which can then indirectly immunize others within the population against polio infection [18, 95, 101, 116, 127, 223]. Furthermore, historical sewage sampling has revealed that OPV shedding is maximal at the same peak timing of WPV incidence [66]. Thus, by identifying which environmental, demographic, or social factors are most strongly associated with seasonal WPV transmission, public health officials will be able to geographically tailor OPV campaigns to facilitate OPV shedding, thus maximizing immunization coverage.

In Chapter 4, I utilized models to examine which factors were associated with herpesvirus transmission and reactivation. This work has not been previously done due to the lack of spatio-temporal data, despite nearly every person in the world being infected with at least one herpesvirus in their lifetime. There are nine herpesviruses which vary slightly in their biology, however they are all defined by their latent, recurring infections. Chickenpox, caused by the varicella zoster virus (VZV) is the only herpesvirus studied at the population scale. In the approximately 20% of individuals, VZV reactivates, causing shingles. Here, I used spatiotemporal data from Thailand to examine the potential drivers of chickenpox transmission and shingles reactivation, including an exploration of potential immune-modulated interactions between the two disease manifestations. Preliminary models indicate that the mechanisms driving transmission and reactivation are independent, with a strong association between ultraviolat irradiation and reactivation. Additionally, the models suggested that more than a half-million chickenpox cases could have been prevented during the 9-year

study period had immunization been initiated, including more than 2000 hospitalizations and 84 deaths. The current best-fit model also indicated that the vaccine would have little-to-no effect on shingles incidence. Had chickenpox vaccination demonstrated that shingles reactivation would increase, the shingles vaccine could be included into the immunization schedule to prevent VZV reactivation. With the global controversy surrounding VZV immunization, and its potential effects on VZV reactivation, this work recommends that all countries introduce mandatory chickenpox and shingles immunization.

In this dissertation, I have identified previously unknown spatiotemporal patterns in birth seasonality and infectious disease dynamics. In addition to describing these patterns, I have attempted to discern (work-in-progress) the mechanisms driving childhood infectious disease dynamics. While most the data I use are historical, they can be applied to modern-day outbreaks. Diphtheria had $30,000$ reported cases and $3,000$ reported deaths in the year 2000 [226], and recent scarlet fever outbreaks in England are the highest levels in 24 years [143]. Typhoid fever, a summer occurring disease, is estimated to have infected more than 21 million and killed more than $200,000$ worldwide in the year 2000 alone [53], and while more than 400 cases of polio were reported as recently as 2013 [153], recent studies have indicated that silent transmission, where visible infection is lacking, may raise this estimate by multiple orders of magnitude [133]. Despite the high global morbidity and mortality of childhood infectious disease, the mechanisms driving most childhood infections remain poorly understood. It is my goal in the coming years to assist in the final eradication of wild poliovirus transmission, by gaining insights into the mechanisms driving the seasonal transmission patterns of this virus, and others. By employing large datasets, in terms of demographic, environmental, and clinical case reports, I hope to provide public health officials with the information they need to decrease the worldwide burden of infectious disease.

APPENDICES

APPENDIX A

Supporting Information: Human Birth Seasonality: Latitudinal Gradient and Interplay with Childhood Disease Dynamics

A.1 Human Birth Seasonality: Latitudinal Gradient and Interplay with Childhood Disease Dynamics: Supporting Information

A.1.1 Birth Timing and Amplitude

In our analyses we followed the work of Rosenberg, who stated that adjusting for the differing number of days in each month had little effect on analyses of birth seasonality [169]. Thus, we did not make any adjustments of our time series to account for the different number of days in each month.

In the wavelet spectral analysis we tested for birth periodicity with periods ranging from 2 months to one-third the length of each data series. Since a significant 1 year period was observed, we constructed monthly phase angle time series for each data series using an 11-13

month period. The phase angle time series were subsequently used to determine the timing of the annual birth peak for each location. Peak-birth months were then averaged for each individual data series, with the U.S. states mapped to visualize the geographical variation in the timing of the annual birth peak. When biannual peak-births occurred in the U.S., they were separated into two 6 month periods: summer (May-Oct) and winter (Nov-April) (Figs. A.1 & A.2).

The analysis of seasonal birth amplitude, or percent deviation from the annual mean, was done using seasonally decomposed time series. The stl function in the stats package in R was used to decompose the data into seasonal ($\mathcal{S}$), trend ($\mathcal{T}$), and noise ($\mathcal{N}$) components for each data series. The noise free time series were constructed as:

$$\mathcal{F} = \mathcal{S} + \mathcal{T} \tag{A.1}$$

The deviation from the mean during the birth peak was calcualted for each year, i, as:

$$x = \max(\mathcal{F}_i) - \text{mean}(\mathcal{T}_i) \tag{A.2}$$

The deviation from the mean during the birth trough was calculated for each year, i, as:

$$y = \min(\mathcal{F}_i) - \text{mean}(\mathcal{T}_i) \tag{A.3}$$

Thus, the one-half peak-trough difference is:

$$z = \frac{x - y}{2} \tag{A.4}$$

The seasonal amplitude, measured as a percent deviation from the mean, was calculated

as:

$$\text{amplitude} = \frac{z}{\text{mean}(\mathcal{T}_i)} \tag{A.5}$$

A.1.1.1 Simulation Study

For the simulation study we used a daily discrete-time SEIR model of measles adopted from Earn et al. 2000 [69]. The model has a daily time step and uses school-term forcing of seasonal transmission based on the school terms of England & Wales. The models assume transition probabilities follow a Poisson process. The difference equations are as follows:

$$S_{t+1} = \mu_t N_t + S_t e^{-(\beta_t I_t + \delta)} \tag{A.6}$$

$$E_{t+1} = S_t \left(1 - e^{-(\beta_t I_t + \delta)}\right) \frac{\beta_t I_t}{\beta_t I_t + \delta} + E_t e^{-(\phi + \delta)} \tag{A.7}$$

$$I_{t+1} = E_t \left(1 - e^{-(\phi + \delta)}\right) \frac{\phi}{\phi + \delta} + I_t e^{-(\gamma + \delta)} \tag{A.8}$$

$$R_{t+1} = I_t \left(1 - e^{-(\gamma + \delta)}\right) \frac{\gamma}{\gamma + \delta} + R_t e^{-\delta} \tag{A.9}$$

$$N_t = S_t + E_t + I_t + R_t \tag{A.10}$$

$$Incidence_t = E_t \left(1 - e^{-(\phi + \delta)}\right) \frac{\phi}{\phi + \delta} \tag{A.11}$$

$$\mu_t = \frac{\nu + A \sin(\omega t + \sigma)}{30} \tag{A.12}$$

$$\beta_t = \frac{B_0}{\frac{1}{365}\left((1+b_1)273 + (1-b_1)92\right)}\left(1 + b_1 Term_t\right) \tag{A.13}$$

$Term_t$ is based off the the school term schedule. When school is in session $Term_t = 1$ and when students are on holiday $Term_t = -1$. See Table A.1 for the school term schedule. The parameter values used for the simulation study can be found in Table A.2. All simulations were run for 100 - 150 yrs. to ensure that the trajectories were past the transient phase.

Holiday	*Model Days*	*Calendar Days*
Christmas	356 - 6	Dec 21 - Jan 6
Easter	100 - 115	Apr 10 - 25
Summer	200 - 251	Jul 19 - Sept 8
Autumn Half Term	300 - 307	Oct 27 - Nov 3

Table A.1: School term schedule. When students are on holiday $Term_t = -1$ otherwise 1.

Parameter	*Value*	*Parameter*	*Value*
$\mathcal{R}_0$	16 (basic reproductive no.)	b_1	0.25
D	1 - 365 (birth peak day)	μ_0	$\frac{1}{18250}$ day^{-1}
δ	$\frac{1}{18250}$ day^{-1} (50 yr life span)	β_0	$\frac{\mathcal{R}_0}{5}$ day^{-1}
ϕ	$\frac{1}{8}$ day^{-1} (8 day latent period)	S_0	0.06
γ	$\frac{1}{5}$ day^{-1} (5 day infectious period)	E_0	0.001
ν	$\frac{30}{18250}$ month^{-1} (balances δ)	I_0	0.001
A	0 - 0.0009208 month^{-1} (0 - 56% birth amp.)	R_0	0.938
ω	$\frac{2\pi}{365}$ $\frac{\text{radians}}{\text{day}}$	N_0	$S_0 + E_0 + I_0 + R_0$
σ	$\frac{\pi}{2} - \frac{2\pi}{365}D$	$Incidence_0$	0
B_0	$\frac{\mathcal{R}_0}{5}$		

Table A.2: Parameters used in simulation study, main text Figure 1.4A & 1.4B.

A.1.2 Inference Study Using Simulated Data

For the inference study we coupled our SEIR model (Eqs. A.6 - A.13) with a stochastic measurement model. The measurement model is as follows:

$$\text{cases}_t \sim \text{normal}(\rho Incidence_t, \rho\tau) \tag{A.14}$$

In order to test whether the seasonality in births influences parameter estimation, we simulated case data using three parameterizations of our model, each differing in the timing of the birth peak. The parameters used to generate the data are given in Table A.3. Simulations were run to year 50 to ensure the transient phase had passed.

Parameter	*Value*	*Parameter*	*Value*
$\mathcal{R}_0$	$\frac{6250}{365}$ (basic reproductive no.)	b_1	0.25
D	162, 295, or 351 (birth peak day)	μ_0	$\frac{1}{18250}$ day^{-1}
δ	$\frac{1}{18250}$ day^{-1} (50 yr life span)	β_0	$\frac{\mathcal{R}_0}{5}$ day^{-1}
ϕ	$\frac{1}{8}$ day^{-1} (8 day latent period)	S_0	0.06
γ	$\frac{1}{5}$ day^{-1} (5 day infectious period)	E_0	0.001
ν	$\frac{30}{18250}$ month^{-1} (balances δ)	I_0	0.001
A	0.000456 month^{-1} ($\sim$ 28% birth amp.)	R_0	0.938
ω	$\frac{2\pi}{365}$ $\frac{\text{radians}}{\text{day}}$	N_0	1
σ	$\frac{\pi}{2} - \frac{2\pi}{365}D$	$Incidence_0$	0
B_0	$\frac{\mathcal{R}_0}{5}$		

Table A.**3:** Parameters used to generate data for study, main text Figure 1.4C.

The three time series generated using the stochastic SEIR model were then fit to the SEIR model with a 0% birth amplitude, i.e. $A = 0$. The mean transmission rate, B_0, was the only free parameter. All other parameters, aside from A and B_0, were fixed at the values used to generate the data. Only the last 6 years of the time series were used for fitting, thus, the initial conditions were set to match those at the beginning of data used for inference.

For each time series B_0 was profiled and the likelihoods of the parameter sets with varying values of B_0 were calculated using a particle filtering in the R package pomp [112]. A particle filter (a.k.a. Sequential Monte Carlo) is a method of integrating state variables of a stochastic system and estimating the likelihood of the model for a fixed parameter set, given the data. However, since our model lacked process noise, we were able to obtain the exact likelihood for each parameter set.

A.1.3 Inference Study Using New York City Measles Data

For the New York inference study we utilized a Partially Observed Markov Process (POMP) model which are suited for dealing with epidemiological data where the state variables (susceptible, infected, recovered individuals) are not observed in the data, rather the infected individuals are partially observed through case reports [112]. For our process model we used a stochastic biweekly discrete-time SIR model. Similar to the model used for the simulation study, transitions were modeled using a Poisson process. The process model is as follows:

$$\lambda_t = \left(\beta_t \frac{I_t}{N_t} + \psi\right) \epsilon_t \tag{A.15}$$

$$\varrho_t = e^{-dt(\lambda_t + \delta)} \tag{A.16}$$

$$S_{t+1} = dtB_t + \varrho_t S_t \tag{A.17}$$

$$I_{t+1} = (1 - \varrho_t)\, S_t \frac{\lambda_t}{\lambda_t + \delta} \tag{A.18}$$

The transmission rate β_t was modeled using a periodic B-spline with 6 bases, a degree of 2, and a period of 1 year. The process noise ϵ_t was modeled as $\epsilon_t \sim \text{normal}(1, \beta_{sd})$. The covariates B_t, monthly number of individuals entering the susceptible class, and N_t, population size, were taken from data. All parameters were estimated using iterated particle filtering [112] (discussed later), with the exception of the death rate which was fixed at $\delta = \frac{1}{600}\ month^{-1}$, i.e. 50 yr. life span, and $dt = \frac{1}{2}$ fixing the time step to biweekly. In order to couple our model with measles case data we overlaid the process model with a stochastic measurement model. The measurement model is as follows:

$$cases_t \sim \text{normal}(\rho I_t, \tau I_t) \tag{A.19}$$

The case data, gathered from [69], consisted of monthly measles cases for New York City from January 1949 - December 1962. Although we did not have birth data for New York City we did have per capita monthly births for the state of New York. Thus, we assumed the per capita monthly birth rate for New York City was equal to the per capita monthly birth rate for New York state. The population size of New York City was taken from the decadal census and the population size was interpolated for non-census years. Taking the New York City population size together with the time series of per capita monthly births we constructed a time series of the number of monthly births, $\mathcal{B}_t$, in New York City (not to be confused with B_t in Eqn. A.16, which will be explained in the next section).

In order to test whether the seasonality in births influences model parameterization, we used four variants of our model (Eqn. A.20-A.23), each differing in the susceptible recruitment covariate, B_t. The first three models contain birth seasonality and account for the

existence of maternal antibodies for 3-9 months. Whereas, in the fourth model we removed the birth seasonality. The first model variant lags the births by 3 months to account for a scenario where maternal antibodies confer protection from measles for the first 3 months of life:

$$B_t = \mathcal{B}_{t-3} \tag{A.20}$$

In the second model variant we lag births by 6 months.

$$B_t = \mathcal{B}_{t-6} \tag{A.21}$$

In the third model variant we lag births by 9 months.

$$B_t = \mathcal{B}_{t-9} \tag{A.22}$$

In the fourth model variant we removed the seasonality of births by making the monthly births constant within each year by setting them equal to the mean monthly births for the year:

$$B_t = \frac{\sum_{i=1}^{12} \mathcal{B}_{i,j}}{12}; j \in [1948 : 1962], \tag{A.23}$$

where i is the month and j is the year.

Each of the four model variants were independently fit to the data using Maximization by Iterated particle Filtering (MIF) using the R package POMP. MIF is a state-of-the-art simulation based method for parameter estimation that uses likelihood as the objective function. The basis of MIF is particle filtering (a.k.a. Sequential Monte Carlo), which is a method of integrating state variables of a stochastic system and estimating the likelihood of the model for a fixed parameter set, given the data. Unlike particle filtering, which uses

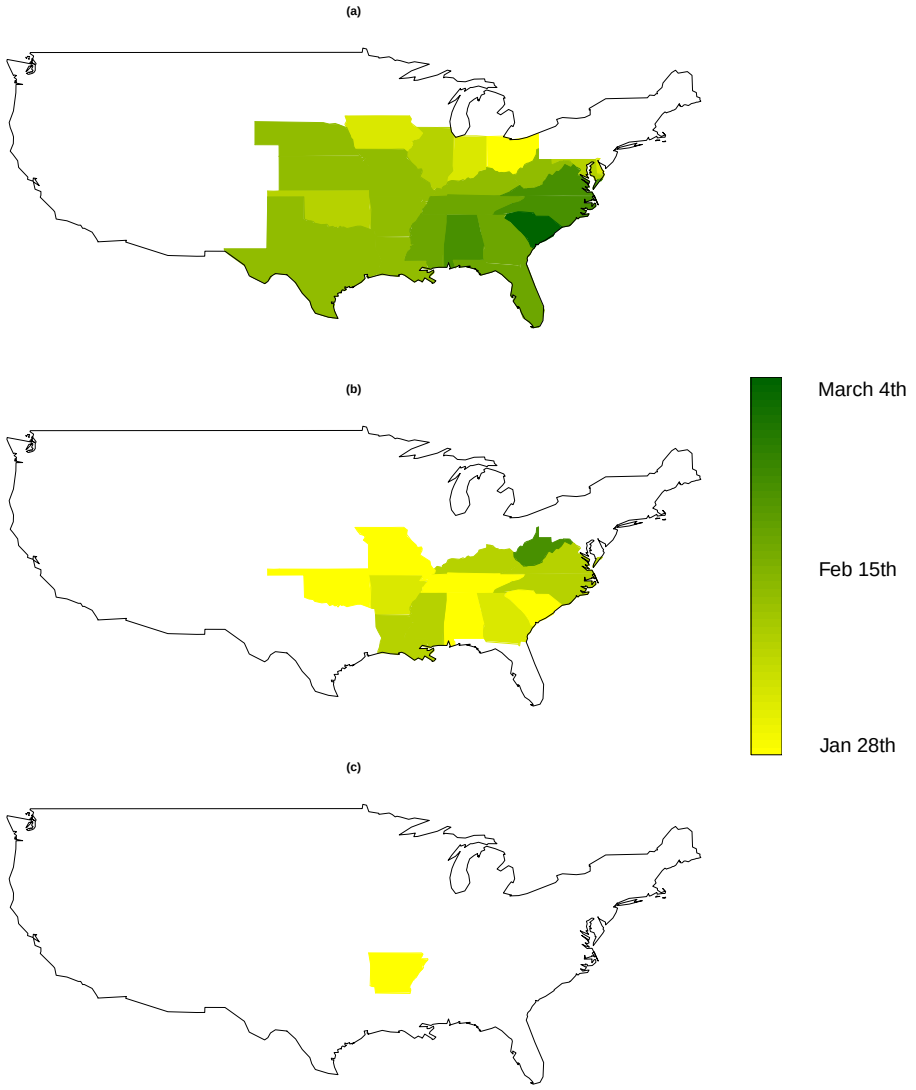
fixed parameter values, MIF varies parameter values throughout the filtering process and selectively propagates particles (in the simplest sense, parameter sets) that have the highest likelihood. Thus, by initializing MIF throughout parameter space one can get a picture of the likelihood surface and identify the maximum likelihood parameter combinations within that space. For each of our models, MIF was initialized with 80000 parameter sets generated using a Sobol design, which pseudo-randomly samples parameters across parameter space in order to evenly sample the space. After this initial phase of MIF, parameter sets were passed through 15 successive stages of MIF, which included profiling. In total, for each model MIF was initialized at over 424000+ locations in parameter space to estimate the shape of the likelihood surface and identify the maximum likelihood parameter set(s). Table A.4 provides the maximum likelihood estimate (MLE) parameter set for each model.

A.1.4 Results.

A.1.4.1 Detailed Results

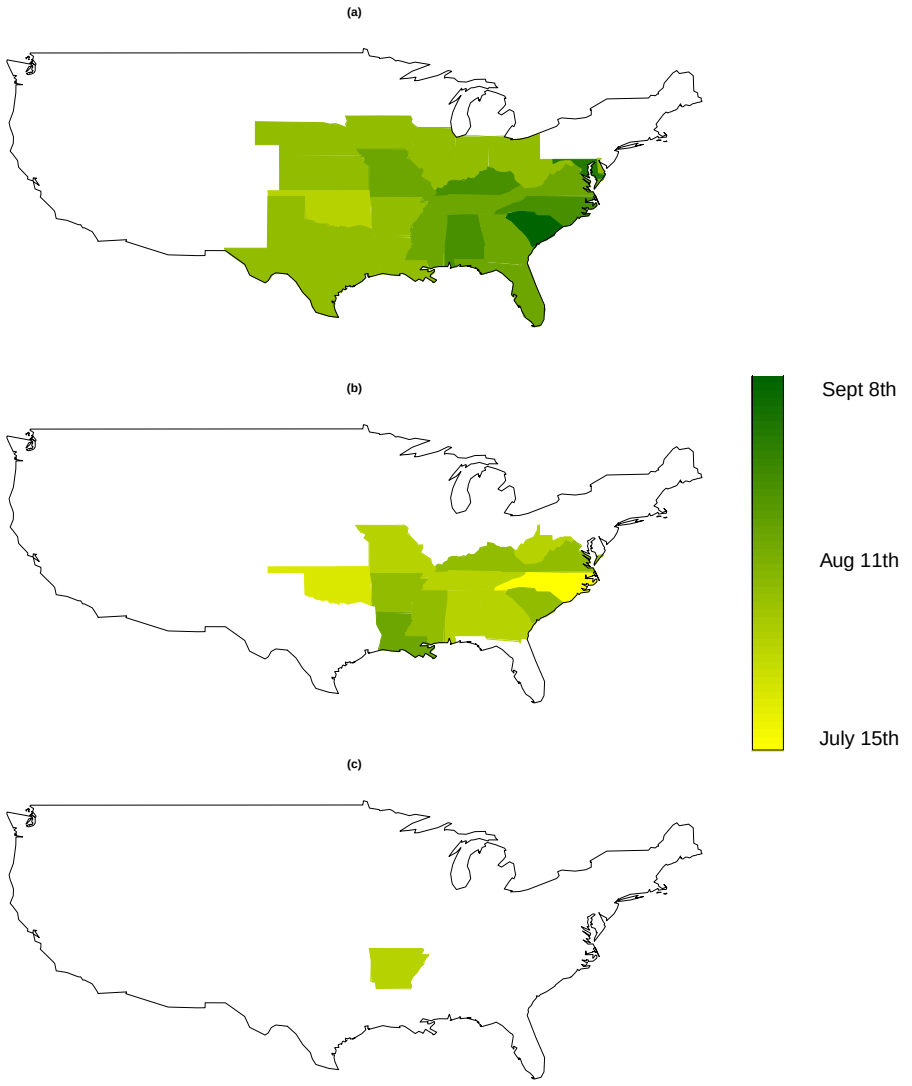
Biannual birth pulses. During the pre-baby boom and baby boom eras, we found that some states had two birth pulses per year, i.e. they had a significant biannual period. All states significant for the biannual period were clustered together in the lower-midwest, deep south, and southeast (Figs. A.1 & A.2). In the baby boom era, some of the states lost their significant biannual period and transitioned to having only a single seasonal birth pulse (Figs. A.1, & A.2). In the modern era, Arkansas remained the only state with a biannual period. The clustering of the states with a significant 6 month period in the southeastern U.S. may have been due to now defunct cultural factors (Figs. A.1 & A.2).

Birth rates. Raw birth rates in the pre-baby boom era ranged from 0.89/1000/month in Nevada (February, 1936) to 2.80/1000/month in New Mexico (May, 1932), with the mean and median both approximately 1.60/1000/month; while in the baby boom era Maryland had the lowest birth rate at 1.13/1000/month (April, 1950), New Mexico with the highest



Supplementary Figure A.1: Bi-annual winter (November-April) birth peak timing by period. (a) pre-baby boom (1931-1945), (b) baby boom (1946-1965), and (c) present period (1965-2008). States shown in white were not significant.

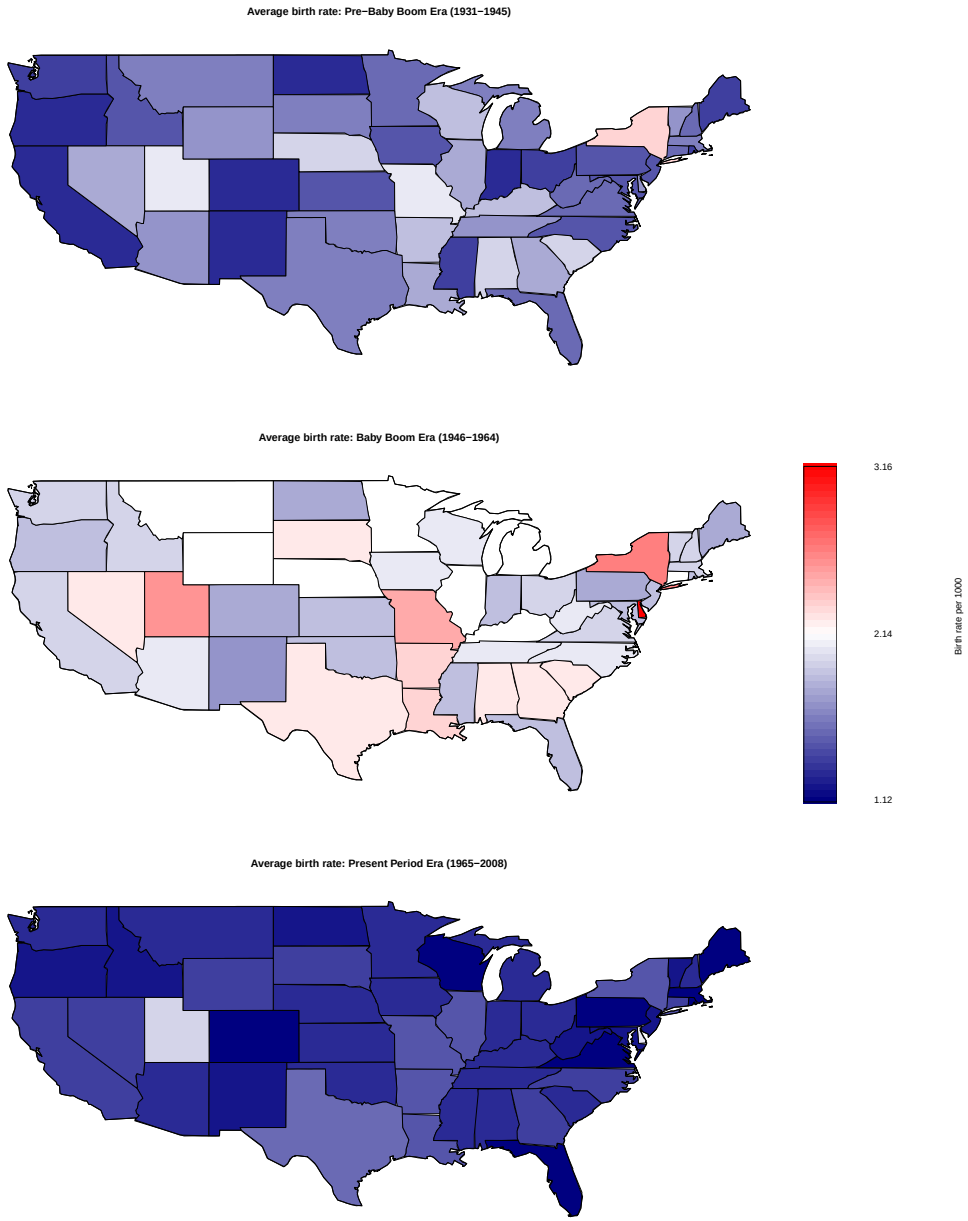
birth rates at 3.36/1000/month (October, 1946), and the mean and median both approximately 2.04/1000/month. In the present period Vermont had the lowest birth rate at 0.67/1000/month (July, 2005), and Utah had the highest at 2.61/1000/month (July, 1977) with the mean and median falling in around 1.27/1000/month. Worldwide birth rates were not calculated, because we did not have population size data for our 200+ countries, rather raw birth values per month were used for wavelet spectral analysis. See Figure A.3 for maps



Supplementary Figure A.2: Bi-annual summer (May-October) birth peak timing by period. (a) pre-baby boom (1931-1945), (b) baby boom (1946-1965), and (c) present period (1965-2008). States shown in white were not significant.

of the mean birth rates in each state and each era.

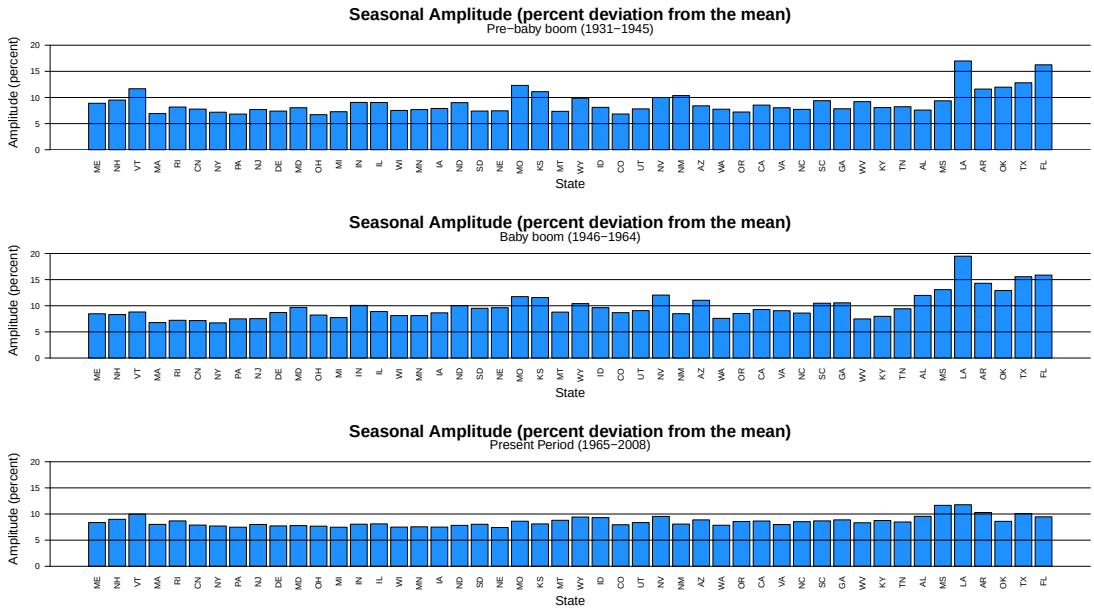
The seasonal birth pulse. Examining the phase angle time series at a period ranging from 11-13 months, the U.S. data had twenty four states significant in the pre-baby boom era, whereas all were significant in the baby boom and present eras. Of the 210 worldwide data series analyzed, 132 (63%) were significant at an 11-13 month period. Many of those found insignificant were shorter time series (5-7 years) or countries with extremely low birth



Supplementary Figure A.**3:** Maps of mean birth rates for each state, in each era. Top to bottom: pre-baby boom, baby boom, and present era. No geographic pattern could be easily discerned.

values (<100 individuals/month). Those states and countries found to be significant were then analyzed for the timing of the peak birth month using a wavelet spectral analysis.

U.S. timing of the seasonal birth peak. During the pre-baby boom era, of the states with a significant 1 year period, Oregon (June 12th) and Maine (June 10th) had the earliest



Supplementary Figure A.4: Seasonal amplitude of births in U.S. states during each era. Note the high amplitude in the southern states (far right in all panels).

peak birth timing —excluding New Mexico, which is an outlier because it has an early peak in all eras yet is located in the southern U.S. Florida (Nov 10th) had the latest peak birth timing in early November. The median peak birth timing was July 3rd, and the mean was July 26th. The range of peak birth timing was at a maximum in the pre-baby boom era with a length of 156 days, approximately 5 months. In this era there was a subtle pattern in the birth peak timing, with the northeast and northwest having earlier peaks than that of the deep south, the southeast and California.

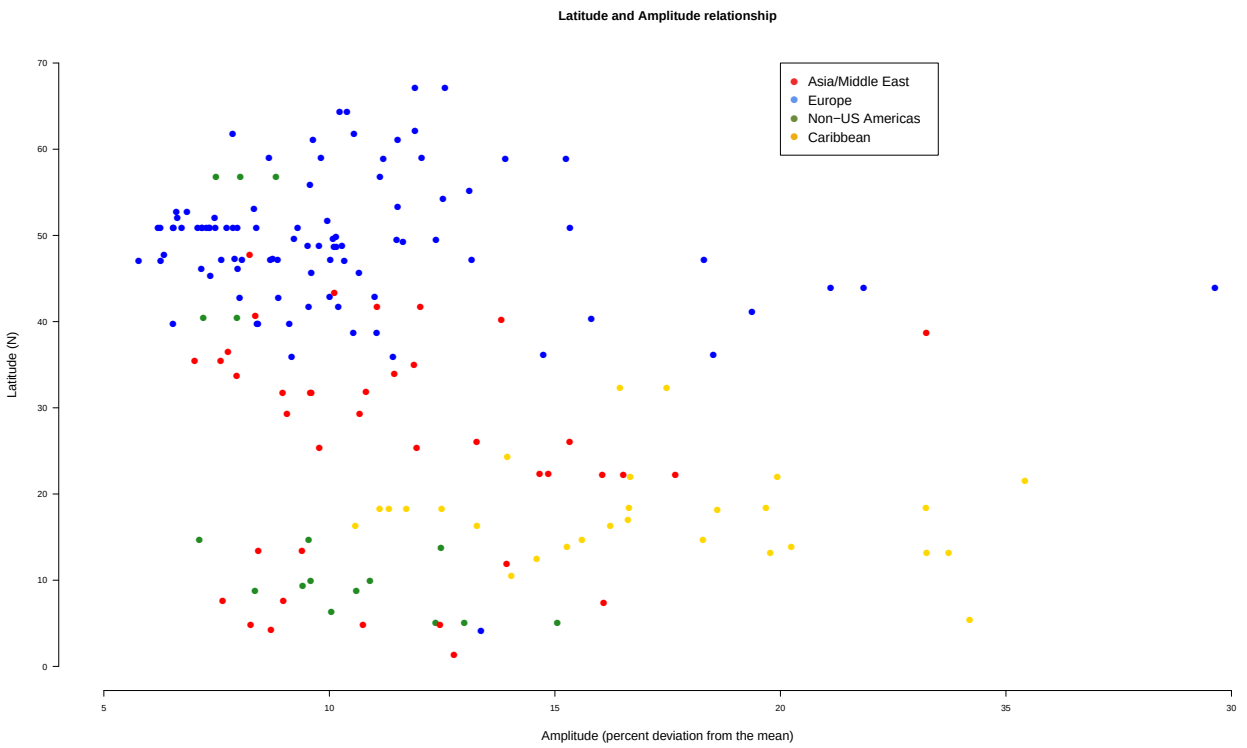
In the baby boom era, when all states were significant at the 1 year period, there was a clear latitudinal gradient in birth amplitude and peak birth timing. Northern states saw an earlier peak birth timing; Utah beginning with a peak birth timing in mid-July (July 13th), followed by Washington with a peak birth timing a few days later (July 18th) with no other peak birth timing occurring for at least another week after that. Again, Florida had the latest peak birth timing 3 months later in mid-October (Oct 21st). The mean (Sept 4th) and median (Sept 8th) both occurred in early September. In the baby boom era, other than

Maine and those states already mentioned, the only states to have a peak prior to August 1st were also located in the northwestern U.S. (Idaho, Montana, North Dakota, and Oregon).

In the modern era, the spatial pattern of peak births was further elucidated with the mid-latitude states acting as a gradient for the northern and southern states. Utah again had the earliest peak birth timing in late June (June 26th), while Florida had the latest peak birth timing in early October (Oct 5th). The mean (Aug 11th) and median (Aug 7th) in this era both occurred in early August. Maine and Vermont also had peak birth timing prior to July 15th, which were the earliest peaks east of the Rocky Mountains. As with the baby boom era, many of the northwest states (Utah, Idaho, Montana, Oregon, Washington, and Wyoming) had peaks prior to July 15th in the present era. Over all eras, Florida consistently had the latest peak ranging from early October in the present period, to early-November in the pre-baby boom era.

Worldwide timing of the seasonal birth peak. The worldwide birth peak timing followed a similar pattern as observed in the U.S., with countries at higher latitudes having an earlier birth peak than those closer to the equator. The earliest peak was in Italy during the period 1970-1985 (March 22nd) followed closely by Tajikstan in the period 1989-1994 (April 15th). The latest birth peak occurred in Saint Vincent and the Grenadines during the period 1992-2005 (November 17th). The mean worldwide peak (U.S. states not included) was mid-August (August 17th). The overall pattern is clear, with Europe (high latitude) having an earlier birth peak, and the Caribbean having a later peak. Both the Asian/Middle Eastern and Non-U.S. data are difficult to categorize as they span broad geographical ranges.

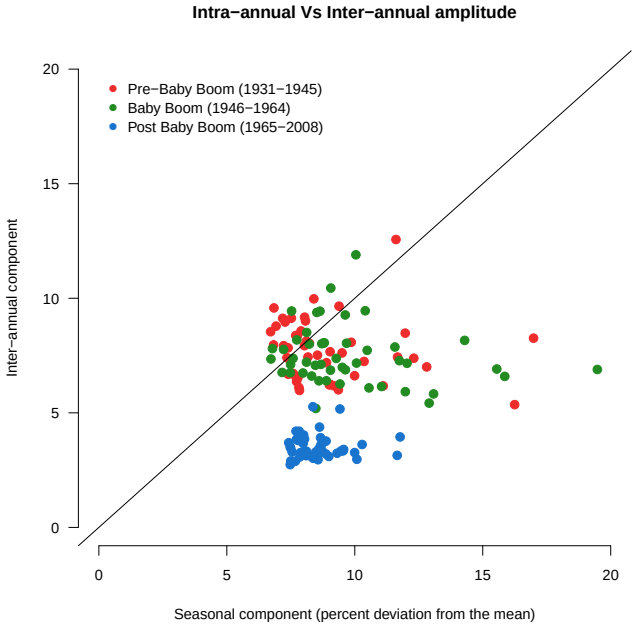
Seasonal birth amplitude. In the U.S., the largest seasonal amplitude in the pre-baby boom era was 20% from the mean (Louisiana, 1945), and the minimum was 4.2% from the mean in 1931 South Dakota (Fig A.4 top panel). These values increased during the baby boom era with a maximum 21.3% variation from the mean (Louisiana, 1954), and a minimum



Supplementary Figure A.5: Seasonal (intra-annual) amplitude of northern hemispheric data plotted vs latitude.

of 5.5% (Connecticut, 1957) (Fig A.4 middle). In the modern era, the maximum variation from the mean dropped to 17.4% (Louisiana, 1965) with the minimum staying relatively similar 5.4% (Delaware, 2004) (Fig A.4 bottom). The mean percent deviation from the mean in all states during the pre-baby boom era was 9.0%, 9.8% during the baby boom, and 8.5% in the modern era.

As shown in the main text Figure 1.2, we observed a latitudinal gradient in the seasonal birth amplitude in each era in the U.S. However, the latitudinal gradient in birth amplitude was not observed outside of the U.S. (Fig A.5). European seasonal amplitudes tended to be low, with a mean of 10.3%. Non-U.S. Americas had an approx. 9.8% amplitude, Asian/Middle Eastern countries having approx. 12.6%, and Caribbean countries approx. 17.2% (Fig A.5). However, due to the high variation in countries grouped into regions it is difficult to draw any conclusions from this.



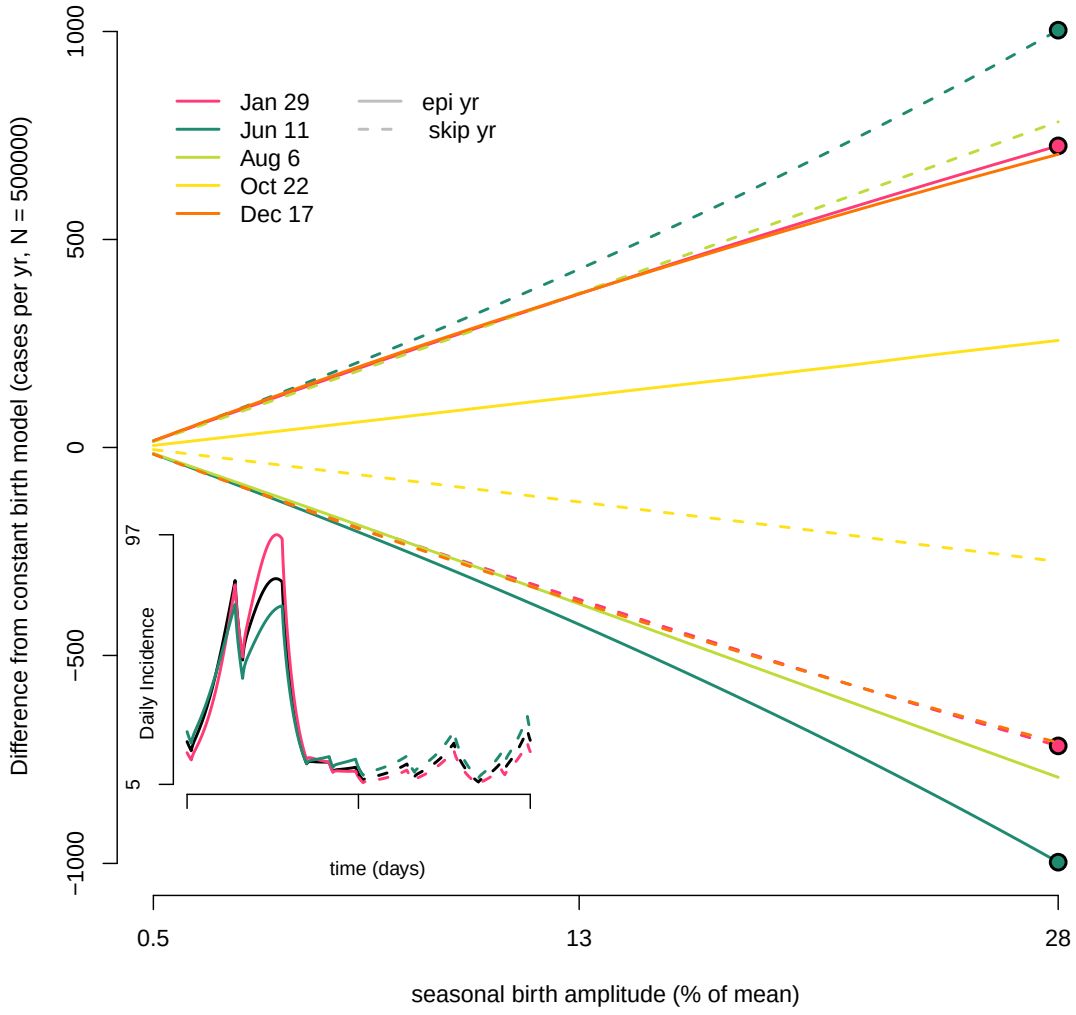
Supplementary Figure A.6**:** The seasonal (intra-annual) amplitude vs the annual (inter-annual) amplitude. In both the pre-baby boom and baby boom most states had a stronger seasonal component, whereas in the present period all states have a stronger seasonal component.

Intra- vs. inter-annual variation in birth rates. We examined the seasonal (intra-annual) variation in births and compare that to the inter-annual variation (percent change from one year to the next) for each state in every era. We found that the seasonal (intra-annual) variation is generally larger (Fig A.6), with the intra-annual variation in the modern era 2-3 times larger than the inter-annual value.

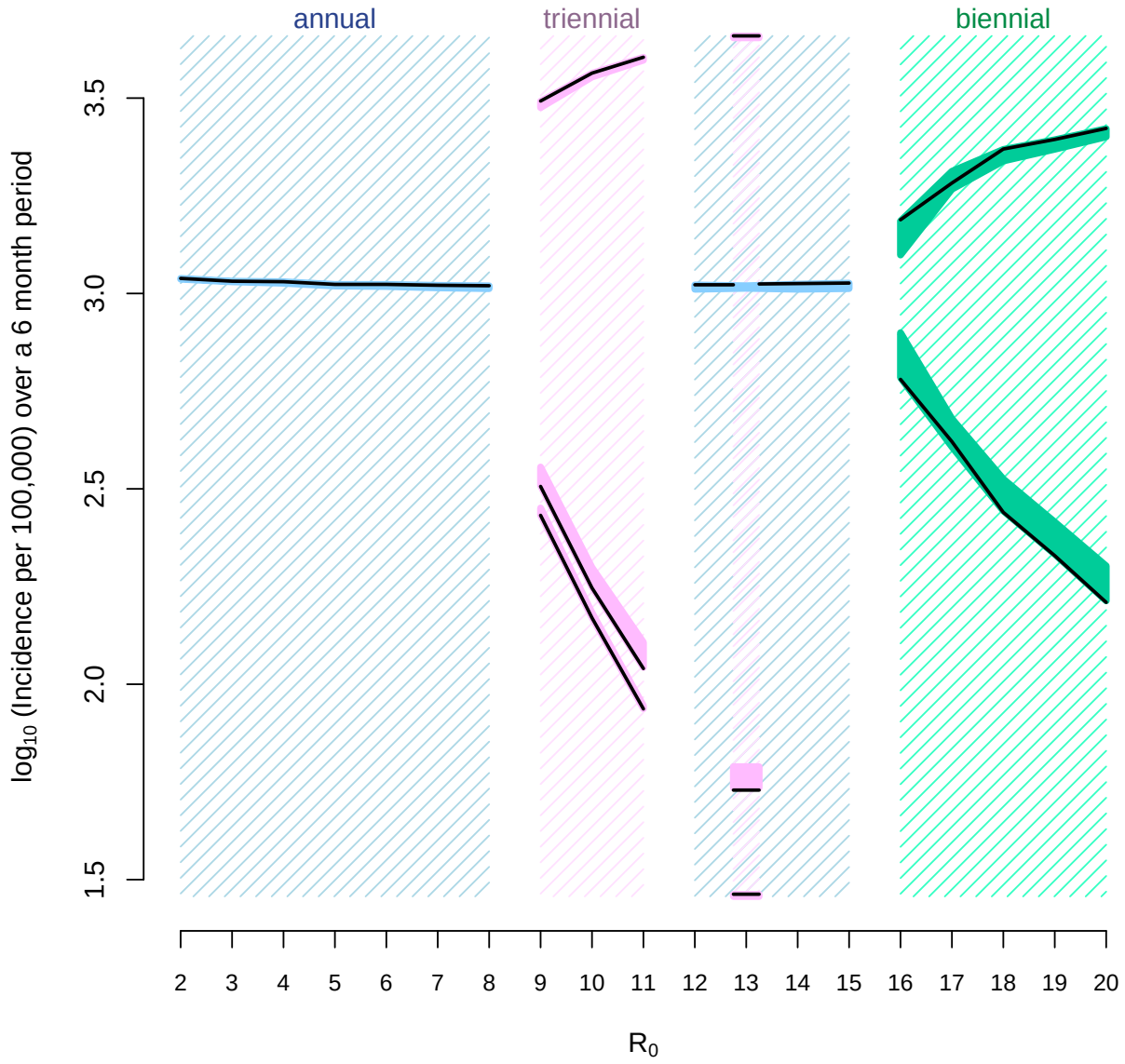
A.1.5 Simulation study

Fig A.7 shows the effect of increasing birth amplitude on measles incidence with the birth peak occurring at various times of the year. The difference in incidence resulting from measles models with either constant births or seasonal births increases with birth amplitude. However, depending on the phase relationship between peak susceptible recruitment (i.e. the birth peak) and the peak in seasonal transmission, birth seasonality can have the effect of either enhancing or dampening the epidemic year incidence.

Fig A.8 shows the effect of birth seasonality for various values of $\mathcal{R}_0$. Birth seasonality in this range of amplitude, $< 28\%$, has a pronounced effect on incidence when epidemics are biennial or triennial, as opposed to annual.



Supplementary Figure A.7: Effect of increasing seasonal birth amplitude on measles incidence. The main graph show the change in epidemic and skip year incidence as a function of birth amplitude for 5 different phases of births seasonality. Phases were set such that the birth peak occurred in either Jan, Jun, Aug, Oct, or Dec. The turquoise and the fuschia points in the main graph correspond to the turquoise and fuchsia time series in the inset. Here $\mathcal{R}_0 = 17$ and the birth amplitude ranged from 0-28%, all parameters are those from Table A.2.



Supplementary Figure A.8: The effect of birth seasonality on diseases with varying basic reproductive numbers ($\mathcal{R}_0$). This is a bifurcation diagram our SEIR model with varying $\mathcal{R}_0$ and varying seasonal birth amplitude. The black lines are the incidence for the model with no birth seasonality (i.e. birth amplitude = 0%). The solid shaded intervals indicate the regions containing the incidence of the model with seasonal births, where the birth peak is in early June and the amplitude ranges from 1.4 - 27.7%. Birth seasonality in this range of amplitude has a pronounced effect on incidence when epidemics are biennial or triennial, as opposed to annual. Here $\mathcal{R}_0 \in [2:20]$, and $S_0 = 1/\mathcal{R}_0$, otherwise all other parameters are those in Table A.2.

A.1.5.1 Inference Study Using New York City Measles Data

Maximum likelihood parameter estimates were obtained for each of our four measles models: seasonal births with a lag of 3 months from birth to susceptible to account for maternal antibodies, seasonal births with a lag of 6 months, seasonal births with a lag of 9 months, and constant births throughout the year.

Parameter	*Model Seas-3*	*Model Seas-6*	*Model Seas-9*	*Model NoSeas*
LogLikelihood	-1080.99 ± 0.20	-1080.86 ± 0.12	-1081.14 ± 0.12	-1080.95 ± 0.14
$\mathcal{R}_0$	19.3	19.5	20.3	19.7
β_{coef_1}	55.4	55.6	58.9	56.5
β_{coef_2}	45.8	47.2	48.6	46.8
β_{coef_3}	45.6	45.5	48.5	47.8
β_{coef_4}	44.0	44.2	46.8	45.3
β_{coef_5}	17.5	18.5	18.7	18.3
β_{coef_6}	28.7	27.6	28.3	27.7
ψ	3.4×10^{-4}	3.8×10^{-4}	4.2×10^{-4}	4.1×10^{-4}
β_{sd}	0.121	0.118	0.121	0.124
dt	1/2	1/2	1/2	1/2
δ	1/600	1/600	1/600	1/600
ρ	0.238	0.236	0.239	0.238
τ	3.8×10^{-2}	3.8×10^{-2}	3.9×10^{-2}	3.8×10^{-2}
S_0	819214	947306	937014	1719583
I_0	16602	22758	17265	27629
R_0	6238289	6104041	6119826	5326893

Table A.4: Maximum likelihood parameter estimates for each model. Rates are given in units of $month^{-1}$. All parameters were estimated using MIF with the exception of δ and dt, which were fixed. Note, $\mathcal{R}_0$ is the basic reproductive number and R_0 is the initial number of recovered individuals.

Country	Years	No. Years	Latitude	Mean peak month	Significant	Group	Amplitude
Albania	1981-2007	27	41.17	6.15	Yes	Europe	19.37%
Algeria	1998-2002	5	33.10	8.80	No	Africa	10.11%
American Samoa	1984-1988	5	-14.30	5.20	Yes	Asia	14.77%
American Samoa	1996-2006	11	-14.30	4.45	Yes	Asia	14.81%
Antigua and Barbuda	1979-1986	8	17.05	10.50	Yes	Caribbean	16.62%
Armenia	1987-1999	13	40.29	7.31	No	Europe	15.81%
Aruba	2002-2007	6	12.52	9.17	Yes	Caribbean	14.60%
Australia	1973-2008	36	-32.35	6.56	No	Asia	6.06%
Austria	1973-2011	38	47.77	6.50	Yes	Europe	6.33%
Azerbaijan	1992-2004	13	40.18	2.31	No	Asia	13.81%
Bahamas	1972-1979	8	24.32	10.50	Yes	Caribbean	13.95%
Bahrain	1975-1985	11	26.03	6.09	No	Asia	15.33%
Bahrain	1986-2002	17	26.03	10.35	Yes	Asia	13.26%
Barbados	1969-1976	7	13.16	7.86	No	Caribbean	23.73%
Barbados	1982-1991	10	13.16	11.00	Yes	Caribbean	19.78%
Belarus	1987-1999	13	53.33	5.31	Yes	Europe	11.52%
Belgium	1971-1995	25	50.84	6.08	Yes	Europe	7.18%
Belgium	1998-2008	11	50.84	7.64	No	Europe	6.54%
Belgium-Bruxelles	1998-2008	11	50.84	8.09	No	Europe	7.35%
Belgium-Flamande	1998-2008	11	50.84	6.73	No	Europe	6.54%
Belgium-Anvers	1998-2008	11	50.84	6.64	No	Europe	7.33%
Belgium-Limbourg	1998-2008	11	50.84	6.45	No	Europe	7.96%
Belgium-Flandreorientale	1998-2008	11	50.84	7.18	No	Europe	6.25%
Belgium-Brabantflamand	1998-2008	11	50.84	6.91	Yes	Europe	7.47%
Belgium-Flandreoccidentale	1998-2008	11	50.84	5.64	No	Europe	7.17%
Belgium-Wallonne	1998-2008	11	50.84	8.55	Yes	Europe	6.73%
Belgium-Germanophone	1998-2008	11	50.84	5.18	No	Europe	15.34%
Belgium-Brabantwallon	1998-2008	11	50.84	6.91	Yes	Europe	7.86%
Belgium-Hainaut	1998-2008	11	50.84	9.09	Yes	Europe	6.20%
Belgium-Liege	1998-2008	11	50.84	8.27	No	Europe	7.08%
Belgium-Luxembourg	1998-2008	11	50.84	8.55	Yes	Europe	8.38%
Belgium-Namur	1998-2008	11	50.84	8.91	Yes	Europe	7.27%
Bermuda	1984-1991	8	32.30	9.38	Yes	Caribbean	17.48%
Bermuda	1995-2001	7	32.30	9.00	Yes	Caribbean	16.45%
BVI	1980-1986	7	18.43	10.43	Yes	Caribbean	23.23%
Brunei Darussalam	1972-1976	5	4.82	8.00	No	Asia	12.46%
Brunei Darussalam	1980-1992	13	4.82	9.31	No	Asia	10.75%
Brunei Darussalam	1996-2002	7	4.82	9.71	Yes	Asia	8.26%
Bulgaria	1973-1978	6	42.75	4.83	No	Europe	8.01%
Bulgaria	1980-1990	11	42.75	6.18	Yes	Europe	8.87%
Canada	1973-1990	18	56.76	6.06	Yes	Americas	7.49%
Canada	1992-1997	6	56.76	6.00	Yes	Americas	8.82%
Canada	1999-2008	10	56.76	6.80	Yes	Americas	8.03%
Cape Verde	1968-1975	8	15.11	3.63	No	Africa	12.94%
Cape Verde	1980-1985	6	15.11	11.50	No	Africa	17.40%
Caymen Islands	1986-1995	10	5.36	10.20	Yes	Caribbean	24.20%
Chile	1967-2008	42	-35.12	9.29	Yes	Americas	7.83%
China-Hong Kong	1973-1977	5	22.30	10.00	Yes	Asia	14.66%
China-Hong Kong	1979-2009	31	22.30	10.00	Yes	Asia	14.85%
China-Macao	1971-1975	5	22.17	9.60	Yes	Asia	16.52%
China-Macao	1984-1989	6	22.17	9.83	Yes	Asia	17.67%
China-Macao	1991-2010	20	22.17	9.70	Yes	Asia	16.06%
Cook Islands	1983-1988	6	-21.20	6.50	No	Asia	19.36%
Costa Rica	1987-1991	5	9.92	10.40	Yes	Americas	9.59%
Costa Rica	2003-2010	8	9.92	9.88	Yes	Americas	10.90%
Croatia	1988-2004	17	45.32	8.18	No	Europe	7.36%
Cuba	1976-1988	13	22.03	10.00	Yes	Caribbean	16.68%
Cuba	1990-2009	20	22.03	10.20	Yes	Caribbean	19.93%
Cyprus	1973-2009	37	34.97	8.49	Yes	Asia	11.88%
Czech Republic	1991-2010	14	49.82	5.36	Yes	Europe	10.15%
Denmark	1972-2005	34	55.85	6.03	Yes	Europe	9.57%
Egypt	1972-1982	11	28.80	7.00	Yes	Africa	30.08%
Egypt	1987-1999	13	28.80	10.85	No	Africa	35.30%
Egypt	2003-2009	7	28.80	9.29	No	Africa	12.60%
El Salvador	1973-2007	35	13.72	10.46	Yes	Americas	12.48%
Estonia	1989-1997	9	58.96	4.89	Yes	Europe	12.05%
Estonia	1999-2005	6	58.96	5.67	Yes	Europe	8.67%
Estonia	2007-2011	5	58.96	6.40	Yes	Europe	9.82%
Faeroe Islands	1972-1987	16	62.09	7.25	No	Europe	11.90%

Table A.5: Data used in national-level analyses of birth seasonality. Significance refers to the annual period. Mean birth peak timing and amplitude were estimated from the data.

Country	Years	No. Years	Latitude	Mean peak month	Significant	Group	Amplitude
Finland	1972-1988	17	61.76	4.94	Yes	Europe	10.55%
Finland	1994-2004	14	61.76	5.93	Yes	Europe	7.86%
France	1974-1989	16	47.14	5.75	Yes	Europe	10.02%
France	1991-1997	7	47.14	7.00	Yes	Europe	8.69%
France	1999-2004	6	47.14	7.83	Yes	Europe	8.06%
French Guiana	1977-1986	10	5.09	10.10	No	Americas	13.00%
French Guiana	1997-2003	7	5.09	10.71	No	Americas	12.35%
French Polynesia	1985-1992	8	-17.53	4.63	Yes	Asia	9.81%
Germany	1991-1997	7	50.86	7.14	Yes	Europe	7.72%
Germany	2004-2010	7	50.86	8.00	Yes	Europe	9.30%
Gibraltar	1973-1988	16	36.14	7.50	No	Europe	14.75%
Gibraltar	2002-2008	7	36.14	10.00	No	Europe	18.51%
Greece	1974-1985	12	38.69	6.00	Yes	Europe	10.53%
Greece	1990-2001	12	38.69	7.67	Yes	Europe	11.05%
Greenland	1972-1987	16	67.10	7.63	No	Europe	12.57%
Greenland	1992-2010	19	67.10	6.95	No	Europe	11.90%
Guadeloupe	1975-1980	6	16.27	9.33	No	Caribbean	10.58%
Guadeloupe	1982-1986	5	16.27	11.00	Yes	Caribbean	13.27%
Guadeloupe	1997-2003	7	16.27	10.57	Yes	Caribbean	16.23%
Guam	1973-1982	10	13.45	10.10	No	Asia	8.43%
Guam	1988-1992	5	13.45	9.80	No	Asia	9.40%
Guatemala	1972-1979	8	14.72	7.38	No	Americas	7.12%
Guatemala	1981-1999	19	14.72	10.79	No	Americas	9.54%
Guernsey	1973-1979	7	49.48	2.71	No	Europe	12.37%
Guernsey	1992-2000	9	49.48	8.22	No	Europe	11.49%
Guyana	1967-1971	5	6.35	9.60	Yes	Americas	10.04%
Hungary	1973-1992	20	47.29	6.40	Yes	Europe	8.74%
Hungary	1994-2004	11	47.29	7.73	No	Europe	7.90%
Iceland	1972-1980	9	64.37	5.67	Yes	Europe	10.39%
Iceland	1982-2004	22	64.37	6.77	Yes	Europe	10.23%
Iran	1999-2004	6	33.68	2.83	No	Asia	7.95%
Ireland	1972-2004	33	53.11	5.94	Yes	Europe	8.33%
Isle of Man	1973-1988	16	54.19	7.38	No	Europe	12.52%
Israel	1973-1981	9	31.78	9.33	Yes	Asia	9.58%
Israel	1983-1988	6	31.78	9.00	Yes	Asia	9.60%
Israel	1990-2009	20	31.78	9.60	Yes	Asia	8.97%
Italy	1970-1985	15	42.87	3.20	Yes	Europe	10.01%
Italy	1988-2009	22	42.87	8.14	Yes	Europe	11.01%
Jamaica	1999-2007	9	18.13	10.78	Yes	Caribbean	18.60%
Japan	1972-1992	21	35.41	7.48	Yes	Asia	7.59%
Japan	1994-2010	17	35.41	8.24	Yes	Asia	7.01%
Jersey	1973-1989	17	49.22	6.59	No	Europe	11.64%
Kazakhstan	1987-2008	22	43.35	6.27	No	Asia	10.11%
Korea Republic	1996-2009	24	36.47	4.83	No	Asia	7.75%
Kuwait	1975-1987	13	29.33	10.23	No	Asia	9.06%
Kuwait	1991-2008	18	29.33	9.44	No	Asia	10.68%
Kyrgyzstan	1985-2004	20	41.76	4.85	No	Asia	11.06%
Kyrgyzstan	2005-2009	5	41.76	7.40	Yes	Asia	12.02%
Latvia	1989-2005	17	56.83	5.12	Yes	Europe	11.12%
Lebanon	2003-2010	8	33.93	8.88	No	Asia	11.44%
Libya	1972-1981	10	29.96	1.00	Yes	Africa	12.01%
Libya	1989-1996	8	29.96	7.00	No	Africa	10.04%
Liechtenstein	1978-1987	10	47.15	6.80	No	Europe	13.16%
Liechtenstein	2000-2005	6	47.15	5.83	No	Europe	18.31%
Lithuania	1987-2011	25	55.22	5.72	Yes	Europe	13.11%
Luxembourg	1973-1989	17	49.64	6.47	No	Europe	9.22%
Luxembourg	1998-2010	13	49.64	6.23	Yes	Europe	10.08%
Malaysia	1994-2008	15	4.19	7.67	No	Asia	8.71%
Maldives	1978-2009	32	4.17	7.31	No	Europe	13.36%
Moldova Republic	1987-1992	6	47.17	5.67	No	Europe	7.60%
Moldova Republic	1998-2010	13	47.17	8.62	No	Europe	8.86%
Malta	1973-1988	16	35.90	9.06	Yes	Europe	11.41%
Malta	1992-2004	13	35.90	9.85	No	Europe	9.17%
Martinique	1975-1992	18	14.67	10.94	Yes	Caribbean	15.60%
Martinique	1998-2003	6	14.67	11.00	Yes	Caribbean	18.29%
Mauritius	1994-2010	17	-20.16	5.76	Yes	Africa	11.34%
Mongolia	1994-2003	10	47.77	5.10	No	Asia	8.24%
Netherlands	1973-1988	16	52.07	6.06	Yes	Europe	7.46%
Netherlands	1990-2010	21	52.07	7.62	Yes	Europe	6.63%
New Caledonia	1970-1977	8	-21.50	5.63	Yes	Asia	11.36%
New Caledonia	1982-2007	26	-21.50	5.54	Yes	Asia	12.12%
New Zealand	1972-2009	39	-41.44	7.46	Yes	Asia	23.52%
Niue	1982-1987	6	-19.06	7.83	No	Asia	45.89%
Norway	1976-1987	12	61.13	5.00	Yes	Europe	11.51%
Norway	1995-2004	10	61.13	5.50	Yes	Europe	9.64%
Occ. Palestinian Territory	1997-2007	11	31.88	10.64	Yes	Asia	10.81%

Country	Years	No. Years	Latitude	Mean birth month	Significant	Group	Amplitude
Palau	1997-2003	7	7.35	7.29	No	Asia	16.08%
Panama	1973-1999	27	8.75	10.78	Yes	Americas	8.35%
Panama	2005-2009	5	8.75	10.00	Yes	Americas	10.60%
Phillipines	1997-2007	11	11.87	10.00	Yes	Asia	13.93%
Poland	1978-2005	28	51.71	5.43	Yes	Europe	9.95%
Portugal	1973-1993	21	39.75	6.67	No	Europe	8.42%
Portugal	1999-2009	11	39.75	8.64	Yes	Europe	9.11%
Puerto Rico	1967-1985	19	18.26	10.05	Yes	Caribbean	11.33%
Puerto Rico	1987-1992	6	18.26	10.00	Yes	Caribbean	12.49%
Puerto Rico	1996-2000	5	18.26	10.00	Yes	Caribbean	11.71%
Puerto Rico	2002-2008	7	18.26	10.00	Yes	Caribbean	11.12%
Qatar	1985-1990	6	25.30	9.50	No	Asia	11.93%
Qatar	1999-2009	11	25.30	10.18	No	Asia	9.78%
Reunion	1977-1986	10	-21.11	6.40	Yes	Africa	7.53%
Reunion	1998-2003	6	-21.11	4.33	No	Africa	6.77%
Romania	1986-1992	7	45.69	6.29	Yes	Europe	9.60%
Romania	1994-2010	16	45.69	6.25	Yes	Europe	10.66%
Saint Helena ex dep	1981-1986	6	-15.93	6.83	No	Africa	31.44%
Saint Lucia	1976-1986	11	13.90	10.82	Yes	Caribbean	15.27%
Saint Lucia	1994-2002	9	13.90	11.00	Yes	Caribbean	20.24%
Saint Vincent and the Grenadines	1992-2005	14	13.20	11.07	Yes	Caribbean	23.24%
San Marino	1973-1978	6	43.94	7.33	No	Europe	21.85%
San Marino	1984-1989	6	43.94	9.50	No	Europe	29.63%
San Marino	2000-2004	5	43.94	7.60	No	Europe	21.12%
Sao Tome and Principe	1967-1971	5	0.32	4.20	No	Africa	13.27%
Seychelles	1973-1993	21	-4.63	6.24	Yes	Africa	14.81%
Seychelles	1995-2010	16	-4.63	5.94	Yes	Africa	18.63%
Singapore	1973-2009	36	1.37	9.36	Yes	Asia	12.77%
Slovakia	1988-1995	8	48.66	5.75	Yes	Europe	10.10%
Slovakia	1998-2002	5	48.66	6.20	Yes	Europe	10.16%
Slovenia	1988-1996	9	46.17	6.67	No	Europe	7.17%
Slovenia	1998-2005	8	46.17	7.38	Yes	Europe	7.97%
Spain	1970-1985	16	39.72	6.38	Yes	Europe	8.40%
Spain	1991-2005	15	39.72	7.87	No	Europe	6.53%
Sri Lanka	1973-1986	14	7.57	9.14	No	Asia	7.64%
Sri Lanka	2005-2010	6	7.57	10.00	Yes	Asia	8.98%
Suriname	1989-2009	21	5.06	11.05	Yes	Americas	15.05%
Sweden	1973-1990	18	58.91	4.56	Yes	Europe	15.25%
Sweden	1992-2002	11	58.91	4.91	Yes	Europe	13.90%
Sweden	2004-2010	7	58.91	5.71	Yes	Europe	11.19%
Switzerland	1973-1982	10	47.03	5.00	Yes	Europe	10.33%
Switzerland	1984-1990	7	47.03	5.86	Yes	Europe	6.26%
Switzerland	1998-2002	5	47.03	6.80	No	Europe	5.78%
Tajikstan	1989-1994	6	38.70	4.00	Yes	Asia	23.24%
Macedonia	1989-1993	5	41.74	8.20	No	Europe	9.54%
Macedonia	1999-2010	12	41.74	7.83	Yes	Europe	10.20%
Tonga	1993-2000	8	-19.70	4.25	Yes	Asia	16.00%
Trinidad and Tobago	1972-1995	24	10.55	10.83	Yes	Caribbean	14.04%
Tunisia	1971-1976	6	35.42	2.83	Yes	Africa	16.04%
Tunisia	1978-1982	5	35.42	4.00	Yes	Africa	13.83%
Tunisia	1985-1995	11	35.42	5.18	No	Africa	13.40%
Tunisia	2001-2007	7	35.42	7.00	Yes	Africa	16.75%
Turks and Caicos Islands	1997-2005	9	21.51	9.56	No	Caribbean	25.43%
Ukraine	1980-1986	7	48.81	6.00	Yes	Europe	9.52%
Ukraine	1989-1996	8	48.81	5.75	Yes	Europe	9.77%
Ukraine	2003-2010	8	48.81	8.25	Yes	Europe	10.28%
UK + NI	1982-1988	7	52.75	6.71	Yes	Europe	6.85%
UK + NI	1990-2006	15	52.75	7.47	Yes	Europe	6.61%
USA	1969-1975	7	40.42	8.86	Yes	Americas	7.21%
USA	1978-2006	29	40.42	8.07	Yes	Americas	7.95%
USVI	1969-1973	5	18.33	10.80	Yes	Caribbean	19.68%
USVI	1980-1997	18	18.33	10.61	Yes	Caribbean	16.64%
Uruguay	1980-1988	9	-33.00	8.33	No	Americas	5.55%
Uzbekistan	1993-1997	5	40.68	5.80	No	Asia	8.36%
Venezuela	1972-2001	30	9.39	9.40	Yes	Americas	9.41%
Wallis and Futuna Islands	1973-1978	6	-13.30	4.00	No	Asia	10.72%

APPENDIX B

Supporting Information: Digital Epidemiology Reveals Global Childhood Disease Seasonality and the Effects of Immunization

B.1 Digital Epidemiology Reveals Global Childhood Disease Seasonality and the Effects of Immunization: Supporting Information

B.1.1 Acquisition and Analyses of Google Trends Data

Google Trends is a publically available data service provided by Google Inc [86] that allows internet users to view and download global information on internet search behaviour. Google Trends represent the relative number of searches for a specific key word, or combination of search terms. The numbers are standardized within each country such that the values range from 0 to 100. A search volume of 0 is assigned, by Google Trends, to weeks/months with a minimal amount of searches. Google Trends provides time-series of these abundance data, but gives no explanation of how the relative abundances were calculated. In addition to relative abundances, for each country, the downloaded Google Trends csv file provided a

list of 'Top-searches for [the language-specific search term]'. Each top-search list included the context of the search term, each with an integer value of its relative abundance from $0 - 100$. The top searches are listed in descending order. Top searches were informative for determining how to interpret Google query data. Because searches for "shingles" on Google Trends could be referring to "the disease shingles" or "roofing shingles", the top search list could be used to distinguish among the search context when the single-word search term has an ambiguous context. The top-search list is therefore invaluable for ensuring data for specific search terms are being properly interpreted.

We used Google Trends data to evaluate childhood disease information seeking behaviour, and obtained country-specific data from the start of Google Trends, January 2004 to July 2015. We downloaded Google Trends data from 36 countries with high volume searches for chicken pox worldwide (Table B.3). For each country the data were subset within the range that included consecutive weeks with > 0 search volume.

In order to relate Google Trends data to the dynamics of chicken pox (or other diseases of interest), care must be taken to select appropriate search terms. Chicken pox is the classical manifestation of disease, and therefore, language-specific queries of "chicken pox" are a straightforward choice for data-mining. In contrast, infections with generic symptoms, such as fever and diarrhea, could arise from many other diseases, making it difficult to identify appropriate queries. In either case, search terms vary subtly from country to country. For instance, in the US "chickenpox" is typically written as a single word, whereas in the U.K., people refer to "chicken pox" as two words; in Spanish, chicken pox is referred to as "varicela", with a single "l". We accounted for this variation among countries by careful choice of search terms, and downloaded the data for 36 countries using 21 language-specific queries of chicken pox. The csv file downloaded for each country included the top-searches for the country-specific search term. As an example, 48 top-searches were provided for Argentina; the top 5 searches–and their integer value, which we refer to as relative abundance–were:

Top searches for ''varicela"

- la varicela, 100
- varicela sintomas, 35
- varicela vacuna, 30
- varicela contagio, 25
- sintomas de varicela, 20

We evaluated (and translated when needed) the top-search list provided for the US, Australia, Mexico, and Thailand, four of the countries for which we had data on reported cases, and these countries are highly variable in their varicella vaccination policy. Although the top-search lists had no metadata provided by Google, the clear difference between top searches among the four countries indicated that the top searches contained valuable epidemiological information. We therefore decided to systematically evaluate the top-search lists for epidemiological information. The top-search lists closely matched expectation based on each countries vaccination policy.

The 36 countries in our study differed in their VZV vaccination history and current policy (Table B.3). The first VZV vaccine was licensed in the US in 1995 and was incorporated into the measles, mumps, rubella, and varicella vaccine (MMRV) in 2005 [40, 188]. These vaccinations were only implemented by a few countries, and at different times, which led us to expect country-specific differences in chicken pox query motivation. Therefore, for a subset of countries, we evaluated the context of Google Trends searches by categorizing searches based on whether they queried chicken pox as a disease, chicken pox vaccination, or other contexts (Fig B.7, Table B.2).

The significance of information seeking seasonality was tested using Morlet wavelet analyses for each country (Fig 2.1, B.5) [91, 210]. Both wavelet analyses and General Additive Models (GAMs) are powerful methods for detecting periodicity in time series. The Google

data for all countries other than Estonia, and the Czech Republic (which had monthly data) were examined for annual (52 week) periodicity using colored noise. First, we measured wavelet significance at the annual band. Second, of the 33 countries, 18 countries (Colombia, the UK, the US, Argentina, Brazil, Denmark, France, Hungary, India, Ireland, Italy, Mexico, The Netherlands, Poland, Romania, South Africa, Sweden, and Vietnam) had a significance band within the entire cone of influence of the annual band (time periods where you can test for significance). Seven countries (Australia, Finland, New Zealand, Philippines, Portugal, Spain, and Thailand) had significant power at the annual period for $> 50\%$ of their time series, with high-power periodicity (i.e., red banding) at the other non-significant time points. China had $\sim 40\%$ of its time series significant at the annual period, with high power annual periodicity at all other time points. In Germany the significance was lost about halfway through the time series, and the power of the annual period diminished. Chile and Japan did not have significant annual periodicity, although Japan did have high power at the annual period throughout its time series. The remaining four countries (Russia, Iran, Austria, and Venezuela) had time series that were too short to test for annual significance using wavelet analyses (i.e., they did not have 3+ years of data).

In order to characterize the seasonal shape of chicken pox information seeking, we used GAMs, which is a nonparametric extension of generalized linear models (GLMs) in which the linear predictor depends on smooth functions of predictor variables. We used the restricted maximum likelihood (REML) method with the linear predictor being the detrended Google data, while the predictor variables we tested included week number for seasonality and time for the overall trend. REML and maximum likelihood methods are less prone to local minima than the other criteria, and usually preferable. We used a GAM, rather than a generalized linear model or other model, because GAMs are flexible when fitting smooth curves to ecological data, and typically allow a better fit for time series than GLMs. A GAM was fit for all countries with significant annual wavelet periodicity, except Estonia and the Czech Republic, because their data were monthly and not weekly.

B.1.2 Validating Information Seeking in Chicken Pox

Once we characterized the seasonal variation in Google queries for chicken pox, we tested whether variation in information seeking behaviour paralleled variation in chicken pox incidence. The Google Trends data required validation against epidemiological data because variation in information seeking could be driven by cultural events rather than changes in disease incidence. For example, in the US, information seeking using the query "breast cancer" has a sharp seasonal peak in Google Trends each year in October, reflecting October as breast cancer awareness month, rather than a month with elevated incidence. For infectious diseases, the covariation between information seeking behaviour and clinical cases can be established using reported cases. We validated our Google Trends data, and evaluated search term context, using records of clinical cases from countries with active chicken pox surveillance (Fig 2.2).

We obtained data from five countries that report chicken pox: Australia, Thailand, and Estonia – which had monthly reported cases – and Mexico and the US, which reported cases weekly. The data from Australia were collected from the Australian National Disease Surveillance System [8], digitized on May 1st, 2015. Thailand chicken pox case data were downloaded from the Bureau of Epidemiology, Department of Disease Control, MoPH, Thailand [202] on April 12, 2015. The data from Mexico were digitized from the weekly disease surveillance reports of the Mexico General Directorate of Epidemiology, first published in [219] and provided to us by the authors. The US data, both historical (Fig B.8) and modern, were obtained from the Project TYCHO database [218]. Data from Estonia were provided by the Estonian Health Board, Department of Communicable Disease Surveillance and Control [74]. Clinical data from these five countries span different time periods, but each overlapped with the Google Trends data for 4+ years. Clinical data spanned Jan 1995 – Feb 2011 in Mexico, Jan 2003–Dec 2014 in Thailand, Jan 2006–Feb 2015 in Australia, Jan 2006–Aug 2013 in the US, and Jan 1999–Dec 2014 in Estonia.

B.1.3 Forecasting

In order to determine if Google Trends data could be used to predict the magnitude and timing of chicken pox outbreaks, we built forecasting models. The models predicted the force of infection, which we defined as the monthly per capita rate at which children age 0–14 years are infected. We refer to this parameter as the force of infection, which is typically defined as the rate of infection per capita susceptible individual, because we are assuming all susceptible individuals are contained within the the 0–14 year age class, and therefore the number of 0–14 year olds is a surrogate for the number susceptible. The forecasting models containing Google data use Google data from the previous time intervals, $t-1$ and $t-2$, to predict the number of chicken pox cases at the current time interval, t. The models were fit to data from two countries that actively report chicken pox cases, one with active immunization (Australia) and one lacking immunization (Thailand). To determine whether Google Trends, T, was able to forecast the magnitude and timing of chicken pox outbreaks, we built and fitted multiple statistical models to forecast chicken pox case data. The correlation between chicken pox information seeking and chicken pox cases was weaker in Australia compared to Thailand ($R^2 = 0.26$ and 0.81, respectively). We therefore used the case data from Australia to test the power of the forecasting models, since Australia poses a more challenging forecasting problem.

The Google Trends data are weekly, whereas both Australia and Thailand reported chicken pox on a monthly basis. Thus, we forecast on a monthly basis and converted the weekly Google Trends data to monthly values. To do this, we repeated the weekly values at daily intervals (Google Trends data are relative search values and not absolute number of searches). We then assigned the daily values to their appropriate month of the year. For each month, we then found the mean of the daily values, which resulted in the values used for forecasting.

The null and four of our eight forecasting models included a cosine function to help predict chicken pox outbreaks. We discovered that the cosine function is required because it

imposes cyclicity on the outbreaks, acting as a proxy for cyclical changes in (1) the number of susceptible individuals in the population and/or (2) the transmission rate. Although the Google Trends data are cyclical, since the forecasting model predicts one-month-ahead, without the cosine function, the Google Trends data alone would be limited in ability to forecast directionality (i.e., to determine if cases are increasing or decreasing). Including a cosine function with a period of 12-months allowed us to overcome this limitation. We tested eight different forecasting models, all slight variations of each other, and compared the model results to a null model that captured the annual seasonal patterns of chicken pox incidence (Table B.1). It is unknown how the Google data scaled to chicken pox data. We therefore estimated scaling parameters in the various models (i.e. α, β_1, β_2, and β_3).

We evaluated the epidemiological information contained in Google Trends by comparing the Google Trends models with a seasonal null model that did not incorporate Google data (model B). The null model lacked information seeking in the force of infection λ_t. All models were fitted to the case data from a VZV-vaccinated population (Australia), which exhibited reduced seasonality. To estimate the number of symptomatic VZV infections each month, I_t, we used Google Trends data from the previous two months, T_{t-1} and T_{t-2}, where t is time in monthly time steps. The chicken pox process model with the best fit, tracked the force of infection, λ_t,

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right)\mathrm{T}_{t-1} + \beta_2|\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right]\epsilon_t. \tag{B.1}$$

The model contained environmental stochasticity, ϵ_t, which was drawn from a gamma distribution with a mean of 1 and variance θ. We estimated the following parameters for the Google model: the mean and the phase of the seasonality (β_1 and ω), a parameter scaling the Google Trends data (β_2), the baseline force of infection (β_3), the process noise dispersion parameter (θ), and the reporting dispersion parameter (τ) of a normal distribution, with a mean of 1, from which case reports were drawn. The parameters were estimated using maximum likelihood by iterated particle filtering (MIF) in the R-package pomp [111, 112].

In order to estimate the number of symptomatic VZV infections per month, we multiplied the force of infection, λ, with an estimate of the population aged 0–14 years [213], C,

$$I_t = \lambda_t \mathrm{C}. \tag{B.2}$$

We modeled the observation process, which represents the number of cases reported. To account for stochasticity in the reporting of symptomatic VZV infections, case reports were drawn from a normal distribution with a mean report rate, $\rho = 1$, and dispersion parameter (τ) which was estimated from the data.

$$\text{chickenpox}_t \sim \mathcal{N}(\rho I_t, \tau I_t). \tag{B.3}$$

We evaluated the epidemiological information contained in Google Trends by comparing the Google Trends model with a seasonal null model where the force of infection did not incorporate Google Trends data. The null model force of infection was modeled as:

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) + \beta_3\right] \epsilon_t. \tag{B.4}$$

To explore other model possibilities, we tested seven other models that included Google data. All model parameters listed are in reference to the best fit Google Model (model A). Of the additional seven Google models, the first lacked the β_2 parameter;

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1} + \beta_3\right] \epsilon_t. \tag{B.5}$$

The second model lacked the β_2 parameter, but included an additional Google Trends scaling parameter, α;

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1}^{\alpha} + \beta_3\right] \epsilon_t. \tag{B.6}$$

The third model contained the α parameter,

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1}^{\alpha} + \beta_2 |\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right] \epsilon_t. \tag{B.7}$$

The fourth model lacked the cosine function and β_2 parameter;

$$\lambda_t = \left[\beta_1(\mathrm{T}_{t-1}) + \beta_3\right] \epsilon_t. \tag{B.8}$$

The fifth model lacked the cosine function and β_2 parameter, but contained the α parameter;

$$\lambda_t = \left[\beta_1(\mathrm{T}_{t-1}^{\alpha}) + \beta_3\right] \epsilon_t. \tag{B.9}$$

The sixth model lacked the cosine function but included the α parameter;

$$\lambda_t = \left[\beta_1(\mathrm{T}_{t-1}^{\alpha}) + \beta_2 |\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right] \epsilon_t. \tag{B.10}$$

and the seventh model lacked the cosine function;

$$\lambda_t = \left[\beta_1(\mathrm{T}_{t-1}) + \beta_2 |\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right] \epsilon_t. \tag{B.11}$$

Results from all models are listed in Table B.1. Models that included the cosine function, including the null (i.e. model equations B.1, B.4, B.5, B.6, and B.7), fit better than those that did not have the cosine function (i.e. model equations B.8, B.9, B.10, and B.11). The best fit Google forecasting model estimated six parameters and had an AIC of 1120.9, while the null model, which lacked Google Trends data, had an AIC of 1148.9. The best fit forecasting model without the cosine function estimated 4 parameters and had an AIC of 1179.3.

To further examine the difference between the Google model and the Null model, we ran 10000 simulations using the maximum-likelihood parameter set for each the Google model and the null model for Australia. We first examined each of the model fits (Fig B.1) to the

chicken pox case data. Since both models were seasonally forced, they were both able to capture the seasonal timing of outbreaks. However, the Google Trends model was able to predict the interannual variation in outbreak size, while the null model could not because the cosine function did not change interannually (Fig B.1).

These results demonstrate that the Google Trends model was better able to capture the dynamics of chicken pox case data. The stochastic simulations showed more variation (larger standard deviation), and captured the data more often than the Null model. To visualize the relationship between the model simulations and the chicken pox case data, we plotted the mean predicted chicken pox cases (model results) against the actual cases for the Google Trends model and the Null model (Fig B.1). Finally, to get a better understanding of why the Google Trends model fit the chicken pox case data better than the Null model, we explored the distribution densities of the troughs of each model against the data for each year (Fig B.2). The Google Trends model achieved a better fit to chicken pox data (Fig B.2). While the Google Trends model best captured the actual troughs in 2012, 2013, and 2014, its density distribution was always closer to the actual number of cases in the trough month relative to the Null model. The trough in 2006 was difficult to characterize because the model was estimating initial conditions, which could explain why neither the Google Trends model nor the Null model were able to accurately forecast the number of cases in May, 2006.

B.1.4 Information Seeking in other Childhood Diseases

To evaluate whether our findings based on chicken pox were representative of infectious childhood diseases in general, we examined information seeking behaviour for other childhood diseases. We obtained country-specific Google Trends data from the US and Australia for "hand foot and mouth", "croup", and "fifth disease" [86].

Google queries of croup, fifth disease, and HFMD in the US and Australia displayed variation in search volume (1) within and between years for each disease, (2) among diseases, and (3) across geographic locations for a given disease. HFMD displayed seasonal variation

in the US and Australia (Fig B.3). HFMD is caused by enteroviruses, which are notorious for their increased summer transmission in temperate regions [159]. The peak in HFMD information seeking generally occurred between June and August in the US and Australia, and was relatively synchronized between these countries. Seasonal variation in HFMD has been documented in clinical case data with peaks in the US occurring from spring to fall [110], which is in keeping with the seasonal variation observed in information seeking. In contrast, in Australia, the concurrent seasonal peak coincided with the southern hemisphere winter. This unexpected timing requires further investigation.

Croup information seeking also displayed seasonality, but unlike HFMD seasonal information seeking, it was asynchronous between the US and Australia. In the US, croup information seeking seasonally peaked between October and November, at the onset of the northern hemisphere winter; whereas in Australia, croup information seeking peaked from May–July, at the onset of the Australian winter. Croup is caused by Human parainfluenza viruses (HPIVs). HPIV-1 and HPIV-2, which cause croup in children, circulate in autumn, suggesting that the seasonality of croup information seeking in the US and Australia follow the seasonal circulation in HPIV-1 and 2.

Information seeking regarding fifth disease, which is caused by parvovirus B19, was highly seasonal in the US (Fig B.3). The seasonal peak in fifth disease information seeking showed a distinct trough from August–October and peaked from April–May, roughly coinciding with the seasonal peak of clinically diagnosed fifth disease in late winter and early spring [4, 161]. The search volume of fifth disease was not sufficient outside North America for geographic comparison.

These preliminary examples further emphasize the untapped potential of analyzing information seeking behaviour of childhood infectious diseases. Digitally detecting pathogen-specific, large-scale spatio-temporal patterns can provide clues for identifying environmental and physiological drivers of the dynamics of these diseases.

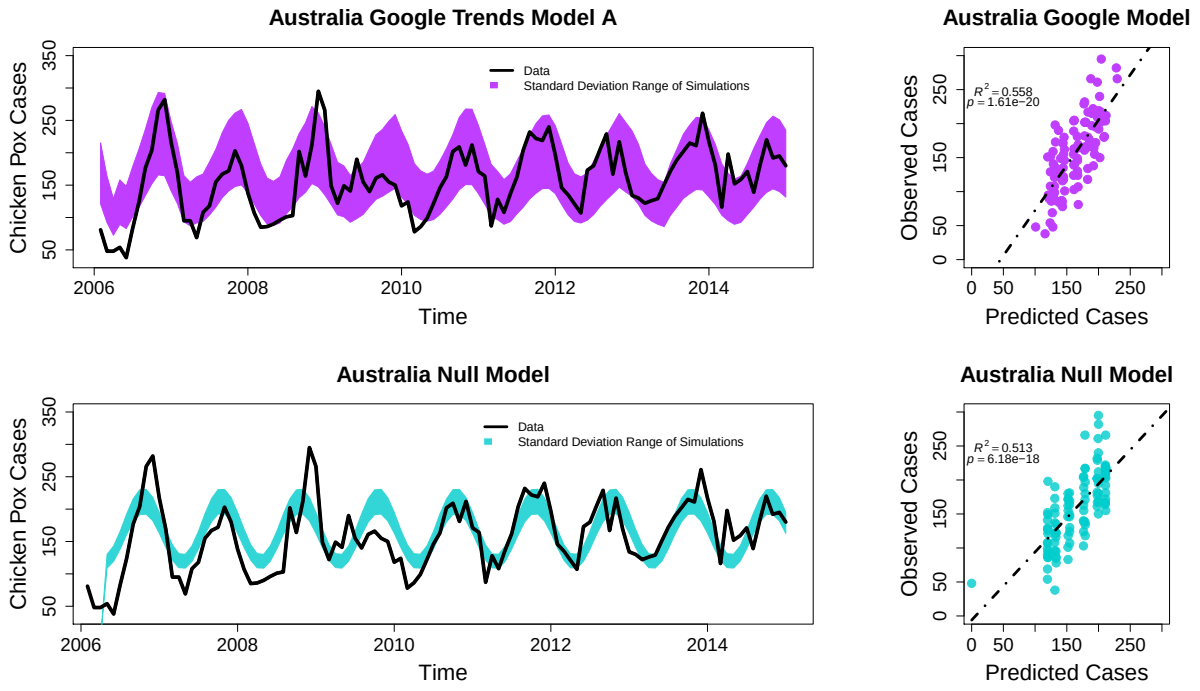
B.1.5 Influenza Information Seeking Behaviour

Our data strongly suggest a signature of immunization on the seasonality of VZV. For the case of varicella, a readily apparent signature of immunization is perhaps to be expected because, when rolled-out into a population at high coverage, the VZV vaccine is highly effective [12, 188]. Therefore, we examined whether observable signatures of immunization can also be detected in diseases with lower vaccine efficacy. In order to determine whether vaccine effects could also be observed in Google data for other vaccine preventable disease, we obtained annual data on inactivated influenza vaccine efficacy and vaccine administration, weekly influenza and pneumonia mortality, and weekly information seeking regarding influenza and flu symptoms in the US state of Wisconsin (Fig B.4). We measured effective immunization as the percent of the population expected to be immunized based on doses administered and vaccine efficacy, which varies substantially from year-to-year. Influenza mortality and information seeking displayed interannual variation not readily attributable to variation in effective immunization. We interpret this to be due to low effective immunization for influenza, which was $< 25\%$ in all years. Although flu and flu symptoms information seeking did not contain a signature of immunization, influenza and pneumonia mortality covaried with information seeking ($R^2 = 0.34$ and $R^2 = 0.50$ for flu and flu symptoms, respectively). This suggests that if seasonal flu immunization accounted for the interannual variation in influenza mortality, the effect of immunization would be reflected in flu information seeking.

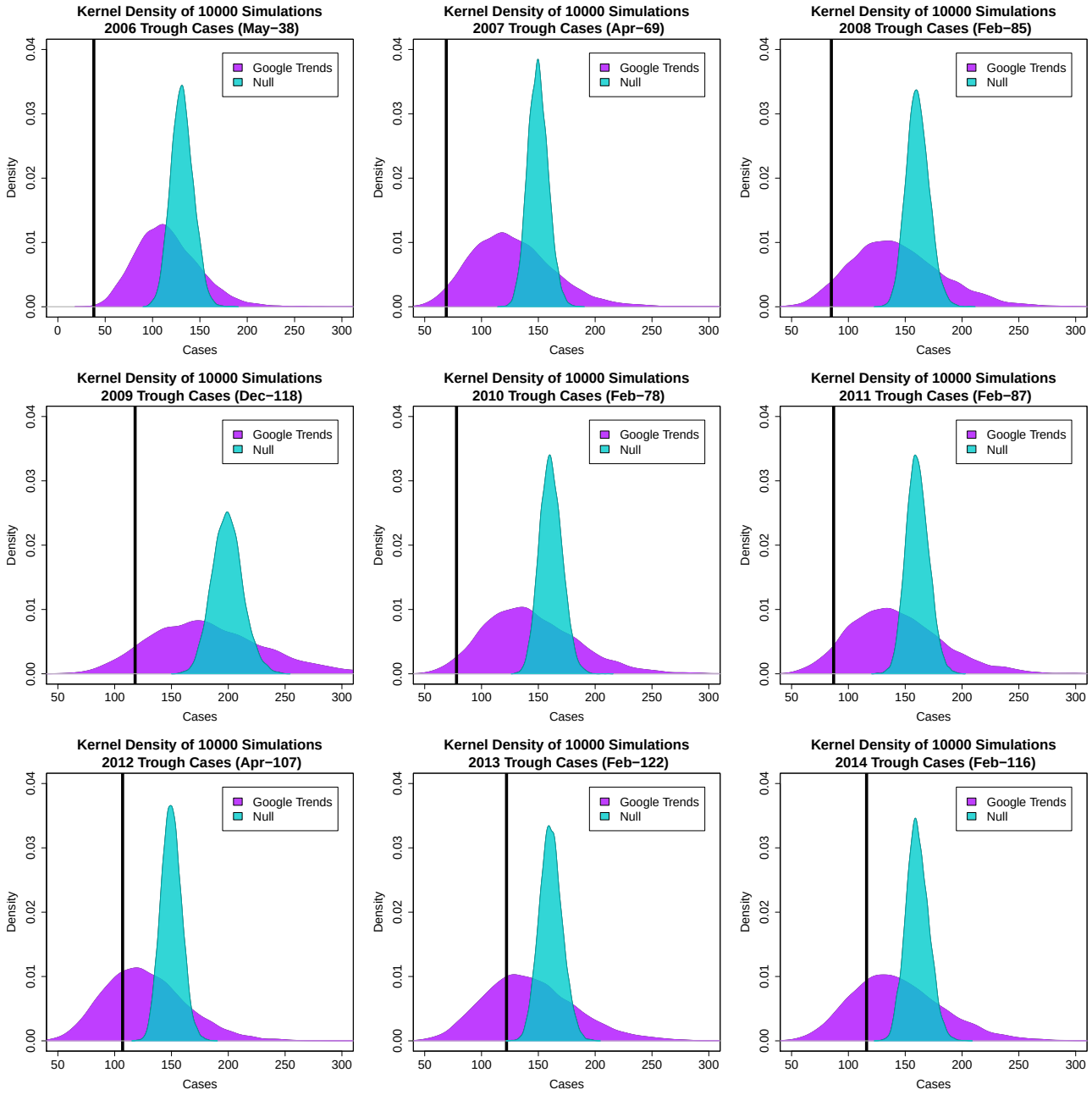
For the state of Wisconsin, weekly influenza information seeking data were obtained from Google trends using the search terms "flu" and "flu symptoms". Wisconsin was chosen because published studies of inactivated influenza vaccine efficacy included patients from Wisconsin. The adjusted vaccine effectiveness estimates for influenza seasons were obtained from the CDC [44]. In years when the lower bound of the 95% CI of vaccine efficacy was negative, the efficacy was set to 0. Weekly influenza and pneumonia mortality was extracted from the Mortality Surveillance Data from the National Center for Health Statistics [43].

Model	Model Structure	LogLik	# Params Est.	AIC	Δ AIC
B.1	Google Model	**-554.47**	6	**1120.9**	**0.0**
B.4	Null Model	−569.47	5	1148.9	28.0
B.5	-β_2	−558.32	5	1128.0	7.1
B.6	-β_2, +α	−563.35	6	1138.7	17.8
B.7	+α	−565.43	7	1144.9	24.0
B.8	- Cosine, -β_2	−585.63	4	1179.3	58.4
B.9	-Cosine, -β_2, +α	−585.02	5	1180.0	59.1
B.10	-Cosine, +α	−584.96	6	1181.9	61.0
B.11	-Cosine	−586.08	5	1182.2	61.3

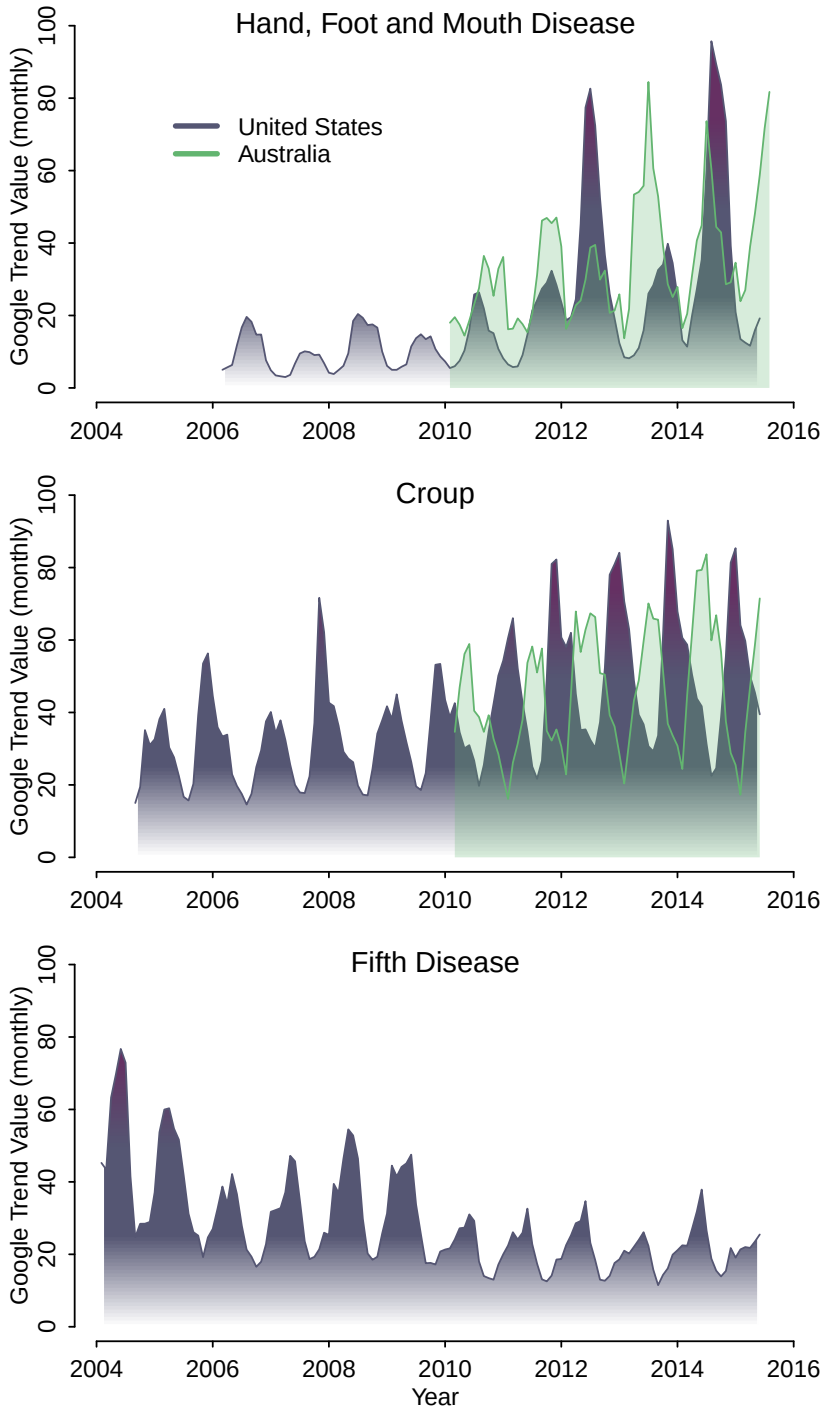
Table B.1: The equation letter, model structure, log–likelihood values, number of estimated parameters, AIC, and difference from top AIC values are shown above. Equation letter matches the equations from Appendix B (main text references to the Google model, refer to the best-fit model, model A). The model structure refers to how the model varies from the top performing Google model. A cosine function, α parameter, or the β_2 parameter were either added (+) or removed (-) from the best fit Google Model. LogLik are the log–likelihood values for each models maximum likelihood parameter set. AIC refers to Akaike Information Criterion which penalizes models that use more parameters. The lowest AIC value represents the best fit model. Δ AIC was the difference in AIC from the best fit model.



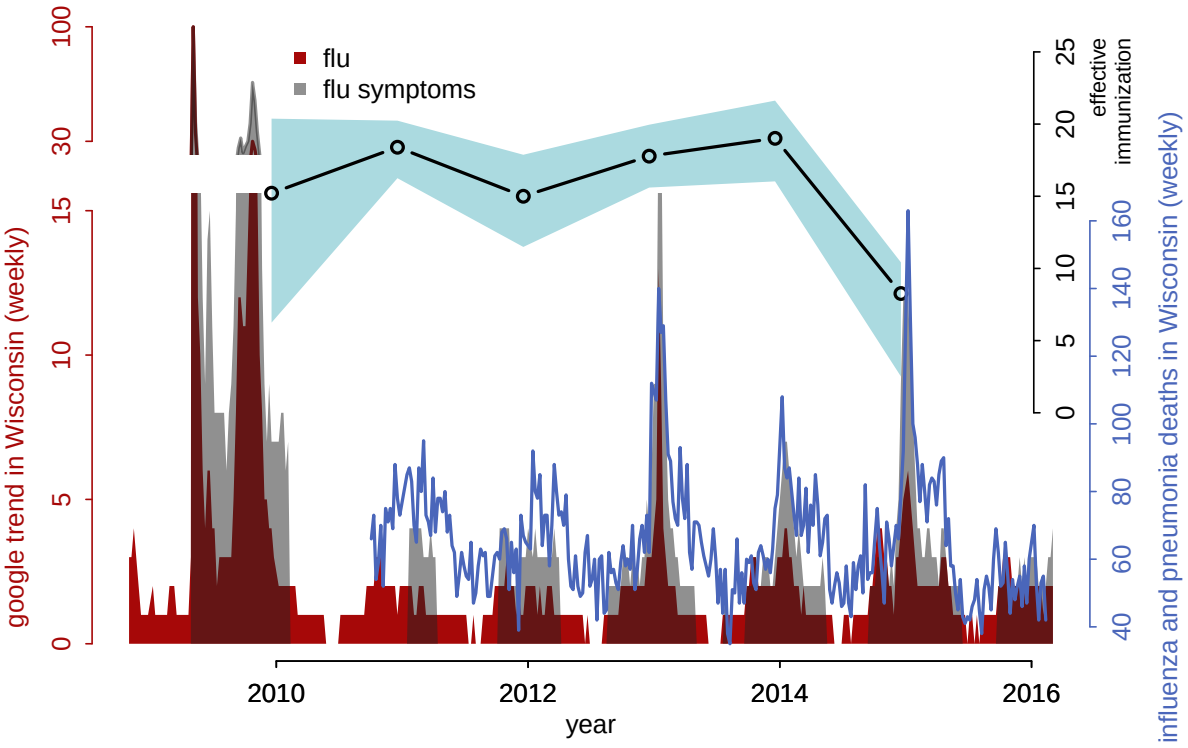
Supplementary Figure B.1: (top left panel)The best–fit Google Trends model (Model A (B.1)), was simulated 10000 times and the range of standard deviation from the mean are plotted in purple against the actual Australian chicken pox case data (black). (top right panel) The relationship between the Google Trends model predicted chicken pox cases and the observed chicken pox cases. (bottom left panel) The Null model (Model B (B.4)), was simulated 10000 times and the range of standard deviation from the mean are plotted in light blue against the actual chicken pox case data (black). (bottom right panel) The relationship between the Null model predicted chicken pox cases and the observed chicken pox cases.



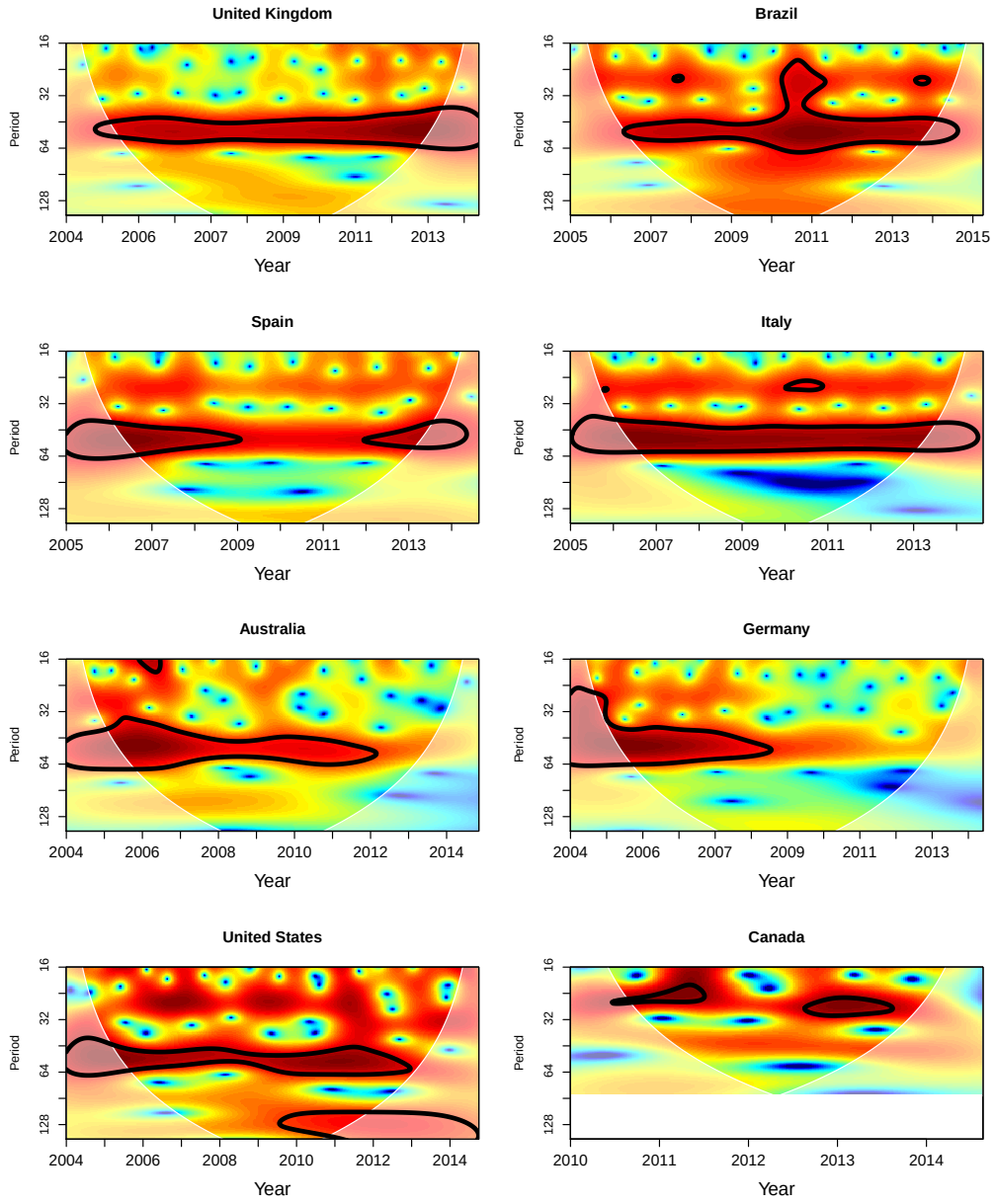
Supplementary Figure B.**2:** Density distributions of the 10000 simulations for each the Google Model (purple) and Null model (light blue) during the trough month in chicken pox cases for each year. The actual number of reported cases are in each panel title, and shown with a vertical black band.



Supplementary Figure B.**3:** Seasonal variation in childhood disease information seeking. Time series of monthly information seeking for Google queries of "hand foot and mouth", "croup", and "fifth disease" in the US and Australia. Hand, foot and mouth disease (HFMD) queries in the US and Australia were relatively in phase with one another, whereas croup queries in the US and Australia were out of phase, with both occurring in the autumn of their respective hemisphere.



Supplementary Figure B.4: Influenza digital epidemiology and mortality in Wisconsin, USA. Weekly influenza and pneumonia mortality (blue line). Weekly Google trends based on the search terms ''influenza" (red) and ''influenza symptoms" (grey). Effective influenza immunization (black). The expected value and range of effective immunization, measured as the percent of the population immunized, was calculated by multiplying the percent of the population vaccinated by the annual vaccine efficacy. The expected values (black line with blue range) are based on the point estimates of vaccine efficacy, the range is based on the 95% CIs for reported efficacy.



Supplementary Figure B.**5:** Wavelet analyses [210] for eight countries. Wavelets are used to identify periodic signals in non-stationary time series data. These signals can vary in amplitude, frequency, and phase over long temporal scales. Wavelets decompose time series data into signals of identifiable period and amplitude, both of which can change with time. For each figure, time is plotted on the x-axis, and the periodicity, in weeks, is plotted on the y-axis, with the colors representing the power of each frequency (blue=low, red=high). Areas circled in black have significant periodicity for that period. In the UK for example, the data are significant at 52 weeks throughout the entire time series (i.e. annual peaks). Meanwhile in both Australia and Germany, significant periodicity was lost during the time period analyzed, while in Spain, significant periodicity was lost between 2009-2012. Canada can only be tested for significance up to 96 weeks because of the short time series available.

Table B.2: Chicken pox search term context. Search terms have been translated from Spanish to English for Mexico and from Thai to English for Thailand. Search terms that could not be properly translated from Thai are indicated by "could not translate" in the category. Note, there were unique searches in Spanish and Thai that resulted in the same English translation.

top chicken pox search terms	relative abundance	category	country	indicator
chicken pox	100	disease (common name or virus)	US	disease
chicken pox vaccine	80	vaccine or vaccination	US	vaccine
shingles chicken pox	75	similar disease	US	other
chicken pox symptoms	45	symptoms	US	disease
chicken pox pictures	30	disease (images)	US	disease
symptoms of chicken pox	30	symptoms	US	disease
what is chicken pox	25	disease (common name or virus)	US	disease
chicken pox adults	25	disease (based on stage of life)	US	disease
chicken pox virus	20	disease (common name or virus)	US	disease
pictures of chicken pox	20	disease (images)	US	disease
chicken pox in adults	20	disease (based on stage of life)	US	disease
chicken pox contagious	20	disease (other)	US	disease
chicken pox rash	20	symptoms	US	disease
chicken pox vaccine	20	vaccine or vaccination	US	vaccine
shingles vaccine	20	similar disease	US	other
varicella chicken pox	15	disease (common name or virus)	US	disease
varicella	15	disease (common name or virus)	US	disease
chicken pox and shingles	15	similar disease	US	other
chicken pox symptoms	15	symptoms	US	disease
signs of chicken pox	15	symptoms	US	disease
chicken pox pictures	10	disease (images)	US	disease
chicken pox in children	10	disease (based on stage of life)	US	disease
shingles from chicken pox	10	similar disease	US	other
never had chicken pox	10	uncategorized	US	other
is chicken pox contagious	10	disease (other)	US	disease
shingles contagious	10	similar disease	US	other
vaccine for chicken pox	10	vaccine or vaccination	US	vaccine
chicken pox treatment	10	care or treatment	US	disease
what is shingles	5	similar disease	US	other
chicken pox virus	5	disease (common name or virus)	US	disease
chicken pox images	5	disease (images)	US	disease
pregnancy and chicken pox	5	disease (based on stage of life)	US	disease

Continued on next page

Table B.2 – *Continued from previous page*

top chicken pox search terms	relative abundance	category	country	indicator
are chicken pox contagious	5	disease (other)	US	disease
chicken pox incubation	5	disease (other)	US	disease
incubation period chicken pox	5	disease (other)	US	disease
chicken pox history	5	uncategorized	US	other
what causes chicken pox	5	disease (other)	US	disease
is shingles contagious	5	similar disease	US	other
incubation for chicken pox	5	disease (other)	US	disease
causes of chicken pox	5	disease (other)	US	disease
cdc chicken pox	5	disease (other)	US	disease
exposure to chicken pox	5	disease (other)	US	disease
cause of chicken pox	5	disease (other)	US	disease
chicken pox transmission	5	disease (other)	US	disease
symptoms for chicken pox	5	symptoms	US	disease
stages of chicken pox	5	symptoms	US	disease
chicken pox vaccination	5	vaccine or vaccination	US	vaccine
varicella vaccine	5	vaccine or vaccination	US	vaccine
treatment for chicken pox	5	care or treatment	US	disease
treatment of chicken pox	5	care or treatment	US	disease
chicken pox	100	disease (common name or virus)	Thailand	disease
a chicken pox	85	disease (common name or virus)	Thailand	disease
chicken pox symptoms	50	symptoms	Thailand	disease
chicken pox treatment	50	care or treatment	Thailand	disease
chicken pox vaccine	40	vaccine or vaccination	Thailand	vaccine
vaccine	40	vaccine or vaccination	Thailand	vaccine
chicken pox medicine	35	care or treatment	Thailand	disease
chicken pox children	20	disease (based on stage of life)	Thailand	disease
chicken pox scars	20	disease (other)	Thailand	disease
prevent chicken pox	15	disease (other)	Thailand	disease
contact chicken pox	15	disease (other)	Thailand	disease
chicken pox blister	15	symptoms	Thailand	disease
chicken pox blisters	15	symptoms	Thailand	disease
treatment of chicken pox	15	care or treatment	Thailand	disease
chicken pox solve	15	care or treatment	Thailand	disease
as chicken pox	10	disease (common name or virus)	Thailand	disease
chicken pox	10	disease (common name or virus)	Thailand	disease
chicken pox infection	10	disease (other)	Thailand	disease
chicken pox is caused by	10	disease (other)	Thailand	disease
chicken pox symptoms	10	symptoms	Thailand	disease
symptoms of chicken pox	10	symptoms	Thailand	disease
itchy chicken pox	10	symptoms	Thailand	disease
chicken pox symptoms	10	symptoms	Thailand	disease
symptoms of chicken pox	10	symptoms	Thailand	disease
price chicken pox vaccine	10	vaccine or vaccination	Thailand	vaccine
chicken pox vaccine	10	vaccine or vaccination	Thailand	vaccine
chicken pox treatment	10	care or treatment	Thailand	disease
treatment of chicken pox	10	care or treatment	Thailand	disease
treatment of chicken pox	10	care or treatment	Thailand	disease

Continued on next page

Table B.2 – *Continued from previous page*

top chicken pox search terms	relative abundance	category	country	indicator
chicken pox scars	10	disease (other)	Thailand	disease
green medicine (traditional medicine)	10	care or treatment	Thailand	disease
chicken pox green medicine (traditional medicine)	10	care or treatment	Thailand	disease
chicken pox cure	10	care or treatment	Thailand	disease
roy chicken pox	10	could not translate	Thailand	other
E-cooked	10	could not translate	Thailand	other
do not eat chicken pox	10	could not translate	Thailand	other
food chicken pox	10	could not translate	Thailand	other
the chicken pox	5	disease (common name or virus)	Thailand	disease
the chicken pox	5	disease (common name or virus)	Thailand	disease
chicken pox in children	5	disease (based on stage of life)	Thailand	disease
topical chicken pox	5	disease (other)	Thailand	disease
measels	5	similar disease	Thailand	other
chicken pox vaccine	5	vaccine or vaccination	Thailand	vaccine
chicken pox vaccine	5	vaccine or vaccination	Thailand	vaccine
to prevent chicken pox	5	disease (other)	Thailand	disease
scar treatment	5	care or treatment	Thailand	disease
chicken pox scar treatment	5	care or treatment	Thailand	disease
chicken pox scars	5	disease (other)	Thailand	disease
chicken pox wound healing	5	care or treatment	Thailand	disease
chicken pox hole	5	could not translate	Thailand	other
the chicken pox	100	disease (common name or virus)	Mexico	disease
chicken pox symptoms	25	symptoms	Mexico	disease
what is chicken pox	15	disease (common name or virus)	Mexico	disease
chicken pox infants	15	disease (based on stage of life)	Mexico	disease
chicken pox in infants	15	disease (based on stage of life)	Mexico	disease
symptoms of chicken pox	15	symptoms	Mexico	disease
measels	15	similar disease	Mexico	other
small pox	15	similar disease	Mexico	other
chicken pox adults	10	disease (based on stage of life)	Mexico	disease
chicken pox in adults	10	disease (based on stage of life)	Mexico	disease
pregnancy chicken pox	10	disease (based on stage of life)	Mexico	disease
treatment chicken pox	10	care or treatment	Mexico	disease
chicken pox vaccine	10	vaccine or vaccination	Mexico	vaccine
rubella	10	similar disease	Mexico	other
zoster chicken pox	5	disease (common name or virus)	Mexico	disease
chicken pox virus	5	disease (common name or virus)	Mexico	disease
chicken pox images	5	disease (images)	Mexico	disease

Continued on next page

Table B.2 – *Continued from previous page*

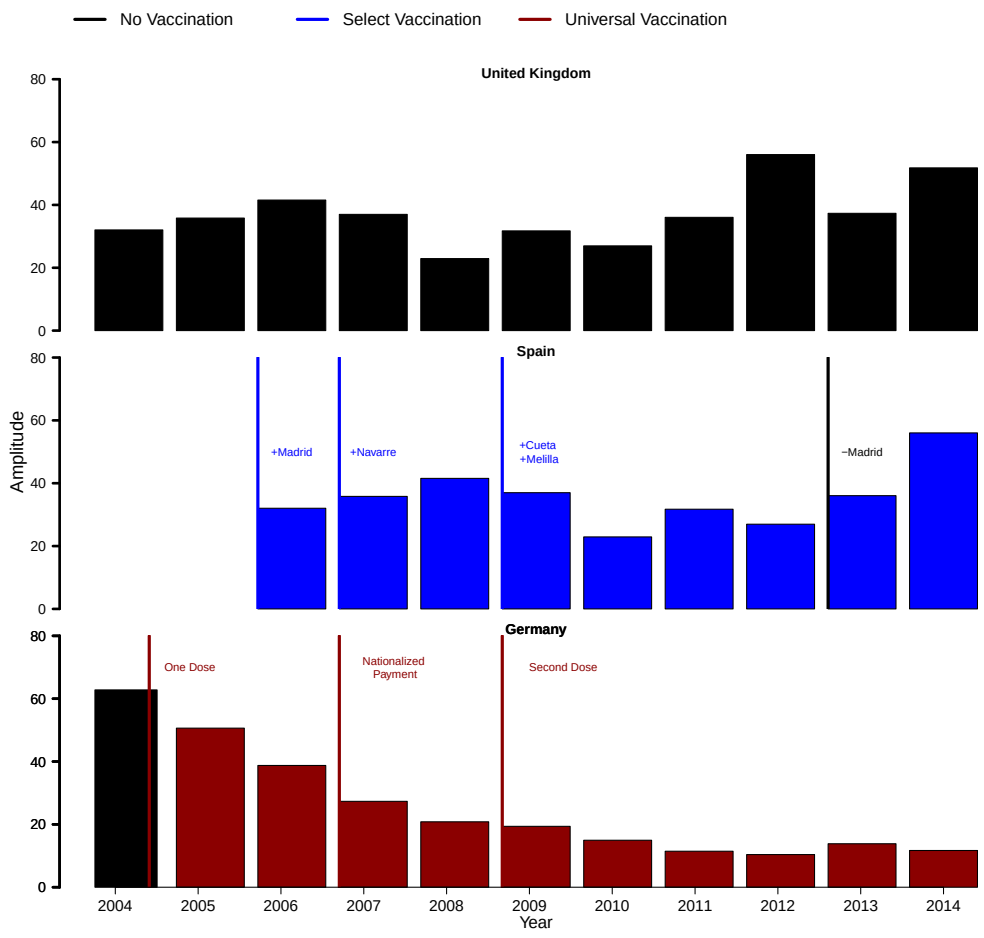
top chicken pox search terms	relative abundance	category	country	indicator
images of chicken pox	5	disease (images)	Mexico	disease
chicken pox pdf	5	disease (images)	Mexico	disease
chicken pox in pregnancy	5	disease (based on stage of life)	Mexico	disease
chicken pox babies	5	disease (based on stage of life)	Mexico	disease
chicken pox in babies	5	disease (based on stage of life)	Mexico	disease
symptoms chicken pox infants	5	disease (based on stage of life)	Mexico	disease
chicken pox spread	5	disease (other)	Mexico	disease
spread of chicken pox	5	disease (other)	Mexico	disease
hemorrhagic chicken pox	5	disease (other)	Mexico	disease
remedies for chicken pox	5	care or treatment	Mexico	disease
care chicken pox	5	care or treatment	Mexico	disease
aciclovir	5	care or treatment	Mexico	disease
treatment of chicken pox	5	care or treatment	Mexico	disease
chicken pox home remedies	5	care or treatment	Mexico	disease
scars of chicken pox	5	disease (other)	Mexico	disease
treatment for chicken pox	5	care or treatment	Mexico	disease
marks of chicken pox	5	disease (other)	Mexico	disease
chicken pox postulates	5	symptoms	Mexico	disease
chicken pox care	5	care or treatment	Mexico	disease
vaccine for chicken pox	5	vaccine or vaccination	Mexico	vaccine
vaccine against chicken pox	5	vaccine or vaccination	Mexico	vaccine
the small pox	5	similar disease	Mexico	other
small pox and chicken pox	5	similar disease	Mexico	other
zoster herpes	5	similar disease	Mexico	other
chicken pox (misspelling)	0	disease (common name or virus)	Mexico	disease
chicken pox (misspelling)	0	disease (common name or virus)	Mexico	disease
photos of chicken pox	0	disease (images)	Mexico	disease
chicken pox and pregnancy	0	disease (based on stage of life)	Mexico	disease
chicken pox twice	0	disease (other)	Mexico	disease
scarlet fever	0	similar disease	Mexico	other
measel symptoms	0	similar disease	Mexico	other
small pox symptoms	0	similar disease	Mexico	other
symptoms chicken pox	100	symptoms	Australia	disease
rash	60	symptoms	Australia	disease
chicken pox rash	60	symptoms	Australia	disease
chicken pox	55	disease (common name or virus)	Australia	disease
chicken pox vaccine	50	vaccine or vaccination	Australia	vaccine
shingles chicken pox	45	similar disease	Australia	other
shingles	45	similar disease	Australia	other
measles	45	similar disease	Australia	other

Continued on next page

Table B.2 – *Continued from previous page*

top chicken pox search terms	relative abundance	category	country	indicator
chicken pox pictures	35	disease (images)	Australia	disease
adults chicken pox	30	disease (based on stage of life)	Australia	disease
chicken pox pregnancy	30	disease (based on stage of life)	Australia	disease
chicken pox children	25	disease (based on stage of life)	Australia	disease
chicken pox australia	25	disease (other)	Australia	disease
chicken pox contagious	25	disease (other)	Australia	disease
chicken pox baby	20	disease (based on stage of life)	Australia	disease
chicken pox spots	20	symptoms	Australia	disease
chicken pox vaccination	20	vaccine or vaccination	Australia	vaccine
chicken pox treatment	20	care or treatment	Australia	disease
chicken pox virus	15	disease (common name or virus)	Australia	disease
chicken pox babies	15	disease (based on stage of life)	Australia	disease
chicken pox pregnant	15	disease (based on stage of life)	Australia	disease
chicken pox immunisation	15	vaccine or vaccination	Australia	vaccine
varicella	10	disease (common name or virus)	Australia	disease
varicella chicken pox	10	disease (common name or virus)	Australia	disease
chicken pox images	10	disease (images)	Australia	disease
chicken pox in adults	10	disease (based on stage of life)	Australia	disease
chicken pox twice	10	disease (other)	Australia	disease
chicken pox incubation	10	disease (other)	Australia	disease
measles rash	10	similar disease	Australia	other
measles symptoms	10	similar disease	Australia	other
mumps	10	similar disease	Australia	other
chicken pox signs	10	symptoms	Australia	disease
chicken pox scars	10	disease (other)	Australia	disease
chicken pox photos	5	disease (images)	Australia	disease
chicken pox picture	5	disease (images)	Australia	disease
chicken pox toddler	5	disease (based on stage of life)	Australia	disease
chicken pox stages	5	disease (other)	Australia	disease
chicken pox herpes	5	disease (other)	Australia	disease
chicken pox mild	5	disease (other)	Australia	disease
small pox	5	similar disease	Australia	other
rubella	5	similar disease	Australia	other
german measles	5	similar disease	Australia	other
symptoms of chicken pox	5	symptoms	Australia	disease
shingles symptoms	5	similar disease	Australia	other
rashes	5	symptoms	Australia	disease

Continued on next page



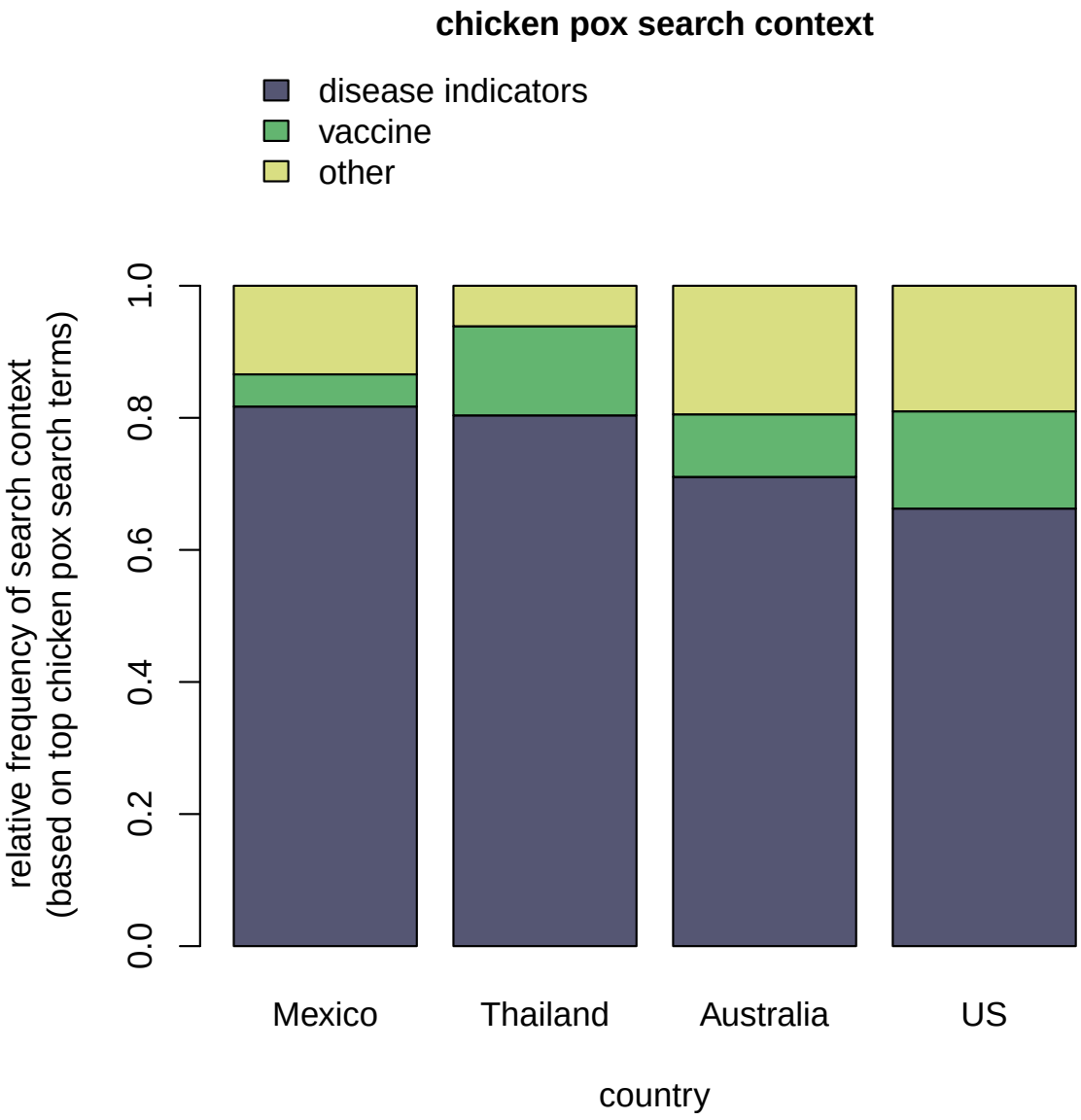
Supplementary Figure B.6: Annual amplitude values of Google searches for the United Kingdom, Spain, and Germany. See text for additional captioning.

Table B.2 – *Continued from previous page*

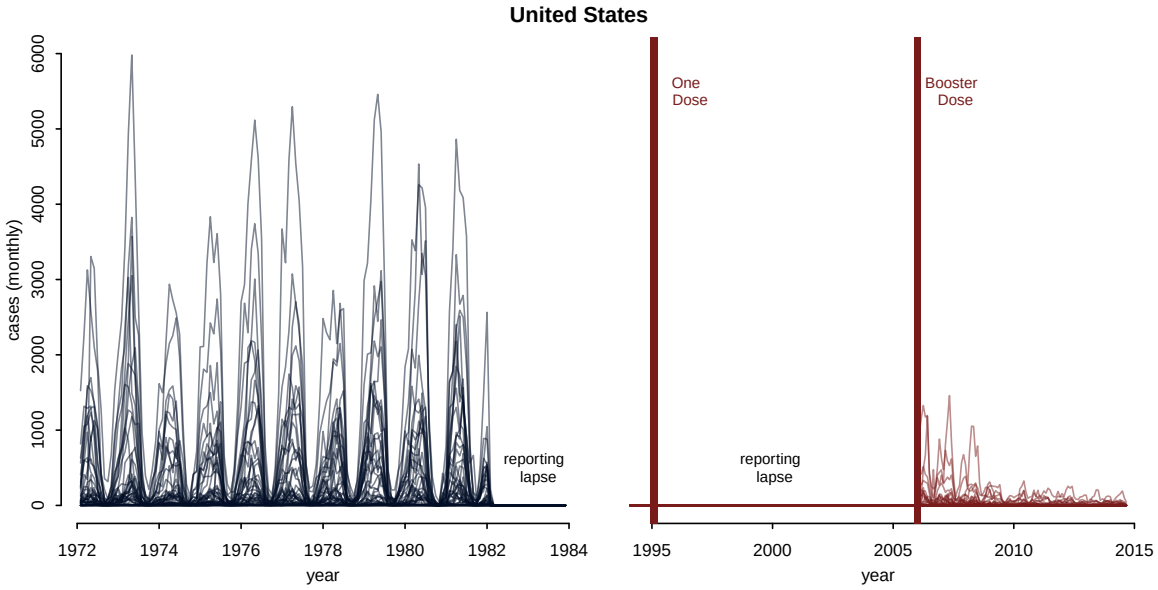
top chicken pox search terms	relative abundance	category	country	indicator
chicken pox rash	5	symptoms	Australia	disease
chicken pox vaccine	5	vaccine or vaccination	Australia	vaccine

For Supplementary Figure B.6, amplitudes were computed by first calculating the difference between the maximum each year and the mean each year. Second, we subtracted the minimum each year from the mean each year. Third, we found the difference between those two values and divided by two to get the final amplitude. Amplitude = $((max(yr1) - mean(yr1)) - (min(yr1) - mean(yr1))/2$. In Spain, municipalities differed in

their implementation of VZV vaccination. The Madrid metro region represents $\sim 14\%$ of the Spanish population (6.5m/46.7m), meaning that the vaccination policy in Madrid will have a large impact on overall chicken pox incidence and chicken pox Google Trends for the country. Spain initially had a significant seasonal period. However, after VZV vaccination was implemented in Madrid and additional cities, the significant seasonal periodicity was lost. Interestingly, the seasonality became significant again when Madrid withdrew VZV vaccination. This is similar to Germany (Figs 2.3 and B.5), where the loss of significant wavelet periodicity followed the implementation of routine immunization after a few years. To examine this loss of seasonality in Google Trends in closer detail, we analyzed the annual amplitude for these two countries and the United Kingdom, which all differed in immunization mandates. The United Kingdom, which has no requirements, Spain, which implemented vaccination in certain municipalities for varying time periods, and Germany, which gradually increased its requirements over the course of a few years: first it required one shot, then made the payments nationalized, and finally required a second dose. In the UK, with no immunization requirements, the annual amplitude of Google searches for chicken pox remains relatively constant. In Spain, when all four municipalities were immunizing, the amplitude decreased from $\sim 40\%$ to $\sim 20\%$ in two years, before Madrid stopped vaccinating, after which the amplitude increased to over 50%. Meanwhile, in Germany pre-vaccine amplitudes in Google searches were $\sim 60\%$, before dropping to $\sim 40\%$ after the requirement of one dose, then dropping to $\sim 20\%$ after instituting nationalized payments, and finally dropping to $\sim 10 - 15\%$ after requiring a second dose. This additional analysis clearly elucidates the impact of immunization on search seasonality.



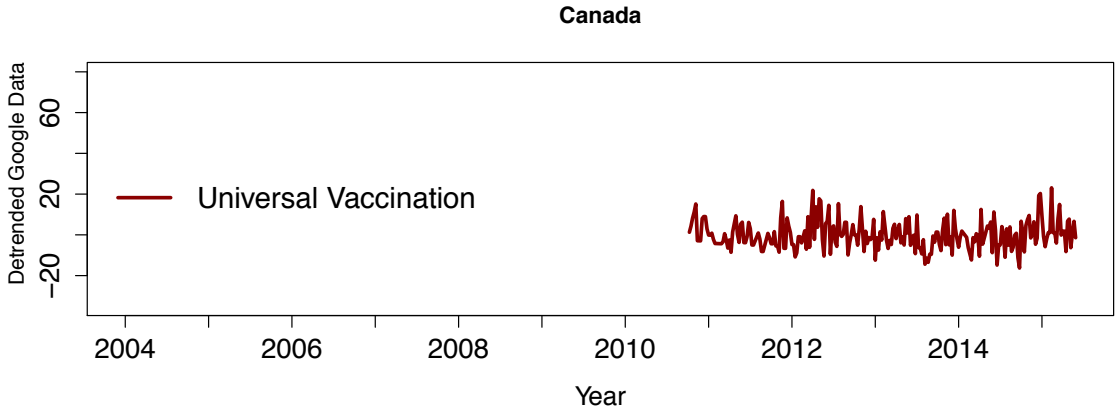
Supplementary Figure B**.7:** Relative frequency of language-specific chicken pox searches. Top searches were categorized into three broad categories: disease indicators, vaccine, or other. Disease indicators were searches considered to indicate chicken pox in the household/community. Vaccine indicators were searches regarding the VZV vaccine, and all other search contexts were placed in the "other" (also see Table B.2).



Supplementary Figure B.**8:** Weekly, state-level, data on chicken pox cases from [218] during the years 1972-2015 for all US states. Each state is plotted as an individual line. Black lines represent reported chicken pox cases during the pre-vaccine era, while red lines represent reported chicken pox cases during the vaccine era in the US

Table B.3: Countries included in our chicken pox information seeking dataset from Google Trends. Countries in black lack nationwide immunization, countries in red began nationwide immunization in 2013 (Brazil) or 2014 (Japan), and countries in green have had nationwide immunization for multiple years. Countries with significant annual (seasonal) periodicities are in *italics*.

Countries			
Argentina	*Australia*	Austria	*Brazil*
Canada	Chile	*China*	*Colombia*
Czech Republic	*Denmark*	*Estonia*	*Finland*
France	Germany	*Hungary*	*India*
Iran	*Ireland*	*Italy*	*Japan*
Mexico	*Netherlands*	*New Zealand*	*Philippines*
Poland	*Portugal*	*Romania*	Russia
South Africa	*Spain*	*Sweden*	*Thailand*
United Kingdom	*United States*	Venezuela	*Vietnam*



Supplementary Figure B.9: Canada Google Trends time series, data only available for the period with Universal Immunization (in red). This figure displays the lack of seasonality in the Google Trends data for Canada, a country with active immunization since 2000.

APPENDIX C

Supporting Information: The Underpinnings of Herpesvirus Dynamics: Transmission & Reactivation of Varicella Zoster Virus

C.1 The Underpinnings of Herpesvirus Dynamics: Transmission & Reactivation of Varicella Zoster Virus: Supporting Information

C.1.1 Introduction

Chickenpox, commonly referred to as varicella zoster (VZ), is a respiratory transmitted infectious disease that causes a characteristic red rash and pox on the skin surface [177]. It is caused by the varicella zoster virus (VZV) which also causes shingles, or herpes zoster (HZ). Symptoms typically arise 1-3 weeks after exposure to an infected individual, and a newly infected individual is infectious for around a week starting 1-2 days prior to the onset of symptoms. Symptoms last approximately two weeks, when the virus then retreats to the nerve ganglia in the spine [222, 234]. By the age of 15, almost all children have antibodies to VZV, either from a natural infection or the vaccine [185, 225]. In approximately 20% of

adults, the latent varicella virus is reactivated and manifests as shingles, typically in adults aged over 60 [208]. The VZV vaccine, typically administered in two doses during childhood, prevents infection with VZV in children, while a booster dose later in life suppresses VZV reactivation in adults [185].

Chickenpox has been studied extensively and is noted by its explosive springtime outbreaks. The mechanism driving increased spring tranmission is still a topic of debate. One of the earliest studies identified temperature as the likely culprit [222], but later examination of school terms identified the seasonal aggregation of children in school (term-time forcing) as the main facilitor of outbreaks [104,127]. However, recent work has argued that term-time forcing cannot be solely responsible for seasonality of chickenpox transmission [136]. Other hypothesized mechanisms include immune related functions, such as seasonal gene expression [63], systemic immunosuppression caused by UV irradiation [212], or seasonal differences in antibodies due to pathogen exposure [55,170].

Shingles is a non-notifiable disease, with epidemiological studies focused on smaller sample populations. It has been noted to lack seasonal patterns in incidence [40,97,182], including small-scale population based studies [24, 46, 79, 156]. Seasonality has been identified in two studies of shingles, both of which examined less than 200 total patients over multiple years [99, 230]. Shingles reactivation is thought to be modulated by ultraviolet (UV) irradiation, which suppresses immunity [57, 81, 185, 238]. Other proposed mechanisms for VZV reactivation include a new exposure to the virus [148], stress [97, 177, 182, 234], a weakened immune system [185], or some amount of time since chickenpox infection due to a decline in cell-mediated immunuty [234]. Regardless of the mechanism, it has been commonly accepted that shingles displays limited to no seasonal patterns.

While the VZV vaccine has been available since the early 1996, it is included as part of the mandated childhood immunization schedule in only a few countries, including the U.S., Canada, Australia, and Germany. Other countries suggest it, or require it in certain regions, but most do not require the vaccine. Because childhood infection with VZV has minimal

complications, low mortality, and symptoms are more mild the younger an individual is, parents encourage their kids to contact infected individuals to expose their children to the pathogen, stimulating the immune system and generating antibodies, early in life [225]. VZV immunization is not on the list of WHO mandated vaccines because implementation could increase shingles incidence, which has more severe symptoms and higher mortality [155]. Other reasons for limiting vaccination include risk to the fetus in pregnant women, immunocomprimised individuals [225], or the knowledge of antibody boosting, where adult exposure to children infected with chickenpox boosts VZV antibodies [26] lessening the chance of reactivation [207], meaning that requiring the vaccine must be carefully considered [120].

Population-level chickenpox and shingles data for Thailand were downloaded to test what mechanisms may drive the seasonal patterns of these two diseases. Previous attempts at modeling the interaction between chickenpox and shingles revealed no association between these two diseases [82]. Since so few countries actively report country-level chickenpox infections, and only two that we could find (Australia and Thailand) reported both chickenpox and shingles, studies examining the dynamics of these two diseases in concert have been limited. Since most VZV clinical case data in Australia were characterized 'unspecified VZV' rather than chickenpox or shingles, this work focused on Thailand, which has an excellent history of disease surveillance.

C.1.2 Methods

C.1.2.1 Data

Country and regional level (Figs 4.4, C.4, C.5) ultraviolet irradiation (UV) data were obtained from the National Center for Atmospheric Research (NCAR) [1]. Two datasets were obtained, one spanned the years 1979-2010 [173] and the other 2011-present [174]. Four, 6-hour daily averages, were extrapolated to monthly data for the time series ranging 2003-2011. Thus a general smoothing occured with the UV data as 1/4 day averages were taken

and then averaged over the entire month. This was done to account for the differences in sun exposre over the course of the year, as there are more hours of sunlight during the summer, and less during the winter. By including the averages, we are inspecting the relationship between sun exposure, tested by UV, and shingles reactivation.

C.1.2.2 Model

To accurately model VZV dynamics, we constructed a Susceptible, Exposed, Infected with chickenpox, Latent (recovered from chickenpox), Infected with shingles, and second Latent (recovered from shingles) state model (Fig 4.1). Once entering the second latent state (shingles recovered), we removed the possibility of a second reactivation due the low natural occurance rate of 0.1% [211]. Population data for Thailand were collected from the United Nations, and interpolated for monthly birth data [213]. All models were fit to the reported chickenpox and shingles clinical case data. Models were fit to the first 8 years of data (2003-2010), with the last year left for out-of-fit estimates. To account for the trend in reported clinical cases of shingles, parameters to estimate the linear increase in reporting were estimated. The generalized model transition probabilities follow a Poisson process, where the fraction of those who remained in the susceptible state were modeled as;

$$pS = e^{-(FOI+\delta)} \tag{C.1}$$

where FOI is the force of infection and δ is the death rate, which was assumed to be constant across all model states;

$$FOI = \beta \left(\frac{(I_{VZ}) + (\omega I_{HZ})}{N} \right)^{\alpha} \epsilon \tag{C.2}$$

where β was the time-varying seasonal force of infection for chickenpox infections (see below - model combinations), I_{VZ} was the number of current individuals infected with chickenpox, ω was a scalar for the relative infectiousness of shingles - allowing for reactivated shingles

individuals to infect a susceptible individual, I_{HZ} was the number of individuals with reactivated shingles, N was the total population, α was a scalar for the force of infection, and ϵ was a noise term which acts as environmental stochasticity. ϵ_t, was drawn from a gamma distribution with a mean of 1 and variance (process noise dispersion parameter) θ. Both the θ and the reporting rate dispersion parameter τ were drawn from a normal distribution, with a mean of 1. The fraction of those who remained in the exposed state were modeled as;

$$pE = e^{-(\phi+\delta)} \tag{C.3}$$

where ϕ was the length of time between when an individual was exposed to chickenpox and before they became infectious, this parameter was fixed at 2 weeks. δ was the fixed death rate for all classes, set for an average lifespan of 60 years. The fraction of those who remained in the infected with chickenpox state I_{VZ} were modeled as;

$$pIVZ = e^{-(\gamma+\delta)} \tag{C.4}$$

where γ was a fixed parameter in which indivduals recovered from chickenpox (7 days). The fraction of those who remained in the first latent state, but remained susceptible to shingles reactivation were modeled as;

$$pL1 = e^{-(\iota\kappa\psi+\delta)} \tag{C.5}$$

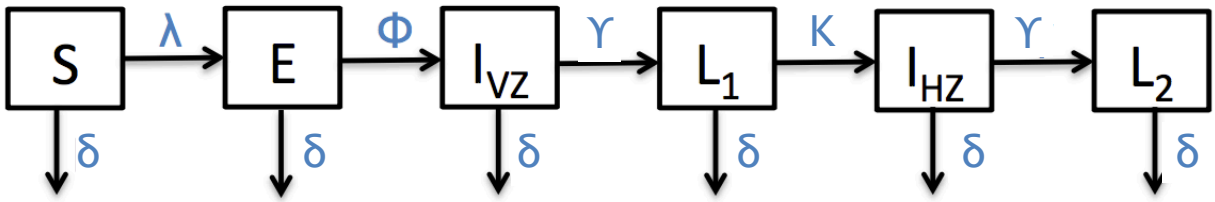
where ι was a fixed parameter which represented the fraction of those infected with chickenpox that would reactivate later in life as shingles (age 50 years), κ is the seasonal reactivation rate of shingles, and ψ is the time-varying antibody boosting from chicken pox infections (see below for both κ and ψ model combinations). The fraction of those who remained infected with shingles were modeled as;

$$pIHZ = e^{-(\gamma+\delta)} \tag{C.6}$$

with the same recovery rate, γ as chickenpox, while the fraction of those who remained in the second latent state were modeled as;

$$pL2 = e^{-(\delta)} \tag{C.7}$$

meaning that death was the only way out of the second latent state. The different states (S-E-IVZ-L1-IHZ-L2) are displayed in Fig C.1, using the transition rates from Equations A-G. Model state trasitions are shown Equations H-M.



Supplementary Figure C.1: Model transitions are described in Equations C.1-C.15, with the seasonal transmission for chickenpox represented as β, and the seasonal reactivation of shingles represented as κ in Equations C.16-C.19, while the interaction of chickenpox incidence on shingles reactivation described in Eq. C.20.

$$S_{t+1} = \mu_t + (S_t * pS) \tag{C.8}$$

where μ_t is the number of children born at time, t. Below are the transitions for states E-L2;

$$E_{t+1} = S_t * (1 - pS)\left(\frac{FOI}{FOI + \delta}\right) + (E_t * pE) \tag{C.9}$$

$$I_{VZ_{t+1}} = E_t * (1 - pE)\left(\frac{\phi}{\phi + \delta}\right) + (I_{VZ_t} * pI_{VZ}) \tag{C.10}$$

$$L_{1_{t+1}} = I_{VZ_t} * (1 - pI_{VZ})\left(\frac{\gamma}{\gamma + \delta}\right) + (L_{1_t} * pL_1) \tag{C.11}$$

$$I_{HZ_{t+1}} = L_{1_t} * (1 - pL_1) \left(\frac{\iota\kappa\psi}{\iota\kappa\psi + \delta} \right) + (I_{HZ_t} * pI_{HZ}) \tag{C.12}$$

$$L_{2_{t+1}} = I_{HZ_t} * (1 - pI_{HZ}) \left(\frac{\gamma}{\gamma + \delta} \right) + (L_{2_t} * pL_2). \tag{C.13}$$

New infections for chickenpox were recorded as;

$$I_{VZ_{new}} = E_t * (1 - pE) \left(\frac{\phi}{\phi + \delta} \right) \tag{C.14}$$

while new shingles infections were recorded as;

$$I_{HZ_{new}} = L_{1_t} * (1 - pL_1) \left(\frac{\iota\kappa\psi}{\iota\kappa\psi + \delta} \right). \tag{C.15}$$

C.1.2.3 Model Combinations

We tested four different models, each examining a unqiue biologically related hypothesis regarding VZV dynamics. Each of these four models were also tested with and without anitbody boosting preventing the reactivation of shingles, thus a total of eight models were tested. Each model represents a specific combination of hypotheses related to to the seasonal transmission of chickenpox (β) and reactivation of shignles (κ). Model 1 tested whether both chickenpox and shingles dynamics were driven by the same seasonal forcing, likley seasonal susceptibility/immunity. This was examined by fitting a single B-Spline function with 6 basis to both chickenpox and shingles dynamics;

$$\beta = \kappa = \exp \sum_{i=1}^{6} q_i \zeta_{A_{i_t}} \tag{C.16}$$

where β is the seasonal forcing for chickenpox, and κ is the seasonal forcing for shingles, and each ζ_A is a B-spline basis with 1 year period. Here, β and κ are equal to eachother, but the FOI for chickenpox (Eqn. C.2) does not equal the reactivation likelihood for shingles (Eqn.

C.5, $\iota\kappa\psi$). Two different B-splines were used for the second model, one for each chickenpox and shingles. Biologically, this represents the hypothesis that chickenpox is driven by school terms (wherein students have increased contact rates), while a different seasonal forcing (possibly some seasonal immunomodulation) drives shingles reactivation. β is the same as in Eq. C.16, while shingles reactivation, κ was ,modeled as;

$$\kappa = \exp \sum_{i=1}^{6} q_i \zeta_{B_{i_t}} \tag{C.17}$$

where each ζ_B is a different (from ζ_A) B-spline basis with 1 year period. The third model combination again depicts a seasonal school-term forcing for chickenpox with a B-spline, while examining the potential role of UV irradiation in shingles reactvation. β is the same as in Eq. C.16, while κ takes the shape of a sigmoidal curve;

$$\kappa = \frac{1}{\pi + e^{(\nu UV + \xi)}} \tag{C.18}$$

where π, ν, and ξ are parameters being estimated for the shape of the sigmoidal curve, as UV varies. The fourth model combination examines whether UV drives both chickenpox transmission and shingles reactivation, where κ is the same as Eq. C.18, but β is;

$$\beta = \frac{1}{\chi + e^{(\upsilon UV + \sigma)}} \tag{C.19}$$

where χ, υ, and σ are parameters being estimated for the shape of the sigmoidal curve as UV varies. Thus, two sigmoidal curves were estimated, one to estimate the effect of UV on chickenpox transmission and another for shingles reactivation.

Finally all four models were fit with and without VZV antibody boosting (ψ); where contact between an individual infected with chickenpox could boost VZV antibodies in individuls who had not developed shingles, making them less susceptible to VZV reactivation. This biological interaction has been previously observed, where shingles reactivation was suppressed by exposure to individuals infected with chickenpox [26]. Models fit without

anitbody boosting, set $\psi = 1$. This forcing was modeled as;

$$\psi = \frac{1}{1 + e^{(\eta I_{VZ} + o)}}. \tag{C.20}$$

where η and o are two parameters being estimated to determine the shape of the boosting curve, and I_{VZ} were the number of individuals infected with chickenpox at that time step. In total, we fit four model combinations, each with and without antibody boosting, for a total of eight different models. All parameters were estimated using maximum likelihood by iterated particle filtering (MIF) in the R-package *pomp* [111, 112].

C.1.3 Results

C.1.3.1 VZV Dynamics

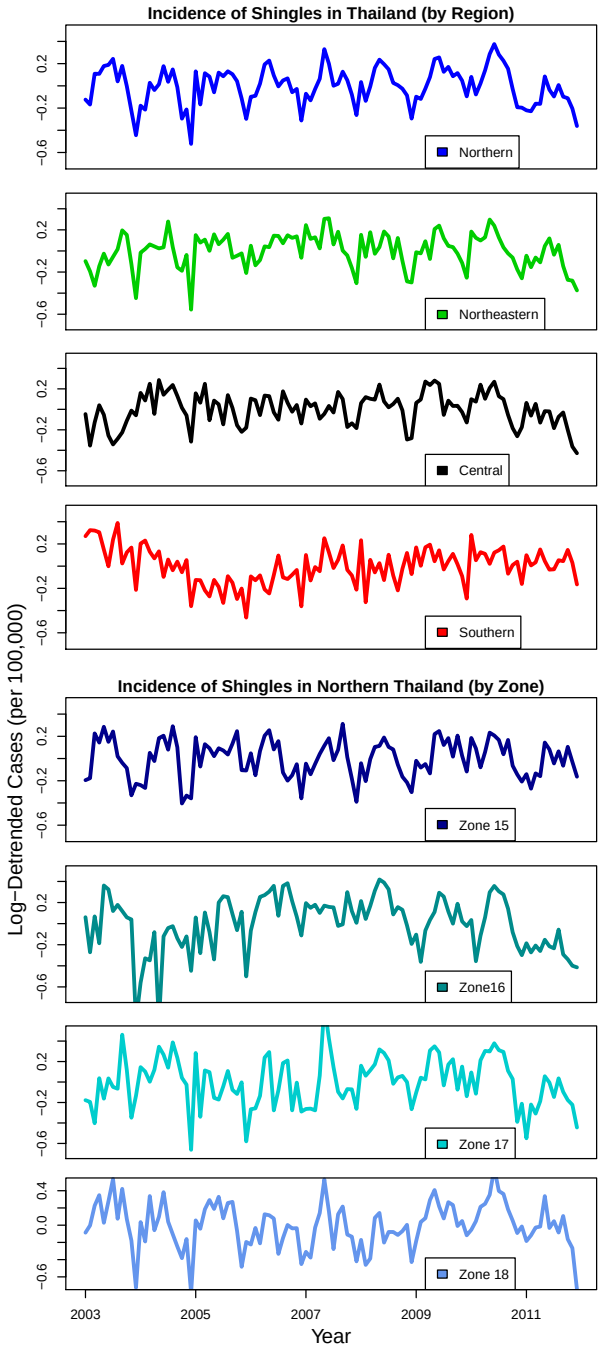
Additional, zonal shingles data can be seen in Fig C.2, where the two northernmost zones (15 & 16) display the clearest seasonality. Unsuprisingly, these were the two zones with highest shingles infection rates (Fig C.3). A further breakdown of the regional UV data and shingles data can be seen in Fig C.4 and Fig C.5.

C.1.3.2 Modeling

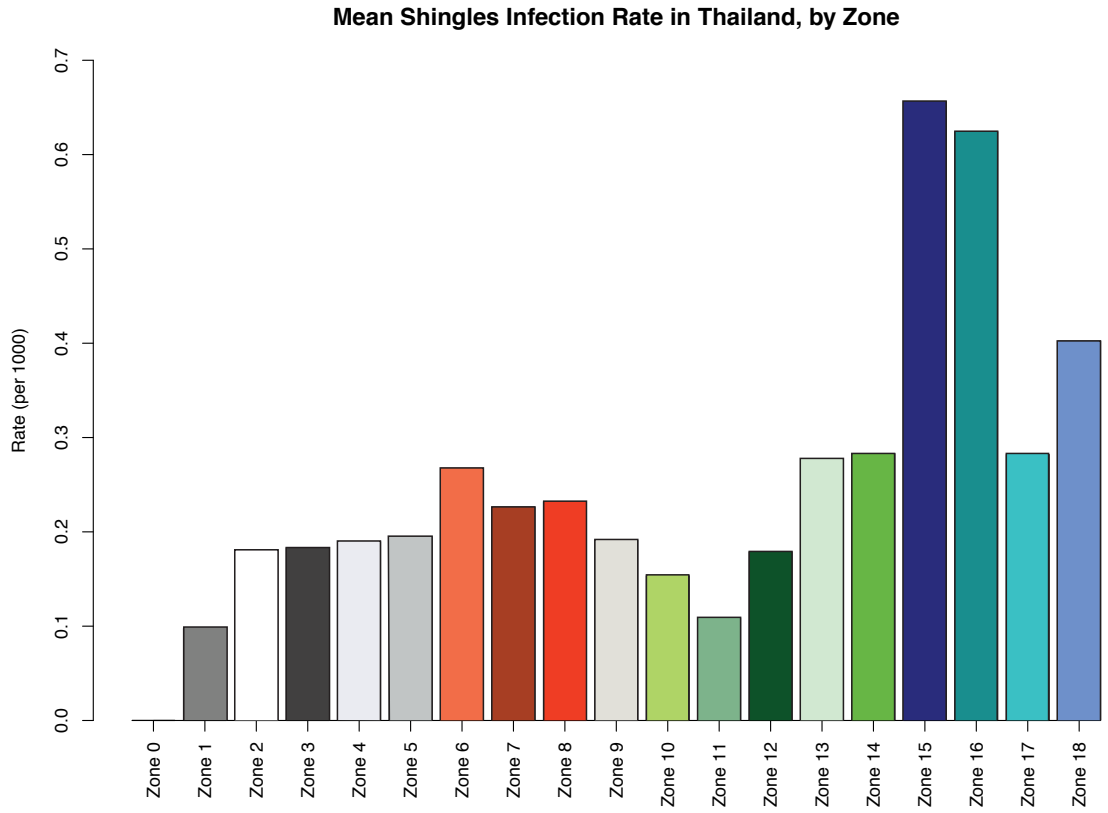
Modeling results from our 8 models can be seen in Table C.1.

ID	Model Description	VZ	HZ	Eqns.	Params	LogLik	AIC
1	1 seasonal forcing	BSplineA	BSplineA	C.16,C.20	12	-1435.1	2894.2
2	1 seasonal forcing, AB	BSplineA	BSplineA	C.16	14	-1436.2	2900.0
3	**2 seasonal forcings**	**BSplineA**	**BSplineB**	**C.16,C.17,C.20**	**18**	**-1329.8**	**2695.6**
4	2 seasonal forcings, AB	BSplineA	BSplineB	C.16,C.17	20	-1332.9	2705.8
5	1 seasonal forcing, UV	BSplineA	UV scalarHZ	C.16,C.18,C.20	15	-1360.8	2751.6
6	**1 seasonal forcing, UV, AB**	**BSplineA**	**UV scalarHZ**	**C.16,C.18**	**17**	**-1359.6**	**2753.2**
7	UV	UV scalarVZ	UV scalarHZ	C.18,C.19,C.20	11	-1417.5	2857.0
8	UV, AB	UV scalarVZ	UV scalarHZ	C.18,C.19	13	-1424.9	2875.8

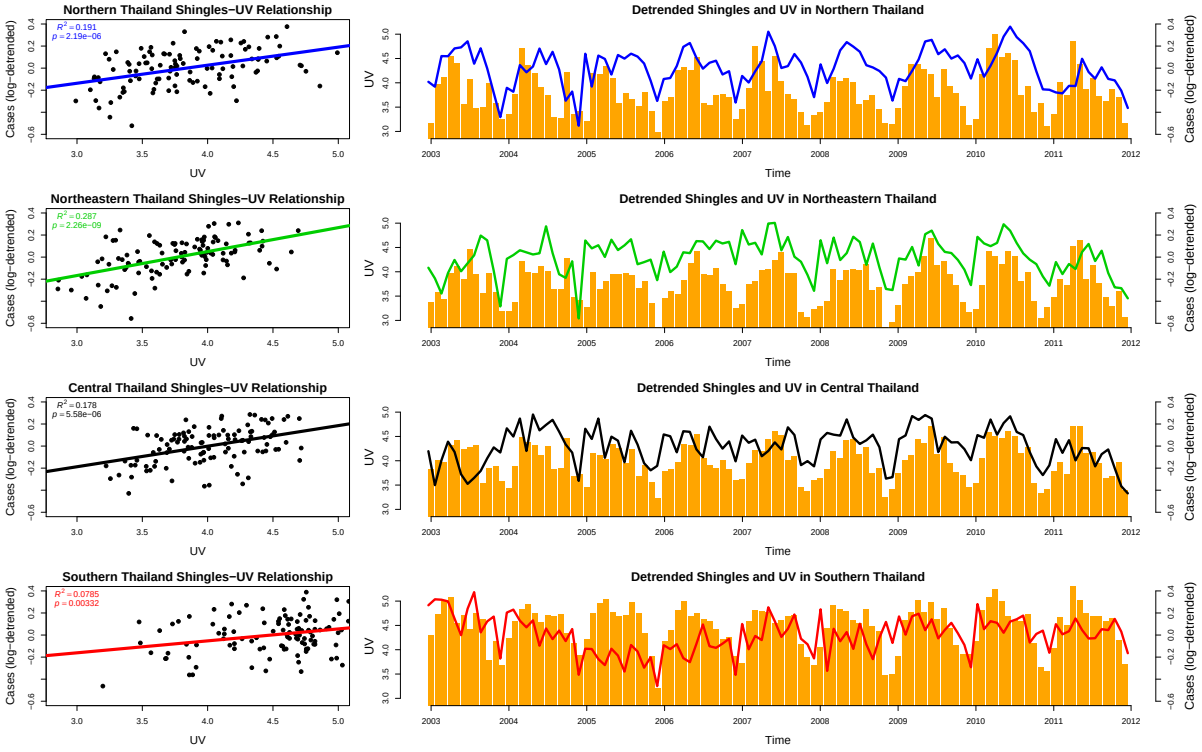
Table C**.1:** A description of the models used to fit VZ and HZ incidence in Thailand. ID is for descriptive purposes in the results section, Model Description is a quick reference to the biology of the model (AB is antibody boosting), VZ is the type of seasonality used to fit varicella zoster, HZ is the type of seasonality used to fit herpes zoster, Eqns references which equations were used for each model combination (Equations C.1-C.15 were used in all models), Params is the number of parameters being estimated, LogLik is the current best log likelihood of that model, and AIC is Akaike Information Criterion, used to compare models with a different number of parameters.



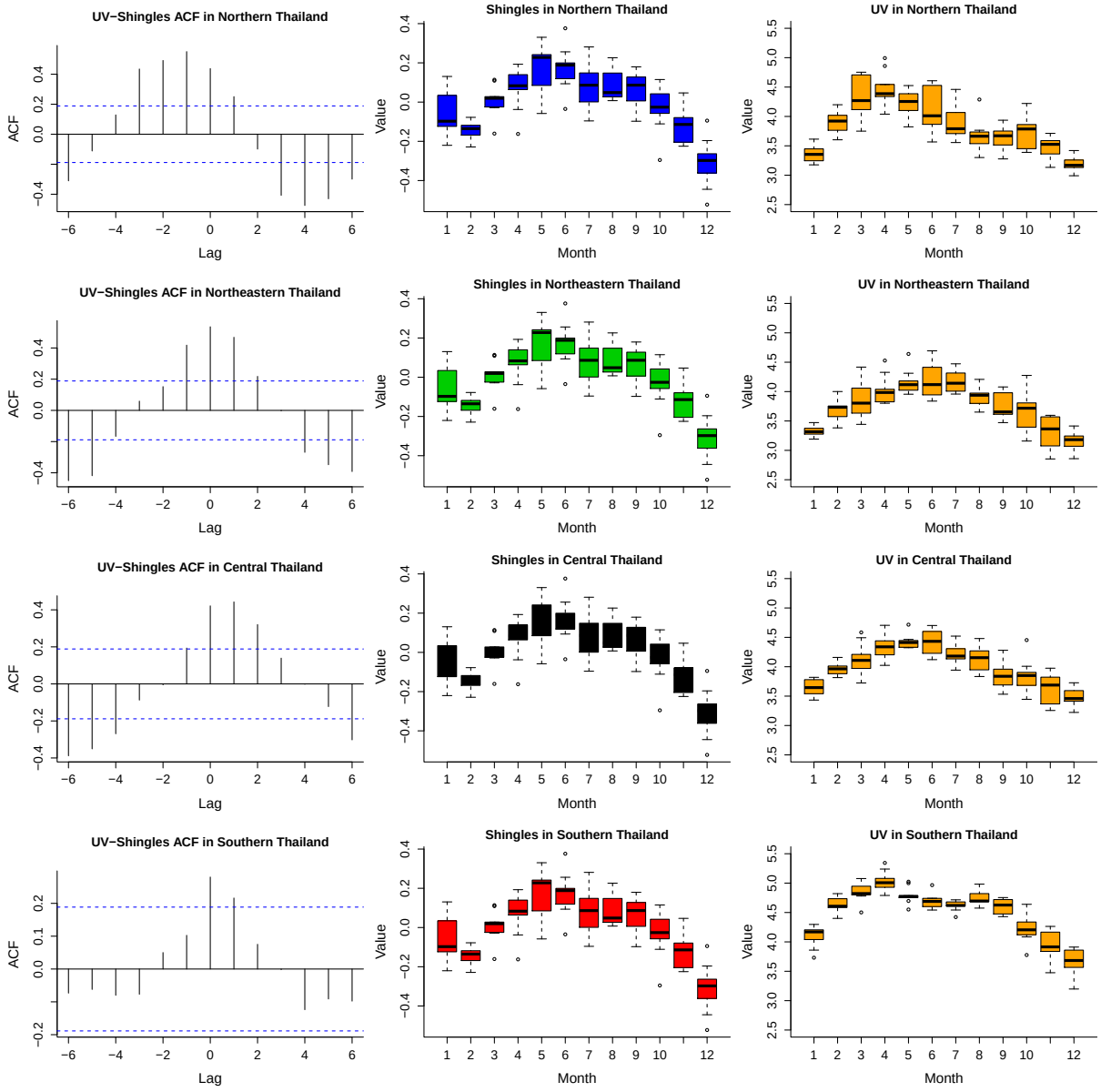
Supplementary Figure C.**2:** (A) (top four panels) Log-detrended monthly incidence of shingles by region demonstrating the seasonality - arranged North to South from top. (bottom four panels) Wavelet analyses of log-detrended monthly shingles incidence of by region - arranged North to South from top.



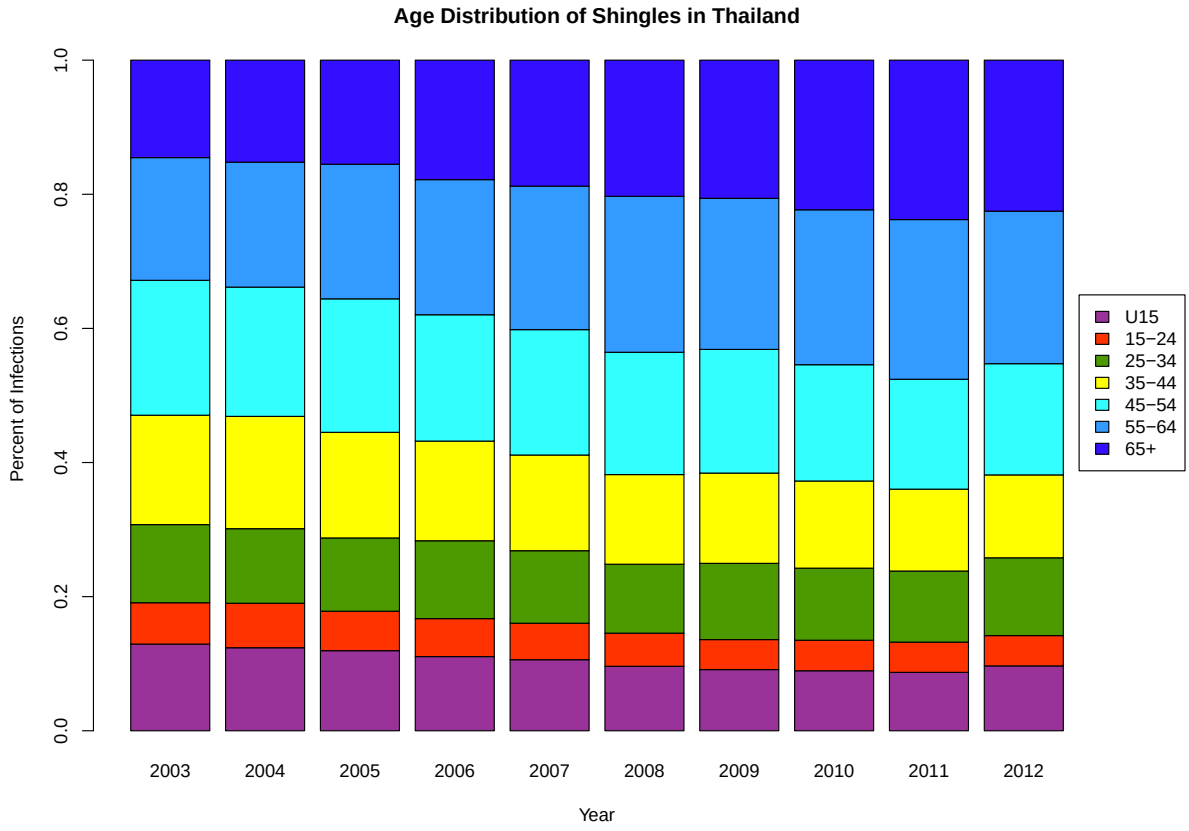
Supplementary Figure C.**3:** Mean shingles infection rates (per 1000 individuals) by zone in Thailand from 2003-2011. Rates were calculated individually, using zone population figures each year.



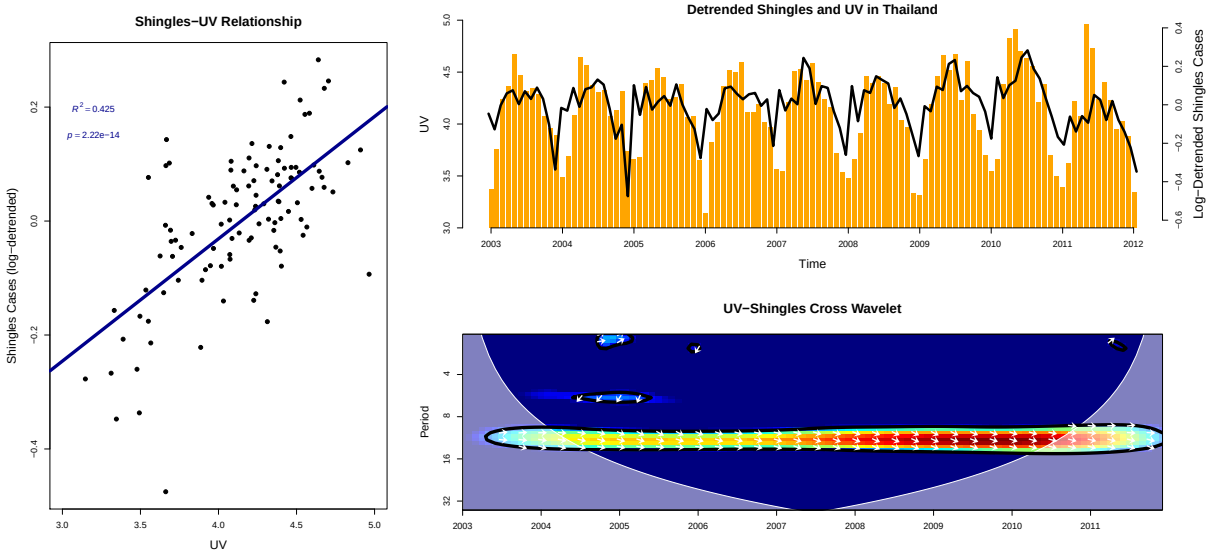
Supplementary Figure C.4: Regional log-detrended shingles cases and regional-specific UV data were examined for correlations.



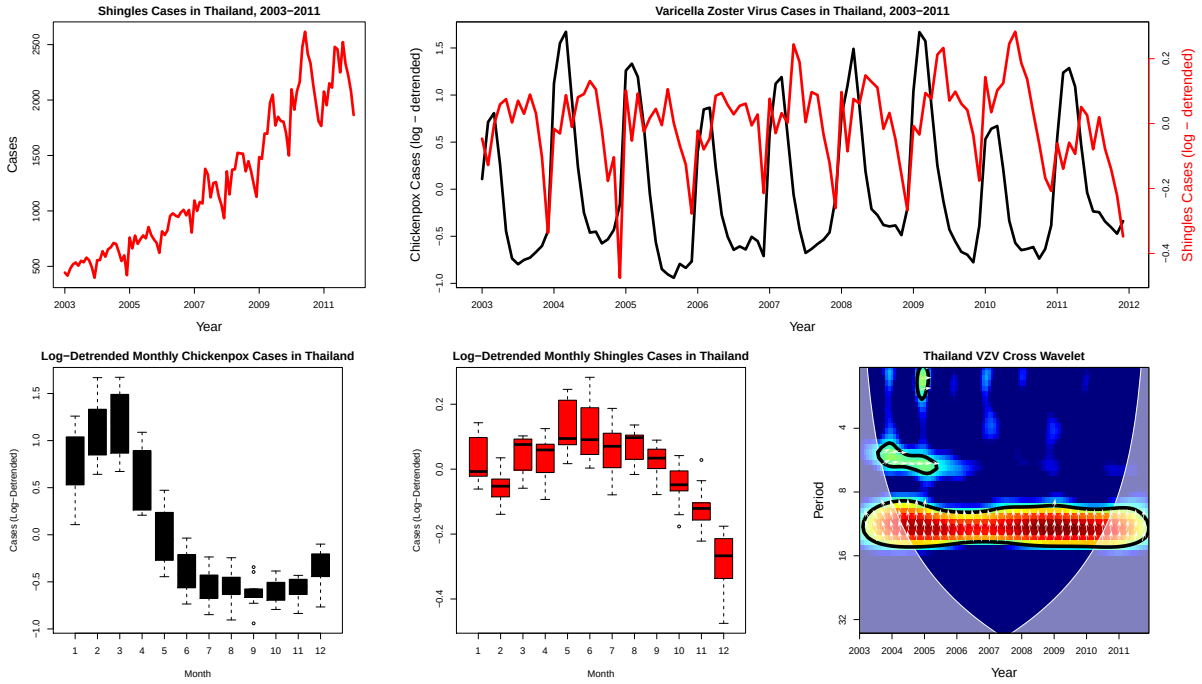
Supplementary Figure C.**5:** UV-Shingles relationships by region. Each row is a region organized by latitude. Top row northern, 2nd row northeastern, 3rd row central, and 4th row southern. Each column is a different analysis. Left column are cross-correlations of UV-shingles, middle column are boxplots of log-detrended shingles cases, and right column are the region-specific UV data.



Supplementary Figure C.6: Age distribution of shingles in Thailand.



Supplementary Figure C.7: National level relationship between UV and shingles in Thailand



Supplementary Figure C.8: Chickenpox-shingles relationship in Thailand

APPENDIX D

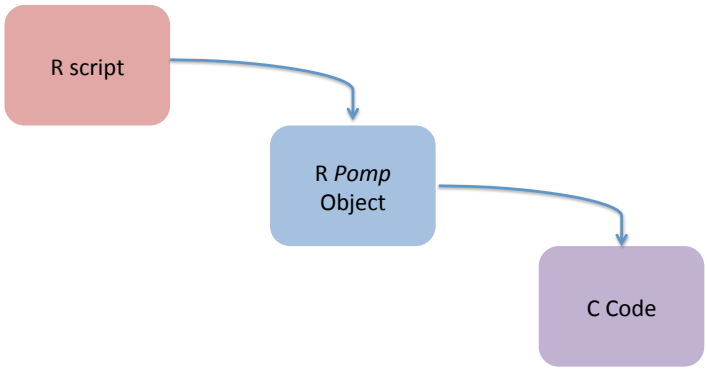
On Reproducibility: *Pomp* Model Settings and Code

D.1 On Reproducibility: *Pomp* Settings and Code

D.1.1 Introduction

For reproducibility purposes, I have included all *pomp* code from Chapters 2, 3 & 4, as well as a tutorial to *pomp*. All data used in this work is either freely available, listed online on my website (kevinmbakker.com), or available on the publishers website if chapter is published. While the general structure of the code follows the basic principles of an R script calling an R *pomp* object which then compiles the C-code, the complexity of the code evolved throughout my dissertation (Fig D.1).

Additionally, each chapter used a different version of *pomp*, with the largest noticible difference between Chapter 2, which used mif1 and Chapters 3 & 4 which used mif2, a vastly improved fitting process (see [111] for details). Finally, to properly search parameter space, all of my models started with a sobol design (sequences) configuration for an initial search of parameter space. This has been shown to be more effective at searching the parameter space than a pseudorandom number sequence. I have included basic code for sobol parameter creation in the first section in below, though the reader will need to create the .csv file which



Supplementary Figure D.1: Basic *pomp* structure, with an R script being run to call the *pomp* object, which then compiles the C++ code.

sets parameter limitations.

D.1.2 A *pomp* Tutorial

Below, one can find a *pomp* tutorial, including a readme file, R script, *pomp* object, and C++ code. Also included are examples on how to create a Sobol Parameter design, and simple examples of how to look at pomp results.

```
#In this section is a readme file on the basic structure of pomp, which aligns
    with the walkthrough I created for measles fitting.


# Fitting models using Maximization by iterated particle filtering (MIF)


#  What you need to run MIF:
#     1. Data
```

```
#    2. Covaritates
#    3. Your model (preferably coded in C)
#    4. R script for running MIF
#    5. Parameter sets

#  We will walk through these one by one.

# Data

#  You will want to have your time series data in a csv or some other text file
    you can load into R.
#  For this example we will work with a short time series of measles in France.
    The data is in the file "data.csv" and is listed in the column "monthly_cases
    ".
#  Note, I define the data as whatever we are fitting the model to: cases for a
    between-host model or virus load for within-host. The data is treated
    differently than covariates in POMP, which is why this distinction is
    important.

# Covariates

#  Covariates are other sets of data or variables that change through time and
    will be factored into the model. For between-host models these may include
    things like population size, temperature, birth rates, vaccine uptake, etc...
    For within host models these may include things like CD4+ concentrations,
    cortisol levels, etc.
#  In our file "data.csv" the covariates are: births, population size, and vaccine
     coverage.
```

```
# The model in C

#  For this example we will be using a discrete-time SIR model. The model is in
    the file "measles_biweekly.c"

#  Your model has to have three components:

#  1. The process model.

#  In our case this is a stochastic discrete-time SIR model. The process model is
    the epidemiological model of the relationship among states.

#  2. The measurement model.

#  This is a model that describes how the process relates to your data. For
    example, an SIR model tracks susceptible, infected, and recovered individuals.
     However, the data is cases, which is different from the total number of
    infecteds. Thus, the measurement model is a formulation of how the data
    relates to the underlying process. In this example, some fraction of infected
    individuals become reported cases and there is some noise involved in the
    measurement.

#  3. The measurement density.

#  This is a model that defines the likelihood calculation. The likelihood is our
    objective function which we want to maximize.


# Parameter definitions and transformations
```

```
#  The first thing you need to do is define each of your model paramters. To
    search parameter space using MIF it is best to transform all of your free
    parameters using either log or logit. Use log for any parameter that needs to
    be > 0 and use logit for parameters that need to be between 0-1.

#  In our example we have 14 parameters. Each parameter is names and is assigned a
     position in a parameter vector p. Single parameters are assignmed a positions
     using the syntax:

   (p[parindex[i]])

#  And a vector of parameters can be assigneed using the syntax:

   (&p[parindex[i]])

#  We will see later where the "parindex" vector comes into play.


#   process model state variable definitions

#    To define the state variables in the process model use the syntax:

      (x[stateindex[i]])


#   covariate definitions

#    Use the syntax:
```

```
      (covar[covindex[0]])

#   measurement model state variable definition

#    Use the syntax:

      (y[obsindex[i]])


# Defining the Process model

#  Let's go to line 76 of the C++ code to look at the process model.
#  Notice in line 77-81 we define

#  state values: double *x
#  parameter values: const double *p
#  etc.

#  These are attributed that will be used in the process model and must be defined
    .

#  The next thing is to define the process model. In our example this is all the
    code in the {} lines 82-127. The first thing to do is to define your parameter
     values on their natural scale. For example our parameter psi is a double and
    it was log transformed in in our parameter vector above so we define psi as:

   double psi;

   psi = exp(LOGPSI);
```

```
#  Once you have all of your parameters defined on their natural scale you can
    define the relationship among the state variables which we did here on lines
    123-127. One thing to keep in mind when writing your process model is that you
     want it to be relatively fast to simulate because you will need to run
    millions upon millions of simulations when using MIF, so if possible use a
    discrete time model because the numerical integration will be much faster with
     fewer time steps.

#  * Notice that the process model has a noise term "epsilon", which is a draw
    from a normal distribution. Particle filtering is built to deal with
    stochastic models so this kind of fitting procedure is overkill if you don't
    have process noise. Common types of process noise include using the draw from
    a normal or gamma distribution.


# Defining the measurement model

#  The measurement model (lines 57-74) can also be thought of as a reporting model
    .
#  The measurement model arguments include:

#  measurement model state variables: double *y
#  process model state variables: double *x
#  parameters: double *p
#  etc.

#  Similar to the process model we first define parameters on their natural scale
    and then define the realtionship between the process model state varables and
```

```
    the measurement model state variables.

#  In our example we assume that the number of individuals that end up observed as
     a reported case in our data is a random draw from a normal distribution with
    a mean of
#  "report_rate * I" and a standard deviation that scales with "I".


# Defining measurement density model

#  The measurement density model (lines 45-54) is essentially a single line
    defining how to calculate the likelihood of the measurement model output.
    Remember, the output of the measurement model is what actually gets compared
    to the data. The array of likelihood values *lik is defined using the
    probability desity that complements our measurement model. We drew our
    reported cases "MEASLES" from rnorm() with a mean of "report_rate*I" and
    standard deviation of "tau*I". Thus, to get the likelihood of drawing "MEASLES
    " cases from our normal distribution we use dnorm. Note, that I added 1e-6 to
    the standard deviation, this is to prevent getting a mean, standard deviation
    and reported cases of 0, which will give you an Inf likelihood. Also, it's
    best to use log liklihoods so you aren't dealing with small numbers.

    *lik = dnorm(MEASLES,report_rate*I,1e-6+tau*I,give_log);


# Creating the POMP object and getting the model running in R


#Now that we have our model coded in C, we need to be able to call this model in
```

```
    R using pomp. We will create an object called a pomp object which will contain
     our model, covariates, and data. Once the pomp object is set up we can run
    simulations which allows us to calculte likelihoods using a vanilla particle
    filter, and we can fit parameters using iterated particle filtering.

#The subsection below listed 'Additional R Code on Simulating Results'
    demonstrates the set up and test of a pomp object.
```

D.1.2.1 R Code for Creating Sobol Parameters

```
####################
# Here we will use a Sobol Design to generate parameter sets that are dispersed
    across parameter space.
# The Sobol design is similar to a Latin Hypercube sampling where you try to
    optimally cover a multidimensional space.
####################

# Load a csv file with the upper and lower bounds of your parameters
# These bounds need to be the upper and lower bounds for each parameter that are
    biologically feasible.
# Don't make these bounds too narrow! Make the bounds as large as you can, you
    can always narrow the parameter space later.
# But you run the risk of not finding a global max Liklihood parameter set if you
     make your bounds too narrow.
# The bounds need to be on the LOG or LOGIT scale for each parameter that will be
     estimated.
# Remember, most parameters will be log transformed, with the exception of those
    bounded between 0-1, those will be logit transformed.
bounds<- read.csv('parameter_bounds_LOG_LOGIT_scale.csv',header=TRUE)
```

```
# lower bounds
lb<- bounds$lower.bound
names(lb)<- bounds$param

# upper bounds
ub<- bounds$upper.bound
names(ub)<- bounds$param

# Generate parameter sets using sobolDesign() function in POMP
# If you have more than a handful of parameters I would suggest starting with 50K
    + parameter sets

sobol.pars<- sobolDesign(lb,ub,nseq=50000)

write.csv(sobol.pars,'sobol_parameters.csv',row.names=FALSE)
```

D.1.2.2 R script

```
rm(list=ls())
require(pomp)
setwd('/Users/kevinbakker/Desktop/pomp_walkthrough/')
run<- c(1) # Split up the work into multiple runs/cores
no.runs<- c(100) # This would be the number of cores you are going to use. I
    usually split analysis like this among 10-80 cores.
# You will need to do speed testing to decide how many cores you want to use. I
    set the number of cores to 10000 here so that this code will run only 5 param
    sets and finish in a few minutes.
```

```
# Each phase of analyses will take 4-30d depending on how many observations,
    parameter sets, cores, the numer of replicates, particles per replicate, etc.

no.MC.reps <- c(1) # Number of Monte Carlo replicates per parameter set.
output.name<- c("MIF_results") # Name of your output csv file
###############
# load pomp object
###############
source('pomp_object.R')

###############
# load Sobol design parameter sets
###############
param.sets<- read.csv(file='sobol_parameters.csv',header=TRUE)

###############
#Set up storage space for results & generate seeds (seeds make your results
    reproducible)
###############
total.params<- dim(param.sets)[1]*no.MC.reps

MIFed.params<-mat.or.vec(total.params,dim(param.sets)[2]+3)
MIFed.params<-as.data.frame(MIFed.params)
names(MIFed.params)<- c(names(param.sets),'LogLik','seLogLik','seed')

seeds <- ceiling(runif(n=total.params,min=1,max=2^30))
MIFed.params[,'seed'] <- seeds

###############
```

```
# List the parameters that you want to estimate
###############
estnames <- c("beta1","beta2","beta3","beta4",'beta5',"beta6",'psi',"beta.sd","
    tau","rho","S.0","I.0","R0")

# Initial valued parameters we want to estimate
estICs <- grep("0",estnames,value=TRUE)

# All other parameters we want to estimate
estpars <- setdiff(estnames,estICs)

# Random-walk standard deviation. On the log and logit scale a value of 0.02 is
    about a 2% variation in the parameter value.
# You don't need to use the same rw.sds for each parameter. I have found that 2%
    works well when you have used a Sobol design.
rw.sds <- rep(0.02,times=length(estnames))
names(rw.sds) <- estnames

###############
# set up run index, this is indexing the parameter set that we are using in this
    run
###############
no.sets<- dim(param.sets)[1]
n<- floor(no.sets/no.runs)
start <- 1 + (run-1)*n # n is the number of param set in each run
end<- n + (run-1)*n
if(run==no.runs){end=dim(param.sets)[1]}

###############
```

```
#MIF
###############
write.seq<-rep(c(rep(0,9),1),length.out=dim(param.sets)[1])

for (i in start:end){ # for each parameter set
   current.params <- unlist(param.sets[i,1:18]) # grab parameter set

   for(j in 1:no.MC.reps){ # for each replicate
      results.index<- ((i-1)*no.MC.reps) + j # calculate the index in our results
         data frame where we will put the results of this replicate
      try({
      save.rng.state <- .Random.seed
      set.seed(seeds[results.index]) # set seed

      # Find some decent initial conditions for this parameter set
      mif.ics <- mif(
             my.model, # name of pomp object
             Nmif=10, # the number of MIF replicates
             start=current.params, # starting parameter set
             ivps=estICs, # list of initial conditions
             rw.sd=rw.sds, # random walk standard deviations
             Np=2000, # number of particles
             var.factor=3, # variance factor. The width of the distribution of
                particles at the start of the first MIF iteration will be rw.sd*
                var.factor and the variance will decrease (i.e. cool) with each
                successive MIF iteration.
             cooling.type='hyperbolic', # type of cooling, see ?mif
             cooling.fraction= 0.7, # When cooling.type="hyperbolic", on the n-th
                MIF iteration, the relative perturbation intensity is (s+1)(s+n),
```

```
            where (s+1)/(s+50)=cooling.fraction. cooling.fraction is the
           relative magnitude of the parameter perturbations after 50 MIF
           iterations.
        method='mif2'
     )

mif.dat <- mif(
              my.model,
              Nmif=45,
              start=coef(mif.ics), # we will initialize using the parameter
                  set resulting from "mif.ics" above.
              pars=estpars, # list of parameters we are eastimating
              ivps=estICs, # list of initial conditions we are estimating
              rw.sd=rw.sds,
              Np=2000,
              var.factor=3,
              cooling.type='hyperbolic',
              cooling.fraction= 0.7,
              method='mif2'
            )

.Random.seed <<- save.rng.state

# If you want to plot the results uncomment the next line. The important
   thing is to make sure the likelihood is climbing. If not try varying rw.
   sds, increasing Nmif, increasing Np, and varying the cooling.fraction
# plot(mif.dat)

MIFed.pars<- coef(mif.dat) # MIFed parameter set
```

```
      # calculate the likelihood of the MIFed parameter set
      loglik.mif<- replicate(n=10,logLik(pfilter(my.model,params=MIFed.pars,Np
          =5000,max.fail=40)))
         bl<- mean(loglik.mif)
         loglik.mif.est<- bl+log(mean(exp(loglik.mif-bl)))
         loglik.mif.se <- sd(exp(loglik.mif-bl))/sqrt(length(loglik.mif))/exp(
             loglik.mif.est-bl)

      # put the results in our results data frame
      MIFed.params[results.index,1:18]<-MIFed.pars
      MIFed.params[results.index,'LogLik'] <-loglik.mif.est
      MIFed.params[results.index,'seLogLik'] <-loglik.mif.se
      MIFed.params <- as.data.frame(MIFed.params)
      })
   }
   # write to file every 10 parameter sets
   if(write.seq[i]==1){write.csv(MIFed.params,file=paste(c(output.name,'_run',run,
       '.csv'),collapse=''),row.names=FALSE)}
}
###############
# Write the final output
#
    ##############################################################################################

write.csv(MIFed.params,file=paste(c(output.name,'_run',run,'.csv'),collapse=''),
    row.names=FALSE)
```

D.1.2.3 Pomp Object

```
###############
# Load Covars & Data
###############
# Notice that we have one more observation of the covariates than data.
# This is because we actually start integrting at timestep==0;
# If you don't have the extra covariate data I would suggest backcasting it by
    fitting
# a spline to the covariates using predict(smooth.spline()).
dat<- read.csv('data.csv',header=TRUE)
dat$time <- dat$year+dat$month/12

###############
# Functions we need to transform params
###############
expit=function(x){1.0/(1.0+exp(-x))}
logit= function(p) {log(p/(1-p))};

###############
# Set up a data frame containing the covariates (don't include the data)
###############
covar.table<- data.frame(timestep=seq.int(from=0,by=1,length=dim(dat)[1]))
covar.table$time<- dat$time
covar.table$pop <- dat$pop_size
covar.table$births<- round(dat$births*(1-(dat$vaccine_coverage*0.01)),digits=0) #
     account for vaccination

# If you have a periodic function in your model such as a sine or cosine,
```

```
# you could instead use a periodic b spline. If you use 4+ basis functions
# then it will give you more flexibility than a sine/cosine. However, every basis
# adds an additional parameter to the model. In this case we use a spline with
# 6 basis and a period of 1, for seasonal transmission.
covar.table$seas <- periodic.bspline.basis(covar.table$time,nbasis=6,degree=2,
    period=1)

###############
# Set up data frame with the data to be fit, this dataframe should start at
    timestep==1
###############
monthly.cases<-data.frame(cases=dat$monthly_cases[2:37],time=covar.table$time
    [2:37],timestep=covar.table$timestep[2:37])

###############
# Compile/load model from C-code
###############
if (is.loaded("measles_meas_dens")) dyn.unload("measles_biweekly.so")
system("R CMD SHLIB measles_biweekly.c")
dyn.load("measles_biweekly.so")

###############
# POMP Object
###############
my.model<- pomp(
            data=monthly.cases[c("timestep","cases")], # data are cases and
                associated timestep, starting w/ timestep==1
            times="timestep", # name of time variable in data dataframe
```

```
rprocess=euler.sim("measles_proc_sim",delta.t=1/2,PACKAGE="measles_
   biweekly"), # measles_proc_sim is the name of our process model
   in C-code; measles_biweekly is the name of our C file; see ?
   plugins for integration options
rmeasure="measles_meas_sim", # name of measurement model in C-code
dmeasure="measles_meas_dens", # name of measurement density model in
   C-code
PACKAGE="measles_biweekly", # name of C-file
t0=0, # time 0, this is the time we initialize the model
obsnames="cases", # name of observations in data dataframe
statenames=c("FOI","pS","S","I"), # process model state variable
   names, this vector must match the order of the states in (x[
   stateindex[i]]) in C-code
covarnames=c("pop","births","seas.1"), # covariate names, must match
   order of (covar[covindex[i]]) in C-code
paramnames=c('alpha','tau','beta1','nbeta','beta.sd','delta','R0','
   psi','rho'), # param names, must match order of (p[parindex[0]])
   in C-code, but notice the actual name doesn't have to match for
   instance here we use "alpha" in C-code we call it LOGALPHA, only
   the indexing matters not the exact name
tcovar="timestep", # name of the time variable in the covariate data
   frame
covar=covar.table, # name of the covariate dataframe
 initializer=function(params,t0, ...){ # the initializer allows you
    to specify how to set up initial conditions using your parameter
     set "params"
p <- expit(params)
  covars<- as.data.frame(covar.table)
    with(
```

```
        as.list(p),
          {
x0=c(FOI.0,pS.0,S.0,I.0)
     x0[1]= 0; # we can set our force of infections and prob(
         remaining susceptible) to 0 for timestep==0 bc these values
         don't get used since we don't integrate to get the initial
         conditions, we are estimating them
    x0[2]= 0; # we can set our force of infections and prob(remaining
         susceptible) to 0 for timestep==0 bc these values don't get
        used since we don't integrate to get the initial conditions,
        we are estimating them
 IC_sum=sum(x0[3:4])+as.numeric(c(R0)) # our initial conditions are S
     .0, I.0, and the parameter R0 (intitial recovered), all as
     proportions estimated on logit scale
    x0[3]= round((x0[3]/IC_sum)*covars[1,'pop']) # we eatimate the
        initial portion in each state, but we need the actual number
        of S, I, & R to initialize the model
    x0[4]= round((x0[4]/IC_sum)*covars[1,'pop']) # we eatimate the
        initial portion in each state, but we need the actual number
        of S, I, & R to initialize the model
names(x0)=c("FOI","pS","S","I");
x0
      })
}
)

###############
#END standard code
###############
```

```
###############
```

D.1.2.4 C++ code

```
// dear emacs, please treat this as -*- C++ -*-
#include <R.h>
#include <Rmath.h>
#include <math.h>

//////////////////////

// define parameter sets (if param is bounded between 0-1 use
   LOGIT otherwise LOG)
#define LOGALPHA       (p[parindex[0]]) // I exponent
#define LOGTAU         (p[parindex[1]]) // observation error
#define LOGBETA        (&p[parindex[2]]) // seasonal beta vector
#define NBETA          (p[parindex[3]]) // number of betas
#define LOGBETA_SD     (p[parindex[4]]) // sd of process noise
#define DELTA          (p[parindex[5]]) // natural death rate
#define LOGITfracR0    (p[parindex[6]]) //logit(initial fraction
   recovered)
#define LOGPSI         (p[parindex[7]]) // log(immigration term)
#define LOGITRHO       (p[parindex[8]]) // the report rate

// define process model state variables
#define FOI         (x[stateindex[0]]) // force of infection
#define pS          (x[stateindex[1]]) // prob(remaining in S)
#define S           (x[stateindex[2]]) // susceptible
#define I           (x[stateindex[3]]) // infected
```

```
// define covariates
#define N           (covar[covindex[0]]) // population size
#define BABIES      (covar[covindex[1]]) // raw births
#define SEAS       (&covar[covindex[2]]) // transmission basis
   vector

// observation model state variables
#define MEASLES       (y[obsindex[0]]) // observed MEASLES cases

////////////////////
// // define LOGIT and EXPIT functions
// static double logit (double p) {
//   return log(p/(1-p));
// }

static double expit (double x) {
  return 1.0/(1.0+exp(-x));
}

/////////////////////
// measurement density model for calculating Likelihood
void measles_meas_dens (double *lik, double *y, double *x, double
    *p, int give_log,
                        int *obsindex, int *stateindex, int *
                           parindex, int *covindex,
                        int ncovar, double *covar, double t) {
  double report_rate, tau;
```

```
  tau = exp(LOGTAU);
  report_rate = expit(LOGITRHO);

   *lik = dnorm(MEASLES,report_rate*I,1e-6+tau*I,give_log);

  if (!isfinite(*lik)) Rprintf("measles_meas_dens %lg %lg %lg %lg
      %lg\n",MEASLES,report_rate,tau,I,*lik);
}

////////////////////
// measurement model
void measles_meas_sim (double *y, double *x, double *p,
                       int *obsindex, int *stateindex, int *
                          parindex, int *covindex,
                       int ncovar, double *covar, double t)
{
  double report_rate, tau;
  tau = exp(LOGTAU);
  report_rate = expit(LOGITRHO);


  MEASLES = rnorm(report_rate*I,tau*I);
  if (MEASLES >= 0) {
    MEASLES == MEASLES;
  } else {
    MEASLES = 0;
  }

}
```

```
//////////////////////
// process model
void measles_proc_sim (double *x, const double *p,
                       const int *stateindex, const int *parindex,
                          const int *covindex,
                       int covdim, const double *covar,
                       double t, double dt)
{
 double beta_sd;
 double epsilon;
 double beta;
 double alpha;
 double psi;
 int nbeta = (int) NBETA;
 int j;
 int fail;


 alpha = exp(LOGALPHA);
 psi=exp(LOGPSI);
 beta_sd = exp(LOGBETA_SD);

  for (j = 0, beta = 0; j < nbeta; j++)
    beta += LOGBETA[j]*SEAS[j];
    beta = exp(beta);

  if (beta_sd > 0) {
  epsilon = rnorm(1,beta_sd);
```

```
  } else {
    epsilon = 1;
}



  if(
     isfinite(epsilon)== FALSE ||
     isfinite(alpha)==FALSE ||
     isfinite(FOI)==FALSE ||
     isfinite(pS)==FALSE ||
     isfinite(S)==FALSE ||
     isfinite(I)==FALSE ||
     isfinite(beta)==FALSE
     )
    {
      Rprintf("non finite value in measles_proc_sim\n");
      return;
    }


  FOI=beta*pow(I/N,alpha)*epsilon+psi*epsilon;


  if(FOI < 0){FOI = 0;}
  pS=exp(-dt*(FOI+DELTA));
  S=dt*BABIES+S*pS;
  I=S*(1-pS)*FOI/(DELTA+FOI);
}
```

D.1.2.5 Additional R Code on Simulating Results

```
#Lets look at our model after some MIF results
rm(list=ls())

setwd("/Users/kevinbakker/Desktop/pomp_walkthrough/")
require(pomp)

#read in results from pomp run
data <- read.csv("MIF_results_run1.csv", header=TRUE)
data1 <- data[order(data$LogLik,decreasing=TRUE),];

#normally you don't want to do this, but here we have 49,995 parameter sets with
    '0' liklihood
data2 <- subset(data1,LogLik<0 );
plot(data2$LogLik)

#grab the best fit parameter set
MLE <- subset(data2,LogLik==max(data2$LogLik))
param.set<- unlist(MLE)

# a single stochastic realization
source('pomp_object.R')
sim<- simulate(my.model,param=param.set)

plot(sim);

#plot the observed vs fitted case reports
plot(time(my.model),obs(my.model),type='n',xlab='time',ylab='cases',bty='n')
  for(i in 1:10){
```

```
    sim<- simulate(my.model,params=param.set)
    lines(time(sim),obs(sim),col='grey',lty=2)
  }
  lines(time(my.model),obs(my.model),lwd=2)


#Other options to look at in pomp

  # make a data frame containing the simulated cases and state variables
  sim.dat<- as.data.frame(sim)

  # make a data frame containing the actual data
  case.data<- as.data.frame(my.model)

  # time
  time(sim)

  # simulated cases
  obs(sim)

  # case data stored in pomp object
  obs(my.model)


###############
# Run a bunch of stochastic realizations and plot with the data
###############

minp<- min(obs(my.model))
```

```
maxp<- max(obs(my.model))


plot(time(my.model),obs(my.model),type='n',xlab='time',ylab='cases',bty='n')
  for(i in 1:10){
    sim<- simulate(my.model,params=param.set)
    lines(time(sim),obs(sim),col='grey',lty=2)
  }
  lines(time(my.model),obs(my.model),lwd=2)
```

D.1.3 Chapter 2 Code: Human Birth Seasonality: Latitudinal Gradient and Interplay with Childhood Disease Dynamics

This work (Chapter 2) utilized *pomp* to estimate what effect birth seasonality would have on measles dynamics. All theoretical simulations were done using a time-series SIR (tSIR) model, not included here. Unfortunately I no longer have the R script used, the basic format is the same seen in the above walkthrough and following two chapters of code. I have included the *pomp* object and C++ code. As with all chapters, I used an R script to call the pomp object, which then called the C++ code. Below are all sets of code, run with *pomp* version 0.41-1.

```
#Here is the pomp object for Chapter 2
######
######
# Covars, data, C-code, and POMP Obj
######
STATE=c('New York') # space between words, choose New York or Maryland
######
# Define Model
######
```

```
birth.model <- BIRTH.MODEL #birth model 1,2, 3, or 4; actual births, trend (from
    stl), seasonal, or annual (step function)
alpha.est<-c(0)#logical 0/1, 1 == alpha on and is estimated in MIF
immigration.est<-c(1)#logical 0/1, 1 == immigration on and is estimated
lag<- BIRTH.LAG

model.code <- c('Measles_biweekly_',STATE,'_',birth.model,alpha.est,immigration.
    est,lag)
######
# Load Covars & Data; case data from 1931-1964, 33yrs of data
######
table1 <- read.csv('births_NYC_Baltimore.csv',header=TRUE)

if(STATE=='New York'){
   ny<- read.table('New_York_measles_cases.txt',header=TRUE)
   ny<-subset(ny,month<1962.999) #measles vaccine licensed in 1963
   ny<-subset(ny,month>=1931) #we only have birth data for 1931+
   ny$month.no<- 1:384
}
if(STATE=='Maryland'){
   bt<- read.table('Baltimore_measles_cases.txt',header=TRUE)
   bt<-subset(bt,month<1962.999)
   bt<-subset(bt,month>=1931)
   bt$month.no<- 1:384
}
######
# Functions we need to transform params
######
expit=function(x){1.0/(1.0+exp(-x))}
```

```
logit= function(p) {log(p/(1-p))};
######
#SET UP COVARS: Pop and seasonality splines
######
demog <- subset(table1,state==STATE)
demog$month <- ordered(demog$month,levels=month.name)
demog$time <- with(demog,year+as.integer(month)/12)
demog<-subset(demog,year<1963)
if(STATE=='New York'){ny$month<-demog$time}
if(STATE=='Maryland'){bt$month<-demog$time}

census <- subset(demog,month=="January")
covar.table <- data.frame(time=c(1931,demog$time,1963.083))
covar.table$timestep <- seq.int(from=0,by=1,length=length(covar.table$time))
covar.table$pop <- predict(smooth.spline(x=census$year[-which(is.na(census$city_
    pop)==TRUE)],y=census$city_pop[-which(is.na(census$city_pop)==TRUE)]),x=covar.
    table$time)$y
covar.table$seas <- periodic.bspline.basis(covar.table$time,nbasis=6,degree=2,
    period=1)
######
# Monthly Cases time series for Measles in NYC or Baltimore
######
if(STATE=='New York'){
   monthly.cases<-data.frame(cases=ny$cases,time=ny$month,timestep=seq(from=1,to
       =384,by=1))
}
if(STATE=='Maryland'){
   monthly.cases<-data.frame(cases=bt$cases,time=bt$month,timestep=seq(from=1,to
       =384,by=1))
```

```
}
######
# Birth data
######
births<-round((c(demog$birth_rate_per1000[1],demog$birth_rate_per1000,demog$birth
    _rate_per1000[384])/1000)*covar.table$pop)
birth.ts.pre.bboom <- ts(births[2:181],frequency=12)
birth.ts.bboom <- ts(births[182:385],frequency=12)
######
# Birth stl()
######
stl1931_45 <- stl(birth.ts.pre.bboom,s.window="periodic", robust=TRUE, inner=1,
    outer=15, na.action=na.fail,l.window=13)
stl1946_64 <-stl(birth.ts.bboom,"periodic", robust=TRUE, inner=1, outer=15, na.
    action=na.fail,l.window=13)

stl1931_45 <- as.data.frame(stl1931_45$time.series)
stl1946_64<- as.data.frame(stl1946_64$time.series)

stl1931_45$sum <- stl1931_45$seasonal+stl1931_45$trend+stl1931_45$remainder
stl1946_64$sum <- stl1946_64$seasonal+stl1946_64$trend+stl1946_64$remainder



stl1931_45$scaled.seasonal.births <-round(mean(stl1931_45$trend)+stl1931_45$
    seasonal)
stl1946_64$scaled.seasonal.births <-round(mean(stl1946_64$trend)+stl1946_64$
    seasonal)

stl1931_45$trend<- round(stl1931_45$trend)
```

```
stl1946_64$trend<- round(stl1946_64$trend)
######
#Turns ON correct birth COVAR
######
if(birth.model==1){covar.table$births <- births}
if(birth.model==2){covar.table$births <- c(stl1931_45$trend[1],stl1931_45$trend,
    stl1946_64$trend,stl1946_64$trend[204])}
if(birth.model==3){covar.table$births <- c(stl1931_45$scaled.seasonal.births[1],
    stl1931_45$scaled.seasonal.births,stl1946_64$scaled.seasonal.births,stl1946_64
    $scaled.seasonal.births[204])}
if(birth.model==4){
   temp.births<- births[2:385]
   annual.births<- rep(NA,384)
   for(i in 0:31){
      jan<- 1+(i*12)
      dec<- 12+(i*12)
      current<- temp.births[jan:dec]
      total<- sum(current)
      monthly<- round(total/12,digits=0)
      annual.births[jan:dec]<- monthly
           }
   annual.births<- c(annual.births[1],annual.births,annual.births[384])
   covar.table$births <- annual.births
}

######
# lag the births according to the lag model - Note, the lag is only meant for the
     models with birth seasonality, i.e. models 1 & 3
######
```

```
if(birth.model==1|birth.model==3){
   if(lag>0){
      length<- dim(covar.table)[1]
      to.remove<-seq(from=(length-(lag-1)),to=length,by=1)
      replacement<- covar.table$births[-to.remove]
      nas<- rep(0,lag) # must use 0s instead of NAs
      replacement<- c(nas,replacement)
      covar.table$births<- replacement
   }
}

######
# now put everything in C for speed - this is the revised c-code w/ beta5=beta1-
    amp
######
if (is.loaded("measles_meas_dens")) dyn.unload("measles_biweekly_amp.so")
system("R CMD SHLIB measles_biweekly_amp.c")
dyn.load("measles_biweekly_amp.so")
######
#POMP Object, this is a modified POMP object, includes the new param "amp"
######
full.measles<- pomp(
             data=monthly.cases[c("timestep","cases")],
             times="timestep",
             rprocess=euler.sim("measles_proc_sim",delta.t=1/2,PACKAGE="measles_
                biweekly_amp"),
             rmeasure="measles_meas_sim",
             dmeasure="measles_meas_dens",
             PACKAGE="measles_biweekly_amp",
```

```
    t0=0,
    obsnames="cases",
    statenames=c("FOI","pS","S","I"),
    covarnames=c("pop","births","seas.1"),
    paramnames=c('alpha','tau','beta1','nbeta','beta.sd','delta','R0','
        psi','rho','amp'),
    tcovar="timestep",
    covar=covar.table,
  initializer=function(params,t0, ...){
    p <- expit(params)
    covars<- as.data.frame(covar.table)
  with(
            as.list(p),
              {
                 x0=c(FOI.0,pS.0,S.0,I.0)
   x0[1]= 0;
   x0[2]= 0;
     IC_sum=sum(x0[3:4])+as.numeric(c(R0))
     x0[3]= round((x0[3]/IC_sum)*covars[1,'pop'])
     x0[4]= round((x0[4]/IC_sum)*covars[1,'pop'])
                 names(x0)=c("FOI","pS","S","I");
              x0
            })
    }
    )

measles <- window(full.measles,start=205)
measles.ic<- window(full.measles,start=205,end=210)
measles.ic.extended<- window(full.measles,start=205,end=228) #extended 2yr
```

```
    section for initial conditions


measles.1949<- window(full.measles,start=217)
measles.ic.1949<- window(full.measles,start=217,end=222)
measles.ic.extended.1949<- window(full.measles,start=217,end=240) #extended 2yr
    section for initial conditions
######
#END standard code
######
######
```

```
// dear emacs, please treat this as -*- C++ -*-

#include <R.h>
#include <Rmath.h>
#include <math.h>

#define LOGALPHA        (p[parindex[0]]) // I exponent
#define LOGTAU          (p[parindex[1]]) // observation error
#define LOGBETA         (&p[parindex[2]]) // seasonal beta vector
#define NBETA           (p[parindex[3]]) // number of betas
#define LOGBETA_SD      (p[parindex[4]]) // sd of process noise
#define DELTA           (p[parindex[5]]) // natural death rate
#define LOGITfracR0     (p[parindex[6]]) //logit(initial fraction
    recovered)
#define LOGPSI          (p[parindex[7]]) // log(immigration term)
#define LOGITRHO        (p[parindex[8]]) // the report rate
```

```
#define AMP             (p[parindex[9]]) // amplitude of seas
   transmission, LOGBETA1-LOGBETA5

#define FOI          (x[stateindex[0]]) // force of infection
#define pS           (x[stateindex[1]]) // prob(remaining in S)
#define S            (x[stateindex[2]]) // susceptible
#define I            (x[stateindex[3]]) // infected

#define N            (covar[covindex[0]]) // population size
#define BABIES       (covar[covindex[1]]) // raw births
#define SEAS        (&covar[covindex[2]]) // transmission basis
   vector

#define MEASLES        (y[obsindex[0]]) // observed MEASLES cases

static double logit (double p) {
  return log(p/(1-p));
}

static double expit (double x) {
  return 1.0/(1.0+exp(-x));
}

void measles_meas_dens (double *lik, double *y, double *x, double
    *p, int give_log,
                        int *obsindex, int *stateindex, int *
                           parindex, int *covindex,
                        int ncovar, double *covar, double t) {
  double report_rate, tau;
```

```
  tau = exp(LOGTAU);
  report_rate = expit(LOGITRHO);
   *lik = dnorm(MEASLES,report_rate*I,1e-6+tau*I,give_log);
  if (!isfinite(*lik)) Rprintf("measles_meas_dens %lg %lg %lg %lg
      %lg\n",MEASLES,report_rate,tau,I,*lik);
}

void measles_meas_sim (double *y, double *x, double *p,
                       int *obsindex, int *stateindex, int *
                          parindex, int *covindex,
                       int ncovar, double *covar, double t) {
  double report_rate, tau;
  tau = exp(LOGTAU);
  report_rate = expit(LOGITRHO);


  MEASLES = rnorm(report_rate*I,tau*I);
  if (MEASLES >= 0) {
    MEASLES == MEASLES;
  } else {
    MEASLES = 0;
  }

}

// our measles model
void measles_proc_sim (double *x, const double *p,
                       const int *stateindex, const int *parindex,
                         const int *covindex,
```

```
                    int covdim, const double *covar,
                    double t, double dt)
{

 double beta_sd;
 double epsilon;
 double beta;
 double alpha;
 double psi;
 int nbeta = (int) NBETA;
 int j;
 int fail;



 alpha = exp(LOGALPHA);
 psi = exp(LOGPSI);
 beta_sd = exp(LOGBETA_SD);

// for (j = 0, beta = 0; j < nbeta; j++)
  //   beta += LOGBETA[j]*SEAS[j];
  //   beta = exp(beta);

 beta = LOGBETA[0]*SEAS[0] + LOGBETA[1]*SEAS[1] + LOGBETA[2]*SEAS
    [2] + LOGBETA[3]*SEAS[3] + (LOGBETA[0]-AMP)*SEAS[4] + LOGBETA
    [5]*SEAS[5];
 beta = exp(beta);

  if (beta_sd > 0) {
  epsilon = rnorm(1,beta_sd);
```

```
  } else {
    epsilon = 1;
}



  if(
     isfinite(epsilon)== FALSE ||
     isfinite(alpha)==FALSE ||
     isfinite(FOI)==FALSE ||
     isfinite(pS)==FALSE ||
     isfinite(S)==FALSE ||
     isfinite(I)==FALSE ||
     isfinite(beta)==FALSE
     )
    {
      Rprintf("non finite value in measles_proc_sim\n");
      return;
    }


  FOI=beta*pow(I/N,alpha)*epsilon+psi*epsilon;
  if(FOI < 0){FOI = 0;}
  pS=exp(-dt*(FOI+DELTA));
  S=dt*BABIES+S*pS;
  I=S*(1-pS)*FOI/(DELTA+FOI);
}
```

D.1.4 Chapter 3 Code: Digital Epidemiology Reveals Global Childhood Disease Seasonality and the Effects of Immunization

D.1.4.1 Introduction

This work (Chapter 3) utilized *pomp* to estimate statistical models, which can be seen in Table B.1. As we can see from this table, the most complex model is B.7, and all the others are nested within that model. To fit any of the other models, including our best fit (Model & Equation B.1), one only needs to remove fit parameters. As with all chapters, I used an R script to call the pomp object, which then called the C++ code. All code is also inlcuded in an R package created for this publication, that R package is called 'chickenpox.pack' Below are all three sets of code, all run with *pomp* version 0.65.1.

D.1.4.2 R script

```
#Clear R
rm(list=ls())

#If using clusters, comment out below file, if running locally, set working
   directory
# setwd('/Users/KMBakker/....')

require('subplex',lib.loc='/home2/bakkerke/pomp_develop');
require('nloptr',lib.loc='/home2/bakkerke/pomp_develop');
require('mvtnorm',lib.loc='/home2/bakkerke/pomp_develop');
require('deSolve',lib.loc='/home2/bakkerke/pomp_develop');
require('coda',lib.loc='/home2/bakkerke/pomp_develop');
require('pomp',lib.loc='/home2/bakkerke/pomp_develop');

#below two lines are for cluster, comment out if running locally
```

```
args<- (commandArgs(TRUE)); # command line arguments that I will enter manually
run<- as.numeric(args[1]);# Split up the work into multiple runs/cores

# no.runs<- c(10) # This would be the number of cores you are going to use. I
    usually split analysis like this among 10-80 cores.
no.nodes<- c(8)
# You will need to do speed testing to decide how many cores you want to use. I
    set the number of cores to 10000 here so that this code will run only 5 param
    sets and finish in a few minutes.
# Each phase of analyses will take a different amount of time depending on how
    many observations, parameter sets, cores, the numer of replicates, particles
    per replicate, etc.

# no.MC.reps <- c(10) # Number of Monte Carlo replicates per parameter set. I
    start with 1 when fitting sobol parameters, and gradually increase as my
    likelihoods get better
no.reps<- 1 # mif Monte Carlo reps per param set, see above line.
output.name<- c("CPOX") # Name of your output csv file
job<- 1;
################
# load pomp object
################
source('pomp_obj_seas_G.R')

################
# load Sobol design parameter sets
################
param.sets<- read.csv(file='sobol.csv',header=TRUE)
param.sets$FOI.0<- 0
```

```
param.sets$I.0<- 0
param.sets<- param.sets[c("alpha","tau","beta1","beta2","beta3","beta.sd","omega"
    ,"rho","FOI.0","I.0")]

################
#Set up storage space for results & generate seeds (seeds make your results
    reproducible)
################
total.params<- dim(param.sets)[1]*no.reps #the total number of params output will
     be the number of start params multiplied by the number of mif replicates
MIFed.params<-mat.or.vec(total.params,dim(param.sets)[2]+3)
MIFed.params<-as.data.frame(MIFed.params)
names(MIFed.params)<- c(names(param.sets),'LogLik','seLogLik','seed');
seeds <- ceiling(runif(n=total.params,min=1,max=2^30))
MIFed.params[,'seed'] <- seeds

################
# List the parameters that you want to estimate
################
# estnames <- c("alpha","tau","beta1","beta2",'beta.sd',"rho","omega")
estnames <- c("tau","beta1","beta3","beta.sd","omega")

# Initial valued parameters we want to estimate
estICs <- grep("0",estnames,value=TRUE)

# All other parameters we want to estimate
estpars <- setdiff(estnames,estICs)
```

```
# Random-walk standard deviation. On the log and logit scale a value of 0.02 is
    about a 2% variation in the parameter value.
# You dont need to use the same rw.sds for each parameter. I have found that 2%
    works well when you have used a Sobol design.
#################
# set up random walk standard deviations
#################

rw.sds <- rep(0.2,times=length(estnames))
names(rw.sds) <- estnames

#################
# set up run index, this is indexing the parameter set that we are using in this
    run
################
n<- floor(dim(param.sets)[1]/no.nodes)
start <- 1 + (job-1)*n #formula is: 1+(job-1)*n| n is the numer of param set in
    each job/run
end<-  n + (job-1)*n #formula is: n + (job-1)*n | n is the number of param sets in
     each job/run
if(job==no.nodes){end=dim(param.sets)[1]} # if job== the last set then end== the
    last param set

#################
#MIF
#################
write.seq<-rep(c(rep(0,9),1),length.out=dim(param.sets)[1]) #write every 10th
    parameter set
```

```
#the two loops commented out are run initially on the clusters to create .o and .
    so files from our C++ code. After these have been created (and I see how long
    single mif iteration takes) I change the loops.
#Also, when starting with sobol, I only run mif once (not the 4 times seen below)
    , but as I get closer, I run mif multiple times in a loop to get closer to the
     MLE

# for (i in 1:3){ # for each parameter set
for (i in start:end){ # for each parameter set
   current.params <- unlist(param.sets[i,1:length(param.sets[1,])]) # grab
       parameter set
# for(j in 1:2){ # for each replicate #how many times you want to repeat each
    parameter set
   for(j in 1:no.reps){ # for each replicate
      results.index<- ((i-1)*no.reps) + j # calculate the index in our results
          data frame where we will put the results of this replicate
      try({
      save.rng.state <- .Random.seed
      set.seed(seeds[results.index]) # set seed

      mif.dat1 <- mif(
                   my.model,
                   Nmif=100,
                   start=current.params, # we will initialize using the parameter
                        set resulting from "mif.ics" above.
                   pars=estpars, # list of parameters we are estimating
                   # ivps=estICs, # list of initial conditions we are estimating
                   rw.sd=rw.sds,
                   Np=2000,
```

```
                var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.7,
                method='mif2',
                max.fail=108
              )
MIFed.pars1<- coef(mif.dat1) # MIFed parameter set

mif.dat2 <- mif(
                my.model,
                Nmif=100,
                start=MIFed.pars1, # we will initialize using the parameter
                    set resulting from "mif.ics" above.
                pars=estpars, # list of parameters we are estimating
                # ivps=estICs, # list of initial conditions we are estimating
                rw.sd=(0.5*rw.sds),
                Np=2000,
                var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.7,
                method='mif2',
                max.fail=108
              )
MIFed.pars2<- coef(mif.dat2) # MIFed parameter set

mif.dat3 <- mif(
                my.model,
                Nmif=100,
```

```
                start=MIFed.pars2, # we will initialize using the parameter
                    set resulting from "mif.ics" above.
                pars=estpars, # list of parameters we are estimating
                # ivps=estICs, # list of initial conditions we are estimating
                rw.sd=(0.25*rw.sds),
                Np=2000,
                var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.7,
                method='mif2',
                max.fail=108
              )
MIFed.pars3<- coef(mif.dat3) # MIFed parameter set

mif.dat4 <- mif(
                my.model,
                Nmif=100,
                start=MIFed.pars3, # we will initialize using the parameter
                    set resulting from "mif.ics" above.
                pars=estpars, # list of parameters we are estimating
                # ivps=estICs, # list of initial conditions we are estimating
                rw.sd=(0.25*rw.sds),
                Np=2000,
                var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.7,
                method='mif2',
                max.fail=108
              )
```

```
      MIFed.pars4<- coef(mif.dat4) # MIFed parameter set
      .Random.seed <<- save.rng.state


      # calculate the likelihood of the MIFed parameter set
      loglik.mif<- replicate(n=10,logLik(pfilter(my.model,params=MIFed.pars4,Np
          =5000,max.fail=108)))
         bl<- mean(loglik.mif)
         loglik.mif.est<- bl+log(mean(exp(loglik.mif-bl)))
         loglik.mif.se <- sd(exp(loglik.mif-bl))/sqrt(length(loglik.mif))/exp(
             loglik.mif.est-bl)

      # put the results in our results data frame
      # MIFed.params[results.index,1:10]<-MIFed.pars
      MIFed.params[results.index,1:length(param.sets[1,])]<-MIFed.pars4
      MIFed.params[results.index,'LogLik'] <-loglik.mif.est
      MIFed.params[results.index,'seLogLik'] <-loglik.mif.se
      MIFed.params <- as.data.frame(MIFed.params)
      })
   }
   # write to file every 10 parameter sets
   if(write.seq[i]==1){write.csv(MIFed.params,file=paste(c(output.name,'_run',run,
       '.csv'),collapse=''),row.names=FALSE)}
}
################
# Write the final output
###############
write.csv(MIFed.params,file=paste(c(output.name,'_run',run,'.csv'),collapse=''),
    row.names=FALSE)
```

D.1.4.3 Pomp Object

```
require(pomp)
setwd('/Users/KMBakker/....')

################
# Load Covars & Data
################
# load the chickenpox and shingles data for Australia
#Data and covariate tables are publicly available on the PNAS website, and also
    included in the chickenpox.pack R package.
dat<- read.csv('CPox_Shing_AUS06_15.csv',header=TRUE)
dat$time<- dat$YEAR + dat$MONTH/12
names(dat)<- c('year','month','chickenpox','shingles','time')
dat<- subset(dat,chickenpox>=0)

# load the google chickenpox data
goog<- read.csv('google_trends_data.csv',header=TRUE)
goog$mo<- NA
for(i in 1:dim(goog)[1]){
  month<- goog$Month[i]
  if(month=='Jan'){goog$mo[i]<- 1}
  if(month=='Feb'){goog$mo[i]<- 2}
  if(month=='Mar'){goog$mo[i]<- 3}
  if(month=='Apr'){goog$mo[i]<- 4}
  if(month=='May'){goog$mo[i]<- 5}
  if(month=='Jun'){goog$mo[i]<- 6}
```

```
  if(month=='Jul'){goog$mo[i]<- 7}
  if(month=='Aug'){goog$mo[i]<- 8}
  if(month=='Sep'){goog$mo[i]<- 9}
  if(month=='Oct'){goog$mo[i]<- 10}
  if(month=='Nov'){goog$mo[i]<- 11}
  if(month=='Dec'){goog$mo[i]<- 12}
}
goog$time<- goog$Year + goog$mo/12
goog<- goog[order(goog$time,decreasing=FALSE),]
goog$time<- round(goog$time,digits=3)

# load the population data
pop<- read.csv('Australia_pop.csv',header=TRUE)
pop2010<- sum(subset(pop,year==2010)$pop) # data from Australian National
    University (http://adsri.anu.edu.au/demo-stats/aust)

# vaccine
vacc<- data.frame(year=2006:2010)
vacc$vaccine[which(vacc$year==2006)]<- 0.3 # account for vaccination from Heywood
     et al 2014 (http://www.scielosp.org/pdf/bwho/v92n8/0042-9686-bwho-92-08-593.
    pdf)
vacc$vaccine[which(vacc$year==2007)]<- 0.82 # account for vaccination
vacc$vaccine[which(vacc$year==2008)]<- 0.88 # account for vaccination
vacc$vaccine[which(vacc$year==2009)]<- 0.90 # account for vaccination
vacc$vaccine[which(vacc$year==2010)]<- 0.91
################
# Functions we need to transform params
################
expit=function(x){1.0/(1.0+exp(-x))}
```

```
logit= function(p) {log(p/(1-p))};


################
# Set up a data frame containing the covariates (dont include the data)
################
covar.table<- data.frame(timestep=seq.int(from=0,by=1,length=dim(dat)[1]+1))
covar.table$time<- c(2006,round(dat$time,digits=3))
covar.table$children <- pop2010
# account for vaccination, still need data form 2010--2015
covar.table$vaccine<- predict(smooth.spline(vacc$year,vacc$vaccine),covar.table$
    time)$y


# plot(covar.table$time,covar.table$vaccine,type='l')
# points(vacc$year,vacc$vaccine)


for(i in 1:dim(covar.table)[1]){
  TIME<- covar.table$time[i]
  if(TIME<=2015){
    covar.table$googleLAG0[i]<- goog$Chicken.Pox.Australia[which(goog$time==TIME)]
    covar.table$googleLAG1[i]<- goog$Chicken.Pox.Australia[which(goog$time==TIME)
        -1]
    covar.table$googleLAG2[i]<- goog$Chicken.Pox.Australia[which(goog$time==TIME)
        -2]
  }
  else{covar.table$googleLAG0[i]<- NA
    covar.table$googleLAG1[i]<- NA
    covar.table$googleLAG2[i]<- NA
  }
}
```

```
covar.table[,c('googleLAG0','googleLAG1','googleLAG2')]
covar.table<- subset(covar.table,time<= 2015)

################
# Set up data frame with the data to be fit, this dataframe should start at
    timestep==1
################
monthly.cases<-data.frame(cases=subset(dat,time<=2015)$chickenpox,time=round(
    subset(dat,time<=2015)$time,digits=3))
monthly.cases$timestep<- 1:108

# # the google trends and the data align well
# par(plt=c(0.1,0.9,0.6,0.9))
# plot(monthly.cases$time,monthly.cases$cases,type='l')
# par(plt=c(0.1,0.9,0.1,0.5),new=TRUE)
# plot(covar.table$time,covar.table$googleLAG0,type='l',col='red')

################
# Compile/load model from C-code
################
if (is.loaded("chickenpox_meas_dens")) dyn.unload("chickenpox_seas_G.so")
system("R CMD SHLIB chickenpox_seas_G.c")
dyn.load("chickenpox_seas_G.so")

################
# POMP Object
################
my.model<- pomp(
```

```
data=monthly.cases[c("timestep","cases")], # data are cases and
    associated timestep, starting w/ timestep==1
times="timestep", # name of time variable in data dataframe
rprocess=euler.sim("chickenpox_proc_sim",delta.t=1,PACKAGE="
    chickenpox_seas_G"), # chickenpox_proc_sim is the name of our
    process model in C-code; chickenpox.c is the name of our C file;
    see ?plugins for integration options
rmeasure="chickenpox_meas_sim", # name of measurement model in C-code
dmeasure="chickenpox_meas_dens", # name of measurement density model
    in C-code
PACKAGE="chickenpox_seas_G", # name of C-file
t0=0, # this is the time we initialize the model
obsnames="cases", # name of observations in data dataframe
statenames=c("FOI","I"), # process model state variable names, this
    vector must match the order of the states in (x[stateindex[i]])
    in C-code
covarnames=c("children","vaccine","googleLAG0","googleLAG1","
    googleLAG2"), # covariate names, must match order of (covar[
    covindex[i]]) in C-code
paramnames=c('alpha','tau','beta1','beta2','beta3','beta.sd','omega',
    'rho'), # param names, must match order of (p[parindex[0]]) in C-
    code, but notice the actual name doesnt have to match for
    instance here we use "alpha" in C-code we call it LOGALPHA, only
    the indexing matters not the exact name
tcovar="timestep", # name of the time variable in the covariate data
    frame
covar=covar.table, # name of the covariate dataframe
 initializer=function(params,t0, ...){ # the initializer allows you
     to specify how to set up initial conditions using your parameter
```

```
        set "params"
     p <- expit(params)
    covars<- as.data.frame(covar.table)
     with(as.list(p),{
         x0=c(FOI.0,I.0)
           x0[1]= 0;
        x0[2]= 0;
         names(x0)=c("FOI","I");
             x0
        })
    }
)
```

D.1.4.4 C++ code

```
// dear emacs, please treat this as -*- C++ -*-

#include <R.h>
#include <Rmath.h>
#include <math.h>

//////////////////////////////////////////////////////////////////
   //

// define parameter sets (if param is bounded between 0-1 use
   LOGIT otherwise LOG)
#define LOGALPHA       (p[parindex[0]]) // exponent for lag1
   google trends
```

```
#define LOGTAU          (p[parindex[1]]) // observation variance
   parameter
#define LOGBETA1        (p[parindex[2]]) // scaler for lag1 google
    trends
#define LOGBETA2        (p[parindex[3]]) // scaler for lag2 google
    trends
#define LOGBETA3        (p[parindex[4]]) // scaler for lag2 google
    trends
#define LOGBETA_SD      (p[parindex[5]]) // sd of process noise
#define LOGITRHO        (p[parindex[7]]) // the mean report rate
#define LOGOMEGA        (p[parindex[6]]) // the phase of the
   cosine function for seasonal forcing
#define M_PI            3.14159265358979323846  /* pi */

// define process model state variables
#define FOI          (x[stateindex[0]]) // force of infection
#define I            (x[stateindex[1]]) // infected

// define covariates
#define CHILDREN     (covar[covindex[0]]) // population size under
    15 yrs of age
#define VACCINE      (covar[covindex[1]]) // percent vaccinated
   against VZV under 15
#define Tlag0        (covar[covindex[2]]) // percent vaccinated
   against VZV under 15
#define Tlag1        (covar[covindex[3]]) // percent vaccinated
   against VZV under 15
#define Tlag2        (covar[covindex[4]]) // percent vaccinated
   against VZV under 15
```

```
// observation model state variables
#define CHICKENPOX       (y[obsindex[0]]) // observed CHICKENPOX
   cases

////////////////////////////////////////////////////////////////////
   //
// define LOGIT and EXPIT functions
// static double logit (double p) {
//   return log(p/(1-p));
// }

static double expit (double x) {
  return 1.0/(1.0+exp(-x));
}

////////////////////////////////////////////////////////////////////
   //
// measurement density model for calculating Likelihood
void chickenpox_meas_dens (double *lik, double *y, double *x,
   double *p, int give_log,
                           int *obsindex, int *stateindex, int *
                              parindex, int *covindex,
                           int ncovar, double *covar, double t) {

  double report_rate, tau;
  double tol = 1.0e-18;
```

```
 tau = exp(LOGTAU);
 // tau = 0;
  report_rate = expit(LOGITRHO);

  if(CHICKENPOX > 0.0){
   *lik = pnorm(CHICKENPOX+0.5,report_rate*I,tau*I,1,0) - pnorm(
      CHICKENPOX-0.5,report_rate*I,tau*I,1,0)+ tol;
  } else{
   *lik = pnorm(CHICKENPOX+0.5,report_rate*I,tau*I,1,0)+ tol;
  }
  if (give_log) *lik = log(*lik);
  if (!isfinite(*lik)) Rprintf("chickenpox_meas_dens %lg %lg %lg
     %lg %lg\n",CHICKENPOX,report_rate,tau,I,*lik);
}

//////////////////////////////////////////////////////////////////
   //
// measurement model
void chickenpox_meas_sim (double *y, double *x, double *p,
                          int *obsindex, int *stateindex, int *
                             parindex, int *covindex,
                          int ncovar, double *covar, double t) {

  double report_rate, tau;
  tau = exp(LOGTAU);
  //tau = 0;

  report_rate = expit(LOGITRHO);
```

```
  CHICKENPOX = rnorm(report_rate*I,tau*I);
  if (CHICKENPOX >= 0) {
    CHICKENPOX = nearbyint(CHICKENPOX);
        } else {CHICKENPOX = 0;}

}

//////////////////////////////////////////////////////////////////
   //
// process model
void chickenpox_proc_sim (double *x, const double *p,
                    const int *stateindex, const int *parindex,
                       const int *covindex,
                    int covdim, const double *covar,
                    double t, double dt)
{
 double beta_sd;
 double epsilon;
 double beta1;
 double beta2;
 double beta3;
 double alpha;
 // int fail;
 double scale;
 double omega;


 alpha = exp(LOGALPHA);
```

```
beta_sd = exp(LOGBETA_SD);
beta1 = exp(LOGBETA1);
beta2 = exp(LOGBETA2);
beta3 = exp(LOGBETA3);
omega = exp(LOGOMEGA);


 if (beta_sd > 0) {
   scale = pow(beta_sd,2);
   epsilon = rgamma(1/scale,scale);
       } else {epsilon = 1;}

 if(
    isfinite(epsilon)== FALSE ||
    isfinite(alpha)==FALSE ||
    isfinite(FOI)==FALSE ||
    isfinite(I)==FALSE ||
    isfinite(beta1)==FALSE ||
    isfinite(beta2)==FALSE
    )
   {
     Rprintf("non finite value in chickenpox_proc_sim\n");
     return;
   }

 FOI = (beta1*cos((2*M_PI/12)*(t+omega))*Tlag1 + beta3)*epsilon;
 if(FOI < 0){FOI = 0;}
 I = CHILDREN*FOI;
```

```
}
```

D.1.5 Chapter 4 Code: The Underpinnings of Herpesvirus Dynamics: Transmission & Reactivation of Varicella Zoster Virus

D.1.5.1 Introduction

As my coding abilities increased, I changed my R scripts and C++ code to include loops, so I didn't have to run different code for each model test. In this work (Chapter 4), I have been using *pomp* version 1.7 to fit mechanistic models. Table C.1 includes all model variations fit to-date. Different variations can be fit based on simple changes to the R script, which can be seen in the comments below. This code specifically runs model 1, which biologically tests whether the same seasonal mechanism (B-Spline) drives both transmission and reactivation of VZV, without antibody-boosting. As with all chapters, I used an R script to call the pomp object, which then called the C++ code.

D.1.5.2 R script

```
rm(list=ls())


# require('subplex',lib.loc='/home2/bakkerke/pomp_develop');
# require('nloptr',lib.loc='/home2/bakkerke/pomp_develop');
# require('mvtnorm',lib.loc='/home2/bakkerke/pomp_develop');
# require('deSolve',lib.loc='/home2/bakkerke/pomp_develop');
# require('coda',lib.loc='/home2/bakkerke/pomp_develop');
# require('pomp',lib.loc='/home2/bakkerke/pomp_develop');


###next three lines needed for local R session
# setwd("/Users/kevinbakker/Desktop/VZV_dynamics")
require(pomp);
```

```
# run=1
###

param.sets<- read.csv(file='VZV1_run4.csv',header=TRUE);
#SET THESE PRIOR TO ALL MODEL ITERATIONS!
#the modelnumber indicates which models to fit without reference to antibody
    boosting (i.e. model 1 will fit both model 1 and 2 from Table C.1)
# 1 = models 1,2, 2 = models 3,4, 3 = models 5,6, 4 = models 7,8
modelnumber <- 1; # range for model 1-4
#below is whether or not boosting will be included.
#Follwing Table C.1, if 0 is set, then the above models will be 1,3,5, or 7 will
    be run, depending on what model number is selected.
#However if boosting IS included (below setting =1), then models 2,3,6, or 8 will
     be run, depending on what model number is selected.
boosting <- 0;  # range for boost 0-1 (no/yes)
param.sets$model <- modelnumber;
param.sets$boost <- boosting;

args <- (commandArgs(TRUE)); # command line arguments that I will enter manually
run <- as.numeric(args[1]); # for FLUX#
no.runs <- c(52); # number of cores (this started around 800 when fitting sobol
    parameters, now I use a core per parameter set)
no.MC.reps <- c(35); #number of times I will repeat each parameter set, this used
     to be 1, and has gone as high as 200
output.name<- c("VZV_results");

# load our pomp object
source('VZV_PO_1.R')
```

```
#total searches
total.params<- dim(param.sets)[1]*no.MC.reps;
MIFed.params<-mat.or.vec(total.params,dim(param.sets)[2]+3);
MIFed.params<-as.data.frame(MIFed.params);
names(MIFed.params)<- c(names(param.sets),'LogLik','seLogLik','seed');
seeds <- ceiling(runif(n=total.params,min=1,max=2^30));
MIFed.params[,'seed'] <- seeds;

#Which parameters to estimate depending on model number
estnames <- c()
if(modelnumber ==1 && boosting == 0){
estnames <- c('zeta1','zeta2','zeta3','zeta4','zeta5','zeta6','rhovz','tauvz','
    tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','IHZ.0','L2.0');
  }else if(modelnumber == 1 && boosting == 1){
estnames <- c('zeta1','zeta2','zeta3','zeta4','zeta5','zeta6','rhovz','tauvz','
    tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','IHZ.0','L2.0','eta','omicron'
    );
  }else if(modelnumber == 2 && boosting == 0){
estnames <- c('zeta1','zeta2','zeta3','zeta4','zeta5','zeta6','rhovz','tauvz','
    tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','IHZ.0','L2.0','zeta1B','
    zeta2B','zeta3B','zeta4B','zeta5B','zeta6B');
  }else if(modelnumber == 2 && boosting == 1){
estnames <- c('zeta1','zeta2','zeta3','zeta4','zeta5','zeta6','rhovz','tauvz','
    tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','IHZ.0','L2.0','zeta1B','
    zeta2B','zeta3B','zeta4B','zeta5B','zeta6B','eta','omicron');
  }else if(modelnumber == 3 && boosting == 0){
estnames <- c('zeta1','zeta2','zeta3','zeta4','zeta5','zeta6','rhovz','tauvz','
    tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','IHZ.0','L2.0','a','nu','xi');
```

```
  }else if(modelnumber == 3 && boosting == 1){
estnames <- c('zeta1','zeta2','zeta3','zeta4','zeta5','zeta6','rhovz','tauvz','
    tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','IHZ.0','L2.0','eta','omicron'
    ,'a','nu','xi');
  }else if(modelnumber == 4 && boosting == 0){
estnames <- c('rhovz','tauvz','tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','
    IHZ.0','L2.0','chi','upsilon','sigma','a','nu','xi');
  }else if(modelnumber == 4 && boosting == 1){
estnames <- c('rhovz','tauvz','tauhz','m1','m2','Beta_SD','S.0','IVZ.0','L1.0','
    IHZ.0','L2.0','eta','omicron','chi','upsilon','sigma','a','nu','xi');
  }else{ print("You are a fucking idiot Kevin")}
# Initial valued parameters we want to estimate
estICs <- grep("0",estnames,value=TRUE); #state variables to estimate (which we
    will do first) - grabs all parameters with a '0' in them
estpars <- setdiff(estnames,estICs); # All other parameters we want to estimate

# Random-walk standard deviation. On the log and logit scale a value of 0.02 is
    about a 2% variation in the parameter value.
# You dont need to use the same rw.sds for each parameter. I have found that 2%
    works well when you have used a Sobol design.
#these are .2 for multiple iterations through mif (below)
# rw.sds.ics <- rep(0.2,times=length(estICs)); #ran walk SDs of the ICs for state
     variables
# names(rw.sds.ics) <- estICs;
rw.sds <- rep(0.1,times=length(estnames)); #ran walk SDs of the ICs for all
    parameters to estimate variables
names(rw.sds) <- estnames;
```

```
# set up run index, this is indexing the parameter set that we are using in this
    run
no.sets<- dim(param.sets)[1]
n<- floor(no.sets/no.runs)
start <- 1 + (run-1)*n # n is the number of param set in each run
end<- n + (run-1)*n
if(run==no.runs){end=dim(param.sets)[1]}

################
#MIF
write.seq<-rep(c(rep(0,9),1),length.out=dim(param.sets)[1]); #how often the file
    writes

for (i in start:end){ # for each parameter set #this basically keeps them in
    chunks of which of the 50k+ param sets to run (i.e. 1-5000, 5001-1000, etc...)
# for (i in 1:2){ # for each parameter set
  current.params <- unlist(param.sets[i,1:length(param.sets[1,])]) # grab
      parameter set
  for(j in 1:no.MC.reps){ # for each replicate #how many times you want to repeat
       each parameter set
  # for(j in 1:1){ # for each replicate #how many times you want to repeat each
      parameter set
    results.index<- ((i-1)*no.MC.reps) + j # calculate the index in our results
        data frame where we will put the results of this replicate
    try({
    save.rng.state <- .Random.seed
    set.seed(seeds[results.index]) # set seed

    # Find some decent initial conditions for this parameter set
```

```
# Also note that each successive loop reduces the random walk standard
    deviation.
mif.dat1 <- mif2(
               my.model,
               Nmif=100,
               start=current.params, # we will initialize using the parameter
                   set resulting from "mif.ics" above.
               pars=estnames, # list of parameters we are estimating
               # ivps=estICs, # list of initial conditions we are estimating
               rw.sd=rw.sds,
               Np=10000,
               # var.factor=3,
               cooling.type='hyperbolic',
               cooling.fraction= 0.5,
               # method='mif2',
               max.fail=110
              )
MIFed.pars1<- coef(mif.dat1) # MIFed parameter set
# MIFed.pars4<- coef(mif.dat1) # MIFed parameter set

mif.dat2 <- mif2(
               my.model,
               Nmif=100,
               start=MIFed.pars1, # we will initialize using the parameter set
                    resulting from "mif.ics" above.
               pars=estnames, # list of parameters we are estimating
               # ivps=estICs, # list of initial conditions we are estimating
               rw.sd=(0.5*rw.sds),
               Np=10000,
```

```
                # var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.5,
                # method='mif2',
                max.fail=110
               )
MIFed.pars2<- coef(mif.dat2) # MIFed parameter set

mif.dat3 <- mif2(
                my.model,
                Nmif=100,
                start=MIFed.pars2, # we will initialize using the parameter set
                    resulting from "mif.ics" above.
                pars=estnames, # list of parameters we are estimating
                # ivps=estICs, # list of initial conditions we are estimating
                rw.sd=(0.25*rw.sds),
                Np=10000,
                # var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.5,
                # method='mif2',
                max.fail=110
               )
MIFed.pars3<- coef(mif.dat3) # MIFed parameter set

mif.dat4 <- mif2(
                my.model,
                Nmif=100,
```

```
                start=MIFed.pars3, # we will initialize using the parameter set
                    resulting from "mif.ics" above.
                pars=estnames, # list of parameters we are estimating
                # ivps=estICs, # list of initial conditions we are estimating
                rw.sd=(0.25*rw.sds),
                Np=10000,
                # var.factor=3,
                cooling.type='hyperbolic',
                cooling.fraction= 0.5,
                # method='mif2',
                max.fail=110
               )
MIFed.pars4<- coef(mif.dat4) # MIFed parameter set
.Random.seed <<- save.rng.state

# calculate the likelihood of the MIFed parameter set
loglik.mif<- replicate(n=10,logLik(pfilter(my.model,params=MIFed.pars4,Np
    =5000,max.fail=110)))
  bl<- mean(loglik.mif)
  loglik.mif.est<- bl+log(mean(exp(loglik.mif-bl)))
  loglik.mif.se <- sd(exp(loglik.mif-bl))/sqrt(length(loglik.mif))/exp(loglik.
      mif.est-bl)

# put the results in our results data frame
MIFed.params[results.index,1:length(param.sets[1,])]<-MIFed.pars4
MIFed.params[results.index,'LogLik'] <-loglik.mif.est
MIFed.params[results.index,'seLogLik'] <-loglik.mif.se
MIFed.params <- as.data.frame(MIFed.params)
})
```

```
  }
  # write to file every 10 parameter sets
  if(write.seq[i]==1){write.csv(MIFed.params,file=paste(c(output.name,'_run',run,
      '.csv'),collapse=''),row.names=FALSE)}
}


################
# Write the final output
###############
write.csv(MIFed.params,file=paste(c(output.name,'_run',run,'.csv'),collapse=''),
    row.names=FALSE)
```

D.1.5.3 Pomp Object

```
require(pomp)

# On Pascual Lab flux:
# require('subplex',lib.loc='/home2/bakkerke/pomp_develop');
# require('nloptr',lib.loc='/home2/bakkerke/pomp_develop');
# require('mvtnorm',lib.loc='/home2/bakkerke/pomp_develop');
# require('deSolve',lib.loc='/home2/bakkerke/pomp_develop');
# require('coda',lib.loc='/home2/bakkerke/pomp_develop');
# require('pomp',lib.loc='/home2/bakkerke/pomp_develop');

################
# Load Covars & Data
################
covariates2 <- read.csv('Thai_Covariates_short.csv', header=TRUE);
```

```
covariates1 <- covariates2[,-c(1)]
covariates <- covariates1[1:97,]
covariates$modelID <- param.sets$model[1];
covariates$boosting <- param.sets$boost[1];

covariates$seas <- periodic.bspline.basis(covariates$timevec,nbasis=6,degree=2,
    period=1);
covariates$seasB <- periodic.bspline.basis(covariates$timevec,nbasis=6,degree=2,
    period=1);

# covariates$modelID <- param.sets$model[1:109];
# covariates$boosting <- param.sets$boost[1:109];



################
# Functions we need to transform params
################
expit=function(x){1.0/(1.0+exp(-x))};
logit=function(p){log(p/(1-p))};

################
# Set up data frame with the data to be fit, this dataframe should start at
    timestep==1
################
CASES1 <- read.csv('Thai_CPOX_SHING_short.csv', header=TRUE); #DO NOT PAD DATA
CASES <- CASES1[1:96,]
################
# Compile/load model from C-code
################
```

```
if (is.loaded("zoster_meas_dens")) dyn.unload("VZV_1.so")
system("R CMD SHLIB VZV_1.c")
dyn.load("VZV_1.so")

################
# POMP Object
################
my.model<- pomp(
      data=CASES[c("step","POX","SHING")], # data are cases and associated
          timestep, starting w/ timestep==1
      times="step", # name of time variable in data dataframe
      rprocess=euler.sim("zoster_proc_sim",delta.t=1/4,PACKAGE="VZV_1"), # polio_
          proc_sim is the name of our process model in C-code; measles_biweekly is
           the name of our C file; see ?plugins for integration options
      rmeasure="zoster_meas_sim", # name of measurement model in C-code
      dmeasure="zoster_meas_dens", # name of measurement density model in C-code
      PACKAGE="VZV_1", # name of C-file
      t0=0, # time 0, this is the time we initialize the model
      obsnames=c("POX","SHING"), # name of observations in data dataframe
      statenames=c("pS","pE","pI","pL1","pL2","FOI","S","E","IVZ","L1","IHZ","L2",
          'IVZNEW','IHZNEW','rep_hz'), # process model state variable names, this
          vector must match the order of the states in (x[stateindex[i]]) in C-
          code
      covarnames=c("population","births","coc1","coc2","UV","modelID","boosting","
          seas.1","seasB.1"), # covariate names, must match order of (covar[
          covindex[i]]) in C-code
      paramnames=c('delta','Neta','zeta1','Omega','Alpha','Beta_SD','Phi','Gamma',
          'Iota','rhovz','tauvz','tauhz','m1','m2','zeta1B','eta','omicron','chi',
          'upsilon','sigma','a','nu','xi'), # param names, must match order of (p[
```

```
    parindex[0]]) in C-code, but notice the actual name doesnt have to match
     for instance here we use "alpha" in C-code we call it LOGALPHA, only
    the indexing matters not the exact name
tcovar="step", # name of the time variable in the covariate data frame
covar=covariates, # name of the covariate dataframe
 initializer=function(params,t0, ...){ # the initializer allows you to
    specify how to set up initial conditions using your parameter set "
    params"
       p <- expit(params)
      covars<- as.data.frame(covariates)
       with(
         as.list(p),
       {
         x0=c(pS.0,pE.0,pI.0,pL1.0,pL2.0,FOI.0,S.0,E.0,IVZ.0,L1.0,IHZ.0,L2
             .0,IVZNEW.0,IHZNEW.0,rep_hz.0) #this is the initial states
           x0[1:6]= 0;
         IC_sum=sum(x0[7:12]);
        x0[7]= round((x0[7]/IC_sum)*covars[1,'population']); #S
           x0[8]= round((x0[8]/IC_sum)*covars[1,'population']); #E
           x0[9]= round((x0[9]/IC_sum)*covars[1,'population']); #IVZ
           x0[10]= round((x0[10]/IC_sum)*covars[1,'population']); #L1
         x0[11]= round((x0[11]/IC_sum)*covars[1,'population']); #IHZ
         x0[12]= round((x0[12]/IC_sum)*covars[1,'population']); #L2
           x0[13:14]= 0; #IVZNEW, IHZNEW
         # x0[15]= x0[15]; #it stays as it is
         names(x0)=c("pS","pE","pI","pL1","pL2","FOI","S","E","IVZ","L1","
             IHZ","L2","IVZNEW","IHZNEW","rep_hz");
         x0
      }
```

```
        )
    }
)
################
# end pomp object
################
```

D.1.5.4 C++ code

```
// dear emacs, please treat this as -*- C++ -*-

#include <R.h>
#include <Rmath.h>
#include <math.h>

// parameters
// process model parameters w/ log or logit transformation
#define DELTA             (p[parindex[0]]) // natural death rate,
    fixed parameter
#define NETA              (p[parindex[1]]) // number of etas for
   the spline
#define LOGZETA           (&p[parindex[2]]) // seasonal eta
   vector for VZimmunomodulation
// #define LOGBETA       (p[parindex[3]]) // beta value,
   chickenpox transmission rate
#define LOGITOMEGA        (p[parindex[3]]) // the relative
   infectiousness of herpes zoster versus varicella zoster, logit
    between 0-1
```

```
#define LOGALPHA          (p[parindex[4]]) // exponent in the
   force of infection
#define LOGBETA_SD        (p[parindex[5]]) // standard deviation
   of gamma distributed process noise
#define PHI               (p[parindex[6]]) // the rate at which
   an exposed person becomes infected with varicella zoster and
   manifests as chickenpox, fixed parameter
#define GAMMA             (p[parindex[7]]) // the recovery rate
   for varicella zoster, fixed parameter
#define IOTA              (p[parindex[8]]) // the average rate at
    which latent infections reactivate to manifest as herpes
   zoster, fixed parameter
// measurement model parameters, independent reporting for
   varicella zoster and herpes zoster
#define LOGITRHOVZ        (p[parindex[9]]) // the mean report
   rate for varicella zoster
#define LOGTAUVZ          (p[parindex[10]]) // observation model
   dispersion parameter for varicella zoster
#define LOGTAUHZ          (p[parindex[11]]) // observation model
   dispersion parameter for herpes zoster
#define LOGM1             (p[parindex[12]]) // observation model
   dispersion parameter for varicella zoster
#define LOGM2             (p[parindex[13]]) // observation model
   dispersion parameter for herpes zoster
#define LOGZETAB          (&p[parindex[14]]) // seasonal eta
   vector for HZ immunomodulation
#define LOGETA            (p[parindex[15]]) // boosting parameter
#define LOGOMICRON        (p[parindex[16]]) // boosting parameter
```

```
#define LOGCHI            (p[parindex[17]]) // scales UV to VZ
   sigmoid
#define LOGUPSILON        (p[parindex[18]]) // scales UV to VZ
   sigmoid
#define LOGSIGMA          (p[parindex[19]]) // scales UV to VZ
   sigmoid
#define LOGA              (p[parindex[20]]) // scales UV to HZ
   sigmoid
#define LOGNU             (p[parindex[21]]) // scales UV to HZ
   sigmoid
#define LOGXI             (p[parindex[22]]) // scales UV to HZ
   sigmoid


// #define LOGSCALARVZ     (p[parindex[17]]) // scales UV to VZ
// #define LOGSCALARHZ     (p[parindex[18]]) // scales UV to HZ

// // initial condition parameters, logit transforms bc they are
   fractions of the population
// #define LOGITfracS0   (p[parindex[15]]) // logit(initial
   fraction susceptible)
// #define LOGITfracE0   (p[parindex[16]]) // logit(initial
   fraction exposed), fix at zero
// #define LOGITfracIV0  (p[parindex[17]]) // logit(initial
   fraction varicella zoster)
// #define LOGITfracL10  (p[parindex[18]]) // logit(initial
   fraction latent1)
// #define LOGITfracIH0  (p[parindex[19]]) // logit(initial
   fraction herpes zoster)
```

```
// #define LOGITfracL20  (p[parindex[20]]) // logit(initial
   fraction latent2)

// state variables
// probabilities realting to state transitions
#define pS                  (x[stateindex[0]]) // prob(remaining
   in Susceptible class), this is time-varying
#define pE                  (x[stateindex[1]]) // prob(remaining
   in Exposed class)
#define pI                  (x[stateindex[2]]) // prob(remaining
   in Infected varicella zoster or herpes zoster class)
#define pL1                 (x[stateindex[3]]) // prob(remaining
   in the first latent class)
#define pL2                 (x[stateindex[4]]) // prob(remaining
   in the second latent class)
// other state variables
#define FOI                 (x[stateindex[5]]) // force of
   infection
#define S                   (x[stateindex[6]]) // susceptible
#define E                   (x[stateindex[7]]) // exposed
#define IVZ                 (x[stateindex[8]]) // infected,
   varicella zoster
#define L1                  (x[stateindex[9]]) // latent after
   varicella zoster infecton
#define IHZ                 (x[stateindex[10]]) // infected/
   reactive herpes zoster
#define L2                  (x[stateindex[11]]) // latent after
   herpes zoster infection
```

```
#define IVZ_NEW            (x[stateindex[12]]) // new infections
   with varicella zoster
#define IHZ_NEW            (x[stateindex[13]]) // new infections
   with herpes zoster
#define report_rate_hz     (x[stateindex[14]]) // time varying
   reporting for herpes zoster

// covariates
#define N           (covar[covindex[0]]) // population size
#define BABIES      (covar[covindex[1]]) // number of births
#define C1          (covar[covindex[2]]) // covariates for the
   trend in herpes zoster reporing
#define C2          (covar[covindex[3]]) // covariates for the
   trend in herpes zoster reporing
#define UV          (covar[covindex[4]]) // UV covariate
#define MODELNUMBER (covar[covindex[5]]) // model number
#define BOOSTING    (covar[covindex[6]]) // boosting on/off
#define SEAS        (&covar[covindex[7]]) // basis vector for
   seasonal VZ immunomodulation
#define SEASB       (&covar[covindex[8]]) // basis vector for
   seasonal HZ immunomodulation

// observations
#define VARICELLA_ZOSTER    (y[obsindex[0]]) // observed cases of
    varicella zoster (i.e., chicken pox)
#define HERPES_ZOSTER       (y[obsindex[1]]) // observed cases of
    herpes zoster (i.e., shingles)

// logit function for transformations
```

```
// static double logit (double p) {
//   return log(p/(1-p));
// }

// expit function for transformations
static double expit (double x) {
  return 1.0/(1.0+exp(-x));
}

// measurement density
void zoster_meas_dens (double *lik, double *y, double *x, double
   *p, int give_log,
                       int *obsindex, int *stateindex, int *
                          parindex, int *covindex,
                       int ncovar, double *covar, double t) {

  //define doubles
  double report_rate_vz;
  double tau_vz;
  double tau_hz;
  double lik1;
  double lik2;
  double templik;
  double tol = 1.0e-18;

  // untransform reporting parameters
  report_rate_vz = expit(LOGITRHOVZ);
  tau_vz = exp(LOGTAUVZ);
  tau_hz = exp(LOGTAUHZ);
```

```
// Varicella zoster likelihood
if(VARICELLA_ZOSTER > 0.0){
  lik1 = pnorm(VARICELLA_ZOSTER+0.5,report_rate_vz*IVZ_NEW,tau_
     vz*IVZ_NEW,1,0) - pnorm(VARICELLA_ZOSTER-0.5,report_rate_
     vz*IVZ_NEW,tau_vz*IVZ_NEW,1,0);
} else{
  lik1 = pnorm(VARICELLA_ZOSTER+0.5,report_rate_vz*IVZ_NEW,tau_
     vz*IVZ_NEW,1,0);
}

// Herpes zoster likelihood
if(HERPES_ZOSTER > 0.0){
  lik2 = pnorm(HERPES_ZOSTER+0.5,report_rate_hz*IHZ_NEW,tau_hz*
     IHZ_NEW,1,0) - pnorm(HERPES_ZOSTER-0.5,report_rate_hz*IHZ_
     NEW,tau_hz*IHZ_NEW,1,0);
} else{
  lik2 = pnorm(HERPES_ZOSTER+0.5,report_rate_hz*IHZ_NEW,tau_hz*
     IHZ_NEW,1,0);
}

 // work on log-scale bc computer can hard time with super
    small numbers
 // calculate the likelihood of observing the cases in both
    data sets, call it templik
  templik = exp(log(lik1) + log(lik2));
  *lik = templik + tol;

// take the log of the likelihood
```

```
  if (give_log) *lik = log(*lik);
  if (!isfinite(*lik)) Rprintf("zoster_meas_dens %lg %lg %lg\n",
     VARICELLA_ZOSTER,HERPES_ZOSTER,*lik);
}

// measurement model
void zoster_meas_sim (double *y, double *x, double *p,
                      int *obsindex, int *stateindex, int *
                         parindex, int *covindex,
                      int ncovar, double *covar, double t) {

  //define doubles
  double report_rate_vz;
  double tau_vz;
  double tau_hz;

  // untransform reporting parameters
  report_rate_vz = expit(LOGITRHOVZ);
  tau_vz = exp(LOGTAUVZ);
  tau_hz = exp(LOGTAUHZ);

  VARICELLA_ZOSTER = rnorm(report_rate_vz*IVZ_NEW,tau_vz*IVZ_NEW)
     ;
  if (VARICELLA_ZOSTER >= 0.0) {
    VARICELLA_ZOSTER = nearbyint(VARICELLA_ZOSTER);
  } else {
    VARICELLA_ZOSTER = 0.0;
  }
```

```
  HERPES_ZOSTER = rnorm(report_rate_hz*IHZ_NEW,tau_hz*IHZ_NEW);
  if (HERPES_ZOSTER >= 0.0) {
    HERPES_ZOSTER = nearbyint(HERPES_ZOSTER);
  } else {
    HERPES_ZOSTER = 0.0;
  }
}

// process model
void zoster_proc_sim (double *x, const double *p,
                      const int *stateindex, const int *parindex,
                         const int *covindex,
                      int covdim, const double *covar,
                      double t, double dt)
{
 // define doubles
 double zeta;
 double omega;
 double alpha;
 double beta_sd;
 double epsilon;
 double scale;
 double m1;
 double m2;
 double zetaB;
 double eta;
 double omicron;
 double scalarvz; //still used to equal sigmoid
 double scalarhz; //still used to equal sigmoid
```

```
double BetaT;
double kappa;
double chi;
double upsilon;
double sigma;
double a;
double nu;
double xi;
double psi;

// double modelnumber;
// double boosting;
 // define integers
int neta = (int) NETA;
int j;
int i;
// int fail;

// construct the B-spline
for (j = 0, zeta = 0; j < neta; j++)  //j++ means go one plus in
    the loop from 0->5
   zeta += LOGZETA[j]*SEAS[j]; //logeta vector with 6
      coefficients for BSpline, seas -> 6 covariates, this is
      the BSpline basis
   zeta = exp(zeta);
// construct the second B-spline
for (i = 0, zetaB = 0; i < neta; i++)  //j++ means go one plus
   in the loop from 0->5
```

```
    zetaB += LOGZETAB[i]*SEASB[i]; //logeta vector with 6
       coefficients for BSpline, seas -> 6 covariates, this is
       the BSpline basis
    zetaB = exp(zetaB);

// untransform parameters for use in the model
 omega = expit(LOGITOMEGA);
 alpha = expit(LOGALPHA);
 beta_sd = exp(LOGBETA_SD);
 m1 = exp(LOGM1);
 m2 = exp(LOGM2);
 eta = exp(LOGETA);
 omicron = exp(LOGOMICRON);
 chi = exp(LOGCHI);
 upsilon = exp(LOGUPSILON);
 sigma = exp(LOGSIGMA);
 a = exp(LOGA);
 nu = exp(LOGNU);
 xi = exp(LOGXI);


//estimate seasonal susceptibility or reactivation from UV
scalarvz = 1/(chi+exp((upsilon*UV)-sigma));
scalarhz =  1/(a+exp((nu*UV)-xi));

// calculate the gamma distributed process noise epsilon
  if (beta_sd > 0.0) {
    // calculte the scale parameter, which here is also the
       variance
```

```
  // in C, rgamma(shape,scale), if shape==(1/scale), then mean
     ==shape*scale==1 & variance==shape*scale*scale=scale
   scale = pow(beta_sd,2);
   epsilon = rgamma(1/scale,scale);
} else {epsilon = 1.0;}

// run some checks, if non-finte values return and do not
   continue
if(
   isfinite(zeta)== FALSE ||
   isfinite(epsilon)== FALSE ||
   isfinite(omega)==FALSE ||
   isfinite(alpha)==FALSE ||
   isfinite(FOI)==FALSE ||
   isfinite(pS)==FALSE ||
   isfinite(pL1)==FALSE ||
   isfinite(S)==FALSE ||
   isfinite(IVZ)==FALSE ||
   isfinite(IHZ)==FALSE ||
   isfinite(zetaB)==FALSE ||
   isfinite(eta)==FALSE ||
   isfinite(omicron)==FALSE ||
   isfinite(scalarvz)==FALSE ||
   isfinite(scalarhz)==FALSE
   )
  {
    Rprintf("non finite value in zoster_proc_sim\n");
    return;
  }
```

```
  // process model equations
    if(MODELNUMBER == 1){
      BetaT = zeta; kappa = zeta;
    }else if(MODELNUMBER == 2){
      BetaT = zeta; kappa = zetaB;
    }else if(MODELNUMBER == 3){
      BetaT = zeta; kappa = scalarhz;
    }else // if(MODELNUMBER == 4)
    {
      BetaT = scalarvz; kappa = scalarhz;
    // }else {Rprintf("You messed up your Model numbering\n");
  }

// if then for antibody boosting on/off
   if(BOOSTING == 1){
     psi = 1/(1+exp(eta*IVZ-omicron));
   }else //(BOOSTING == 0)
   {
     psi = 1;
   // }else{Rprintf("You messed up your Boosting numbering\n");
   }

// KAPPA = kappa;
// PSI = psi;

// General form for FOI

  FOI = BetaT*pow(((IVZ+omega*IHZ)/N),alpha)*epsilon;
```

```
  if(FOI < 0.0){FOI = 0.0;}
  // probabilites based on the per capita rates and the time step
     , dt in months
  // we intend to use a dt of 1/4 month, i.e., weekly time steps
  pS = exp(-dt*(FOI+DELTA));
  pE = exp(-dt*(PHI+DELTA));
  pI = exp(-dt*(GAMMA+DELTA));
  pL1 = exp(-dt*(IOTA*kappa*psi+DELTA));
  pL2 = exp(-dt*(DELTA));
  // epi classes
  S = dt*BABIES + S*pS;
  E = (S*(1-pS)*FOI/(DELTA+FOI)) + (pE*E);
  IVZ = (E*(1-pE)*PHI/(DELTA+PHI)) + (pI*IVZ);
  L1 = (IVZ*(1-pI)*GAMMA/(GAMMA+DELTA)) + (pL1*L1);
  IHZ = (L1*(1-pL1)*(IOTA*kappa*psi)/(IOTA*kappa*psi+DELTA)) + (
     pI*IHZ);
  L2 = (IHZ*(1-pI)*GAMMA/(GAMMA+DELTA)) + (pL2*L2);
  // new infections
  IVZ_NEW = (E*(1-pE)*PHI/(DELTA+PHI));
  IHZ_NEW = (L1*(1-pL1)*(IOTA*kappa*psi)/(IOTA*kappa*psi+DELTA));
  // report rate for herpes zoster is time varying
  report_rate_hz +=  m1*t*C1 - m2*t*C2;
  if(report_rate_hz > 1.0){report_rate_hz = 1.0;}
  if(report_rate_hz < 0.0){report_rate_hz = 0.0;}
}
```

D.1.6 Summary

In the 5 years it took me to complete my Ph.D., I'd estimate than more than 95% of my (non-teaching) time and effort went towards mathematical modeling. Mathematical modeling consists of not just writing and running the code, but developing biologicaly based and sensible models. It is always possible to add more parameters to a model to achieve a better fit, but are they biologically based? This question I faced throughout my dissertation, along with how one picks out the 'best' model. Chapters 3 & 4 both include 8+ models, and both pick one that provides the 'best' fit. While Chapter 3 was mostly straightforward, I am still in the process of choosing the correct model for Chapter 4. One model fits better, but may not be as bioligcally sensible as the next best fit model, thus the quandry.

Pomp may seem inaccessible or incredibly complex, however I was able to pick it up with minimal coding or mathematical background (though it did take a few years). There are further tutorials available on the *pomp* website. Finally, all my code was based off the initial code I looked at in 2012, and *pomp* has evolved into a much cleaner software package since. Among other improvements, *pomp* creators have integrated C-snippets into *pomp*, where one no longer needs 3 different files (e.g. Fig D.1) to execute the code. Over the next few months, I look forward to 'updating' my code for all of my in-progress projects, many of which are summarized in Chapter 5.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] National center for atmospheric research, https://rda.ucar.edu/. Accessed Aug 4th, 2016.

[2] Sonia Altizer, Andrew Dobson, Parviez Hosseini, Peter Hudson, Mercedes Pascual, and Pejman Rohani. Seasonality and the Dynamics of Infectious Diseases. *Ecology letters*, 9(4):467–84, April 2006.

[3] Sonia Altizer, Wesley M Hochachka, and Andre A Dhondt. Seasonal Dynamics of Mycoplasmal Conjunctivitis in Eastern North American House Finches. *Journal of Animal Ecology*, 73:309–322, March 2004.

[4] L.J. Anderson. Role of parvovirus b19 in human disease. *The Pediatric infectious disease journal*, 6(8):711–718, 1987.

[5] C. Armstrong. Seasonal distribution of poliomyelitis. *American Journal of Public Health and the Nation's Health*, 40.10:1296–1304, 1950.

[6] J Aschoff. Annual cycle of reproduction in warm-blooded animals. *General Studies*, 8:742–776, 1955.

[7] J. Aschoff. Annual rhythms in man. *Biological rhythms*, pages 475–487, 1981.

[8] Australian-Government. National Notifiable DIseases Surveillance System. http://www9.health.gov.au/cda/source/rpt_1_sel_a.cfm. Accessed May 1, 2015, 2015.

[9] K.M. Bakker, M. Martinez-Bakker, B. Helm, and T.J. Stevenson. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proceedings of the National Academy of Sciences of the United States of America*, 113:6689–6694, 2016.

[10] E Batschelet. *Statistical rhythm evaluation*. John Wiley, 1974.

[11] C.T. Bauch and D.J. Earn. Transients and attractors in epidemics. *Proceedings of the Royal Society London B: Biological Sciences*, 270:1573–1578, 2003.

[12] Roger Baxter, Trung N Tran, Paula Ray, Edwin Lewis, Bruce Fireman, Steve Black, Henry R Shinefield, Paul M Coplan, and Patricia Saddier. Impact of vaccination on the epidemiology of varicella: 1995-2009. *Pediatrics*, 134(1):24–30, 2014.

[13] S. Becker. Seasonality of fertility in matlab, bangladesh. *Journal of biosocial science*, 13 (01):97–105, 1981.

[14] S. Becker, A. Chowdhury, and H. Leridon. Seasonal patterns of reproduction in matlab, bangladesh. *Population Studies*, 40 (3):457–472, 1986.

[15] Stan Becker. Seasonal Patterns of Births and Conception Throughout the World. *Advances in Experimental Medicine and Biology*, 286:59–72, 1991.

[16] Michael Begon, Sandra Telfer, Matthew J Smith, Sarah Burthe, Steve Paterson, and Xavier Lambin. Seasonal Host Dynamics Drive the Timing of Recurrent Epidemics in a Wildlife Population. *Proceedings of the Royal Society B: Biological Sciences*, 276(1662):1603–1610, May 2009.

[17] L. Bengtsson, X. Lu, A. Thorson, R. Garfield, and J. Von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in haiti. *PLoS medicine*, 8(8):e1001083, 2011.

[18] M. Benyesh-Melnick, J. L. Melnick, W. E. Rawls, I. Wimberly, J. B. Oro, E. Ben-Porath, and V. Rennick. Studies of the immunogenicity, communicability and genetic stability of oral poliovaccine administered during the winter. *American journal of epidemiology*, 86(1):112–136, 1967.

[19] N Bharti, a J Tatem, M J Ferrari, R F Grais, A Djibo, and B T Grenfell. Explaining Seasonal Fluctuations of Measles in Niger using Nighttime Lights Imagery. *Science*, 334(6061):1424–1427, December 2011.

[20] O N Bjornstad, B F Finkenstadt, and B T Grenfell. Dynamics of measleas epidemics: estimating escaling of transmission rates using a time series {SIR} model. *Ecological Monography*, 72:169–184, 2002.

[21] K. Bloom-Feshbach, W.J. Alonso, V. Charu, J. Tamerius, L. Simonsen, M.A. Miller, and C. Viboud. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (rsv): a global comparative review. *PloS one*, 8(2):e54445, 2013.

[22] M. Bobak and A. Gjonca. The seasonality of live birth is strongly influenced by socio-demographic factors. *Human Reproduction*, 16 (7):1512–1517, 2001.

[23] Luca Bolzoni, Andrew P Dobson, Marino Gatto, and Giulio A De Leo. Allometric Scaling and Seasonality in the Epidemics of Wildlife Diseases. *The American Naturalist*, 172(6):818–828, December 2008.

[24] M Brisson, WJ Edmunds, B Law Band N. J. Gay, R. Walld, M. Brownell, L. L. Roos, and G. De Serres. Epidemiology of varicella zoster virus infection in canada and the united kingdom. *Epidemiology Infections*, 127(2):305–314, 2001.

[25] M. Brisson, W.J. Edmunds, N.J. Gay, B. Law, and G. De Serres. Analysis of varicella vaccine breakthrough rates: implications for the effectiveness of immunization programmes. *Vaccine*, 18:2775–2778, 2000.

[26] M Brisson, NJ Gay, WJ Edmunds, and NJ Andrews. Exposure to varicella boosts immunity to herpes zoster: implications for mass vaccination against chicken pox. *Vaccine*, 20:2500–2507, 2002.

[27] DL Britton and MA Brock. Seasonal germination from wetland seed banks, 1994.

[28] F H Bronson. Mammalian reproduction: an ecological perspective. *Biology of reproduction*, 32:1–26, 1985.

[29] F H Bronson. Seasonal Variation in Human Reproduction: Environmental Factors. *The Quarterly Review of Biology*, 70(2):141–164, 1995.

[30] F H Bronson. Climate change and seasonal reproduction in mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364:3331–3340, 2009.

[31] F.H. Bronson. Are humans seseasonal photoperiodic? *Journal of Biological Rhythms*, 19 (3):180–192, 2004.

[32] H Broutin, C Viboud, B T Grenfell, M A Miller, and P Rohani. Impact of Vaccination and Birth Rate on the Epidemiology of Pertussis: a Comparative Study in 64 Countries. *Proceedings of the Royal Society B: Biological Sciences*, 277:3239–3245, November 2010.

[33] J.S. Brownstein, C.C. Freifeld, and L.C. Madoff. Digital disease detection—harnessing the web for public health surveillance. *New England Journal of Medicine*, 360(21):2153–2157, 2009.

[34] John Bryden, Sebastian Funk, and Vincent AA Jansen. Word usage mirrors community structure in the online social network Twitter. *EPJ Data Science*, 1:1–9, 2013.

[35] R. Cancho-Candela, J.M. Andrs-de Llano, and J. Ardura-Fernandez. Decline and loss of birth seasonality in spain: analysis of 33 421 731 births over 60 years. *Journal of Epidemiology & Community Health*, 61(8):713–718, 2007.

[36] John J Cannell, Michael Zasloff, Cedric F Garland, Robert Scragg, and Edward Giovannucci. On the Epidemiology of Influenza. *Virology Journal*, 5:1–12, January 2008.

[37] Kylie S. Carville, Michaela A. Riddell, and Heath A. Kelly. A decline in varicella but an uncertain impact on zoster following varicella vaccination in Victoria, Australia. *Vaccine*, 28(13):2532–2538, 2010.

[38] CDC. Evaluation of varicella reporting to the National Notifiable Disease Surveillance System–United States, 1972-1997. *Morbidity and mortality weekly report January 29, 1999*, 3, 1999.

[39] CDC. Prevention and control of influenza with vaccines: Recommendations of the Advisory Committee on Immunization Practices (ACIP) - United States 2012-13 season. *Morbidity and Mortality Weekly Report August 17, 2012*, 61(32):605–8, 2012.

[40] CDC. Chickenpox (Varicella) Vaccine Safety. http://www.cdc.gov/vaccinesafety/vaccines/varicella-vaccine.html. Accessed Sept 15, 2015, September 2015.

[41] CDC. Conducting Varicella Surveillance. http://www.cdc.gov/chickenpox/hcp/conducting-surveillance.html. Accessed Sept 15, 2015, September 2015.

[42] CDC. Infectious Diseases Related to Travel. http://wwwnc.cdc.gov/travel/yellowbook/2016/infectious-diseases-related-to-travel/varicella-chickenpox. Accessed Sept 9, 2015, 2015.

[43] CDC. National Center for Health Statistics Mortality Surveillance Data. http://www.cdc.gov/flu/weekly/nchs.htm Accessed March 4, 2016, 2016.

[44] CDC. Seasonal Influenza Vaccine Effectiveness, 2005-2015. http://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.htm Accessed March 4, 2016, 2016.

[45] K. S. F. Chang, S. T. Chan, W. D. Low, and C. K. Ng. Climate and conception rates in hong kong. *Human Biology*, pages 366–376, 1963.

[46] R.S. Chapman, K.R. Cross, and D.M. Felming. The incidence of shingles nad its implications for vaccination policy. *Vaccine*, 21:2541–2547, 2003.

[47] M. Choisy and P. Rohani. Changing spatial epidemiology of pertussis in continental USA, 2012.

[48] D.A. Clements, J I Zaref, C L Bland, E B Walter, and P M Coplan. Partial uptake of varicella vaccine and the epidemiological effect on varicella disease in 11 day-care centers in North Carolina. *Archives of Pediatrics & Adolescent Medicine*, 4:455–461, 2001.

[49] J.L. Coleman and D. Shukla. Recent advances in vaccine develop- ment for herpes simplex virus types i and ii. *Human Vaccines and Immunotherapeutics*, 9:729–735, 2013.

[50] R G Condon and R Scaglion. The Ecology of Human Birth Seasonality. *Human Ecology*, 10(4):495–511, December 1982.

[51] S.M. Cook, R.I. Glass, C.W. LeBaron, and M.S. Ho. Global seasonality of rotavirus infections. *Bulletin of the World Health Organization*, 68(2):171, 1990.

[52] U. M. Cowgill. Season of birth in man. contemporary situation with special reference to europe and the southern hemisphere. *Ecology*, pages 614–623, 1966.

[53] John A. Crump, Stephen P. Luby, and Eric D. Mintz. The global burden of typhoid fever. *Bulletin of the World Health Organization*, 82.5:346–353, 2004.

[54] D.A. Cummings, S. Iamsirithaworn, J.T. Lessler, A. McDermott, R. Prasanthong, A. Nisalak, R.G. Jarman, D.S. Burke, and R.V. Gibbons. The impact of the demographic transition on dengue in thailand: insights from a statistical analysis and mathematical modeling. *PLoS medicine*, 6(9):e1000139, 2009.

[55] A Damasceno, AS Moraes, A Farias, BP Damasceno, LM Barbosa dos Santos, and F Cendes. A spring to summer shift of pro-inflammatory cytokine production in multiple sclerosis patients. *Journal of the Neurological Sciences*, 360:37–40, 2015.

[56] A. Dawson, V. M. King, G. E. Bentley, and G. F. & Ball. Photoperiodic control of seasonality in birds. *Journal of Biological Rhythms*, 16 (4):365–380, 2001.

[57] E.C. De Fabo and F.P. Noonan. Mechanism of immune suppression by ultraviolet irradiation in vivo. *Journal of Experimental Medicine*, 157:84–98, 1983.

[58] R. Desai, A.J. Hall, B.A. Lopman, Y. Shimshoni, M. Rennick, N. Efron, Y. Matias, M.M. Patel, and U.D. Parashar. Norovirus disease surveillance using google internet query share data. *Clinical Infectious Diseases*, 55(8):e75–e78, 2012.

[59] R. Desai, B. A. Lopman, Y. Shimshoni, J. P. Harris, M. M. Patel, and U.D. Parashar. Use of internet search data to monitor impact of rotavirus vaccination in the united states. *Clinical Infectious Diseases*, cis121, 2012.

[60] DHS. Demographic Health Surveys, 2014.

[61] J.G Donahue, P.W. Choo, J.E. Manson, and R. Platt. The incidence of herpes zoster. *Archives of Internal Medicine*, 155:1605–1609, 1995.

[62] L. Donaldson and et al. Independent monitoring board of the global polio eradication initiative tenth report october 2014. *Independent Monitoring Board of the Global Polio Eradication Initiative*, 10:1–36, 2014.

[63] Xaquin Castro Dopico, Marina Evangelou, Ricardo C Ferreira, Hui Guo, Marcin L Pekalski, Deborah J Smyth, Nicholas Cooper, Oliver S Burren, Anthony J Fulford, Branwen J Hennig, et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nature communications*, 6, 2015.

[64] Audrey Dorelien, Sebastien Ballesteros, and Bryan Grenfell. Impact of birth seasonality on dynamics of acute immunizing infections in sub-saharan africa. *PLoS ONE*, 8(10), 2012.

[65] S F Dowell. Seasonal Variation in Host Susceptibility and Cycles of Certain Infectious Diseases. *Emerging Infectious Diseases*, 7(3):369–374, 2001.

[66] S.G. Drozdov. *Epidemiological efficiency of mass vaccination against poliomyelitis by live vaccine.* PhD thesis, Retrieved from the Chumakov Polio Institute Library. Moscow, Russia., 1964.

[67] Scott M Duke-Sylvester, Luca Bolzoni, and Leslie A Real. Strong Seasonality Produces Spatial Asynchrony in the Outbreak of Infectious Diseases. *Journal of the Royal Society Interface*, 8(59):817–825, June 2011.

[68] Ken T D Eames, Natasha L Tilston, Ellen Brooks-Pollock, and W John Edmunds. Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. *PLoS Computational Biology*, 8(3):1–8, March 2012.

[69] DJD Earn, Pejman Rohani, BM Bolker, and BT Grenfell. A simple model for complex dynamical transitions in epidemics. *Science*, 287(5453):667–70, January 2000.

[70] ECDC. Varicella vaccine in the European Union. *ECDC. Stockholm.*, 2014.

[71] J. R. Ehrenkranz. Seasonal breeding in humans: birth records of the labrador eskimo. *Fertility and sterility*, 40 (4):485–489, 1983.

[72] S.L. Elliott, A. Suhrbier, J.J. Miles, G. Lawrence, S.J. Pye, T.T. Le, A. Rosenstengel, T. Nguyen, A. Allworth, S.R. Burrows, and J. Cox. Phase i trial of a cd8+ t-cell peptide epitope-based vaccine for infectious mononucleosis. *Journal of Virology*, 82:1448–1457, 2008.

[73] J.R. Enright. The epidemiology of paralytic poliomyelitis in hawaii. *Hawaii Medical Journal*, 13(5):350–354, 1954.

[74] Estonian-Government. Terviseamet Health Board. http://www.terviseamet.ee/. Accessed Sept 1, 2015, Sept 2015.

[75] M. J. Ferrari, B. T. Grenfell, and P. M. Strebel. Think globally, act locally: the role of local demographics and vaccination coverage in the dynamic response of measles infection to control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368 (1623), 2013.

[76] Matthew J Ferrari, Rebecca F Grais, Nita Bharti, Andrew J K Conlan, Ottar N Bjørnstad, Lara J Wolfson, Philippe J Guerin, Ali Djibo, and Bryan T Grenfell. Supplementary Notes: The Dynamics of Measles in Sub-Saharan Africa. *Nature*, 451(7179):679–84, February 2008.

[77] P E Fine and J A Clarkson. Seasonal Influences on Pertussis. *International Journal of Epidemiology*, 15(2):237–247, June 1986.

[78] B. F. Finkenstädt and B. T. Grenfell. Time Series Modelling of Childhood Diseases: a Dynamical Systems Approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(2):187–205, 2000.

[79] D.M. Fleming, C.A. Norbury, and D.L. Crombie. *Annual and Seasonal Variation in the Incidence of Common Diseases*. Birmingham: Royal College of General Practitioners, 1991.

[80] Jessica Forrest and Abraham J Miller-Rushing. Toward a Synthetic Understanding of the Role of Phenology in Ecology and Evolution. *Philosophical Transactions of the Royal Society: Biological Sciences*, 365(1555):3101–3112, October 2010.

[81] M Gallerani and R Manfredini. Seasonal variation in herpes zoster infection. *British Journal of Dermatology*, 142(3):588–589, 2000.

[82] GP Garnett and BT Grenfell. The epidemiology of varicella-zoster virus infections: the iinfluence of varicella on the prevalence of zoster. *Epidemiological Infection*, 108:513–528, 1992.

[83] H S Gear. The Virus of Poliomyelitis : Its Distribution and Methods of Spread. *The Journal of the Royal Society for the Promotion of Health*, 69(3):149–153, January 1949.

[84] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

[85] G. S. Goldman and P.G. King. Review of the united states universal varicella vaccination program: herpes zoster incidence rates, cost-effectiveness, and vaccine efficacy based primarily on the antelope valley varicella active surveillance project data. *Vaccine*, 31:1680–1694, 2013.

[86] Google. Google Trends. https://www.google.com/trends/., 2015.

[87] GPEI. Global polio eradication initiative, 2015.

[88] NC Grassly and Christophe Fraser. Seasonal Infectious Disease Epidemiology. *Proceedings of the Royal Society B: Biological Sciences*, 273(1600):2541–50, October 2006.

[89] B T Grenfell, O N Bjø rnstad, and J Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414:716–723, 2001.

[90] G. Guzzetta, P. Poletti, E. Del Fava, M. Ajelli, G.P. Scalia Tomba, S. Merler, and P. Manfredi. Hope- simpsons progressive immunity hypothesis as a possible explanation for herpes zoster incidence data. *American Journal of Epidemiology*, 177:1134–1142, 2013.

[91] T.J. Hastie and R.J. Tibshirani. *Generalized additive models*, volume 43. CRC Press, 1990.

[92] Daihai He and David J D Earn. Epidemiological Effects of Seasonal Oscillations in Birth Rates. *Theoretical Population Biology*, 72(2):274–291, September 2007.

[93] Health-Department. Surveillance of viral pathogens in Australia: varicella. *Communicable Diseases Intelligence*, 26, 2002.

[94] J. H. Hemmes, K. C. Winkler, and S. M. Kool. Virus Survival as a Seasonal Factor in Influenza and Poliomyelitis. *Antonie van Leeuwenhoek*, 28(1):221–233, December 1962.

[95] D. L. Heymann, K. Murphy, M. Brigaud, M. Aymard, A. Tembon, and G. K. Maben. Oral poliovirus vaccine in tropical africa: greater impact on incidence of paralytic disease than expected from coverage surveys and seroconversion rates. *Bulletin of the World Health Organization*, 65(4):495, 1987.

[96] O. Higgins, J. Sixsmith, M.M. Barry, and C. Domegan. A literature review on health information seeking behaviour on the web: a health consumer and health professional perspective. *ECDC Technical Report, Stockholm*, 2011.

[97] RE Hope-Simpson. The nature of herpes zoster: A long term study and a new hypothesis. *Proceedings of the Royal Society of Medicine*, 58:9–20, 1965.

[98] R.E. Hope-Simpson. The role of season in the epidemiology of influenza. *Epidemiology & Infection*, 86(1):35–47, 1981.

[99] R.E. Hope-Simpson. Epidemiology of shingles. *Journal of the Royal Society of Medicine*, 84:184, 1991.

[100] Parviez R Hosseini, André A Dhondt, and Andy Dobson. Seasonality and Wildlife Disease: How Seasonal Birth, Aggregation and Variation in Immunity Affect the Dynamics of *Mycoplasma gallisepticum* in House Finches. *Proceedings of the Royal Society B: Biological Sciences*, 271(1557):2569–2577, December 2004.

[101] H. Hull, N. A. Ward, B. P. Hull, J. B. Milstien, and C de Quadros. Paralytic poliomyelitis: seasoned strategies, disappearing disease. *The Lancet*, 343.8909:1331–1337, 1994.

[102] T.G. Hull. A graphical study of the epidemiology of poliomyelitis. *American Journal of Public Health*, 7.10:813–823, 1917.

[103] A. Hulth, G. Rydevik, A. Linde, and J. Montgomery. Web queries as a source for syndromic surveillance. *PloS one*, 4(2):e4378, 2009.

[104] C Jackson, P Mangtani, P Fine, and E Vynnycky. The effects of school holidays on transmission of varicella zoster virus, england and wales, 1967-2008. *PLoS One*, 9(6):e99762, 2014.

[105] Craig James and Richard Shine. The seasonal timing of reproduction: - A tropical-temperate comparison in Australian lizards. *Oecologia*, 67:464–474, 1985.

[106] A.O. Jumaan, O. Yu, L.A. Jackson, K. Bohlke, K. Galil, and J.F. Seward. Incidence of herpes zoster, before and after varicella-vaccination-associated decreases in the incidence of varicella, 1992-2002. *Journal of Infectious Diseases*, 191(12):2002–2007, 2005.

[107] M. Karhunen, T. Leino, H. Salo, I. Davidkin, T. Kilpi, and K. Auranen. Modelling the impact of varicella vaccination on varicella and zoster. *Epidemiological Infection*, 138:469–481, 2010.

[108] M.J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

[109] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, 115 (772):700–721, 1927.

[110] Nino Khetsuriani, Ashley LaMonte-Fowlkes, S. Oberst, and Mark A. Pallansch. Enterovirus surveillance United States, 1970D2005. *Morbidity and Mortality Weekly Report surveillance summary*, 8:1–20, 2006.

[111] Aaron A. King, Edward L. Ionides, Carles Martinez Bretó, Stephen P. Ellner, Matthew J. Ferrari, Bruce E. Kendall, Michael Lavine, Dao Nguyen, Daniel C. Reuman, Helen Wearing, and Simon N. Wood. *pomp: Statistical Inference for Partially Observed Markov Processes,*. http://kingaa.github.io/pomp, 2015. R package, version 1.2.1.1.

[112] Aaron A. King, Dao Nguyen, and Edward L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, arXiv preprint arXiv:1509.00503, 2015.

[113] Robert Koch-Institut. Empfehlungen der Stïndigen Impfkommission (STIKO) am Robert Koch-InstitutStand. *Epidemiol Bulletin*, 30:235–254, 2006.

[114] K. Koelle, X. Rod, M. Pascual, M. Yunus, and G. Mostafa. Refractory periods and climate forcing in cholera dynamics. *Nature*, 436(7051):696–700, 2005.

[115] Michael E. Konkel and Kit Tilly. Temperature-regulated expression of bacterial virulence genes. *Microbes and Infection*, 2.2:157–166, 2000.

[116] P. M. Lago, J. R. Bravo, J. K. Andrus, M. M. Comellas, M. A. Galindo, C. A. de Quadros, and E. Bell. Lessons from cuba: mass campaign administration of trivalent oral poliovirus vaccine and seroprevalence of poliovirus neutralizing antibodies. *Bulletin of the World Health Organization*, 72(2):221, 1994.

[117] D A Lam and J A Miron. Global Patterns of Seasonal Variation in Human Fertility. *Annals of the New York Academy of Sciences*, 709:9–28, February 1994.

[118] David Lam and Jeffrey A Miron. The Seasonality of Births in Human Populations. *University of Michigan Population Studies Center Research Report Series*, 87-114:1–61, 1987.

[119] David A Lam and Jeffrey A Miron. Seasonality of Births in Human Populations. *Social Biology*, 38(1-2):51–78, 1991.

[120] Betta Laurino, A. Pugliese, G. Guzzetta, A. Landi, and P. Manfredi. Perspectives on optimal control of varicella and herpes zoster by mass routine varicella vaccination. *Proceedings of the Royal Society B: Biological Sciences*, 283:20160054, 2016.

[121] C.H. Lavinder, A.W. Freeman, and W.H. Frost. Epidemiologic studies of poliomyelitis in new york city and the northeastern united states during the year 1916. *US Government Printing Office*, 91, 1918.

[122] D. Lazer, R. Kennedy, G. King, and A. Vespignani. The parable of google flu: traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.

[123] J. Leung, R. Harpaz, N-A Molinari, A. Jumaan, and F. Zhou. Herpes zoster incidence among insured persons in the united states, 1993 2006: evaluation of impact of varicella vaccination. *Clinical Infectious Diseases*, 52:3332–340, 2011.

[124] R J Levine, R M Mathew, C B Chenault, M H Brown, M E Hurtt, K S Bentley, K L Mohr, and P K Working. Differences in the Quality of Seman in Outdoor Workers During Summer and Winter. *The New England Journal of Medicine*, 323(1):12–16, 1990.

[125] Paul Licht. Regulation of the annual testis cycle by photoperiod and temperature in the lizard anolis carolinensis. *Ecology*, pages 240–252, 1971.

[126] K J Linthicum, A Anyamba, C J Tucker, P W Kelley, M F Myers, and C J Peters. Climate and Satellite Indicators to Forecast Rift Valley Fever Epidemics in Kenya. *Science*, 285(5426):397–400, July 1999.

[127] W P London and J A Yorke. Recurrent Outbreaks of Measles, Chickenpox and Mumps. I. Seasonal Variation in Contact Rates. *American Journal of Epidemiology*, 98(6):453–468, December 1973.

[128] Anice C Lowen, Samira Mubareka, John Steel, and Peter Palese. Influenza Virus Transmission is Dependent on Relative Humidity and Temperature. *PLoS Pathogens*, 3(10):1470–1476, October 2007.

[129] Musawenkosi L H Mabaso, Marlies Craig, Amanda Ross, and Thomas Smith. Environmental Predictors of the Seasonality of Malaria Transmission in Africa: the Challenge. *The American Journal of Tropical Medicine and Hygiene*, 76(1):33–38, January 2007.

[130] V Macfarlane. Seasonality of Conception in Human Populations. *Biometeorology*, 4:167–182, 1970.

[131] M. Marin, H.C. Meissner, and J.F. Seward. Varicella prevention in the united states: a review of successes and challenges. *Pediatrics*, 122:e744–e751, 2008.

[132] M Martinez-Bakker, Kevin Bakker, Aaron A King, and Pejman Rohani. Human Birth Seasonality: Latitudinal Gradient and Interplay with Childhood Disease Dynamics. *Proceedings of the Royal Society B: Biological Sciences*, 281.1783, 2014.

[133] M. Martinez-Bakker, P. Rohani, and A. A. King. Unraveling the transmission dynamics of polio. *PLoS Biology*, 13:e1002172, 2015.

[134] Micaela Martinez-Bakker and Barbara Helm. The influence of biological rhythms on host–parasite interactions. *Trends in Ecology & Evolution*, 2015.

[135] V. Marziano, P. Poletti, G. Guzzetta, M. Ajelli, P. Manfredi, and S. Merler. The impact of demographic changes on the epidemiology of herpes zoster: Spain as a case study. *Proceedings of the Royal Society B: Biological Sciences*, 282:20142509, 2015.

[136] C Jessica E Metcalf, Ottar N Bjornstad, Bryan T Grenfell, and Viggo Andreasen. Seasonality and Comparative Dynamics of Six Childhood Infections in Pre-vaccination Copenhagen. *Proceedings of the Royal Society B: Biological Sciences*, 276(1676):4111–4118, December 2009.

[137] M. Momiyama and K. Katayama. Deseasonalization of mortality in the world. *International journal of biometeorology*, 16 (4):329–342, 1972.

[138] W S Moos and W Randall. Patterns of Human Reproduction and Geographic Latitude. *International Journal of Biometeorology*, 38(2):84–88, January 1995.

[139] David Morley. Severe Measles in the Tropics. -I. *British Medical Journal*, 1(February 1st):297–300, 1969.

[140] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):381–391, March 2008.

[141] N Nathanson and O.M. Kew. From emergence to eradication: The epidemiology of poliomyelitis deconstructed. *American Journal of Epidemiology*, 2010.

[142] A. Newsholme. *The Origin and Spread of Pandemic Diphtheria*. Swan Sonnenschein & Co., 1900.

[143] NHS. http://www.england.nhs.uk/, 2014.

[144] O.A. Odumade, K.A. Hogquist, and H.H. Balfour. Progress and problems in understanding and managing primary epstein-barr virus infections. *Clinical Microbiology Reviews*, 24:193–209, 2011.

[145] B. Ogunjimi, E. Smits, N. Hens, A. Hens, K. Lenders, M. Leven, V. Van Tendeloo, P. Van Damme, and P. Beutels. Exploring the impact of exposure to primary varicella in children on varicella-zoster virus immunity of parents. *Viral Immunology*, 24:151–157, 2011.

[146] B. Ogunjimi, E. Smits, S. Heynderickx, J. Van den Bergh, J. Bilcke, H. Jansens, R. Malfait, J. Ramet, H.T. Maecker, N. Cools, and P. Beutels. Influence of frequent infectious exposures on general and varicella-zoster virus-specific immune responses in pediatricians. *Clincial and Vaccine Immunology*, 21 (3):417–426, 2014.

[147] J. Paireau, A. Chen, H. Broutin, B. Grenfell, and N.E. Basta. Seasonal dynamics of bacterial meningitis: a time-series analysis. *The Lancet Global Health*, 4(6):e370–e377, 2016.

[148] S.R. Palmer, D.E. Donald, E.O. Caul, W. Kwantes, and H. Tillett. An outbreak of shingles? *The Lancet*, 326(8464):1108–1111, 1985.

[149] A J Peel, J R C Pulliam, A D Luis, R K Plowright, T J O Shea, D T S Hayman, J L N Wood, C T Webb, O Restif, and Proc R Soc B. The effect of seasonal birth pulses on pathogen persistence in wild mammal populations The effect of seasonal birth pulses on pathogen persistence in wild mammal populations. *Proceedings of The Royal Society Biological Sciences*, 281:1–9, 2014.

[150] PHAC. Varicella (Chickenpox). http://www.phac-aspc.gc.ca/im/vpd-mev/varicella-eng.php. Accessed Sept 15, 2015, September 2015.

[151] Virginia E Pitzer, Cécile Viboud, Lone Simonsen, Claudia Steiner, Catherine A Panozzo, Wladimir J Alonso, Mark A Miller, Roger I Glass, John W Glasser, Umesh D Parashar, and Bryan T Grenfell. Demographic Variability, Vaccination, and the Spatiotemporal Dynamics of Rotavirus Epidemics. *Science*, 325:290–294, July 2009.

[152] P. Poletti, A. Melegaro, M. Ajelli, E. del Fava, G. Guzzetta, L. Faustini, G.S. Tomba, P. Lopalco, C. Rizzo, S. Merler, and P. Manfredi. Perspectives on the impact of varicella immunization on herpes zoster. a model-based evaluation from three european countries. *PloS one*, 8(4):e60732, 2013.

[153] Eradication Polio. Polioeradication.org, 2014.

[154] A. Quetelet. Memoire sur les lois des naissance et de la mortalitie a bruxelles. *Nouveau Memoires de l'Academie Royale des Sciences et Belles Lettres de Bruxelles*, 3:495512, 1826.

[155] Jorge Quian, Ricardo Rüttimann, Claudia Romero, Patricia Dall Orso, Alfredo Cerisola, Thomas Breuer, Michael Greenberg, and Thomas Verstraeten. Impact of universal varicella vaccination on 1-year-olds in uruguay: 1997–2005. *Archives of Disease in Childhood*, 93(10):845–850, 2008.

[156] M.W. Ragozinno, L.J. Melton III, L.T. Kurland, C.P Chu, and H.O. Perry. Population-based study of herpes zoster and its sequelae. *Medicine*, 61(5):310–316, 1982.

[157] W Randall. A Statistical Analysis of the Annual Pattern in Birth in the USA, 1967 to 1976. *Journal of Interdisciplinary Cycle Research*, 18(3):179–191, 1987.

[158] G Rasch and W Hellenbrand. Germany adds varicella vaccine to the national vaccination programme. *Eurosurveillance*, 8, 2004.

[159] Gregory L Repass, William C Palmer, and Fernando F Stancampiano. Hand, foot, and mouth disease: Identifying and managing an acute viral syndrome. *Clevel*, 81(9):537–543, 2014.

[160] T. Richards. Weather, nutrition, and the economy: Short-run fluctuations in births, deaths, and marriages, france 17401909. *Demography*, 20 (2):197–212, 1983.

[161] John F. Rodis, Theodore J. Hovick Jr, Dorothy L. Quinn, Sally Shulman Rosengren, and Peter Tattersall. Human parvovirus infection in pregnancy. *Obstetrics & Gynecology*, 72:733–738, 1988.

[162] T. Roenneberg. The decline in human seasonality. *Journal of biological rhythms*, 19 (3):193–195, 2004.

[163] T. Roenneberg and J. Aschoff. Annual rhythm of human reproduction: I. biology, sociology, or both? *Journal of Biological Rhythms*, 5 (3):195–216, 1990.

[164] T. Roenneberg and J. Aschoff. Annual rhythm of human reproduction: Ii. environmental correlations. *Journal of Biological Rhythms*, 5 (3):217–239, 1990.

[165] P. Rohani, D.J. Earn, and B.T. Grenfell. Opposite patterns of synchrony in sympatric disease metapopulations. *Science*, 286:968–971, 1999.

[166] P. Rohani, M.J. Keeling, and B.T. Grenfell. The interplay between determinism and stochasticity in childhood diseases. *The American Naturalist*, 159:469–481, 2002.

[167] Pejman Rohani, Xue Zhong, and Aaron a King. Contact network structure explains the changing epidemiology of pertussis. *Science*, 330(6006):982–985, November 2010.

[168] N Rojansky, A Brzezinski, and J G Schenker. Seasonality in Human Reproduction: an Update. *Human Reproduction*, 7(6):735–745, July 1992.

[169] Harry M Rosenberg. Seasonal Variation of Births United States, 1933-63. *National Center for Health Statistics*, 21(09):1–59, 1966.

[170] LS Rosenblatt, M Shifrine, NW Hetherington, T Paglieroni, and ME MacKenzie. A circannual rhytm in rubella antibody titers. *Journal of Interdisciplenary Cycle Research*, 13(1):81–88, 1982.

[171] D. Russell, A. S. Douglas, and T. M. Allan. Changing seasonality of birth–a possible environmental effect. *Journal of epidemiology and community health*, 47 (5):362–367, 1993.

[172] M.L. Russell, D.P. Schopflocher, L. Svenson, and S.N. Virani. Secular trends in the epidemiology of shingles in alberta. *Epidemiological Infection*, 135:908–913, 2007.

[173] S. Saha, S. Moorthi, H. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y. Hou, H. Chuang, H. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. V. Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. van den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg. Ncep climate forecast system reanalysis (cfsr) monthly products, january 1979 to december 2010. Online Dataset. Accessed Aug 4th, 2016.

[174] S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker. Ncep climate forecast system version 2 (cfsv2) monthly products. Online Dataset, 2012. Accessed Aug 4th, 2016.

[175] Marcel Salathé, Linus Bengtsson, Todd J. Bodnar, Devon D. Brewer, John S. Brownstein, Caroline Buckee, Ellsworth M. Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L. Mabry, and Alessandro Vespignani. Digital epidemiology. *PLoS Computational Biology*, 8(7):1–5, 2012.

[176] J.E. Salk. Poliomyelitis vaccine in the fall of 1955. *American Journal of Public Health and the Nation's Health*, 46.1, 1956.

[177] A Sauerbrei. Diagnosis, antiviral therapy, and prophylaxis of varicella-zoster virus infections. *European Journal of Clinical Microbiol Infectious Diseases*, 2016.

[178] N.M. Sawtell and R.L. Thompson. Rapid in vivo reactivation of herpes simplex virus in latently infected murine ganglionic neurons after transient hyperthermia. *Journal of Virology*, 66:2150–2156, 1992.

[179] J. Schachter, D. E. Waggoner, and P. K. Whelpton. Short range birth projections. *Public Health Reports*, 73 (11):989–997, 1958.

[180] Dieter Schenzle. An Age-Structured Model of Pre- and Post-Vaccination Measles. *Journal of Mathematics Applied in Medicine and Biology*, 1:169–191, 1984.

[181] M. Schueete. Modeling the effects of varicella vaccination programs on the incidence of chickenpox and shingles. *Bulletin of Mathematical Biology*, 61:1031–1064, 1999.

[182] FT Scott, RW Johnson, M Leedham-Green, E Davies, WJ Edmunds, and J Breuer. The burden of herpes zoster: A prospective population based study. *Vaccine*, 24(9):1308–1314, 2006.

[183] W.T. Sedgewick and C.A. Winslow. Statistical studies on the seasonal prevalence of typhoid fever in various countries and its relation to seasonal temperature. *Memoirs of the American Academy of Arts and Sciences, New Series*, 12:467–571, 1902.

[184] Daniel A Seiver. Trend and Variation in the Seasonality of U.S. Fertility, 1947-1976. *Demography*, 22(1):89–100, 1985.

[185] Nitu Sengupta and Judy Brewer. A global perspective of the epidemiology and burden of varicella-zoster virus. *Current Pediatric Reviews*, 5.4:207–228, 2009.

[186] R.E. Serfling and I.L. Sherman. Poliomyelitis distribution in the united states. *Public Health Reports*, 68.5:453, 1953.

[187] Jane Seward, Karin Galil, and Melinda Wharton. *Epidemiology of varicella.* Cambridge University Press, 2000.

[188] Jane F Seward, Barbara M Watson, Carol L Peterson, Laurene Mascola, Jan W Pelosi, John X Zhang, Teresa J Maupin, Gary S Goldman, Laura J Tabony, Kimberly G Brodovicz, Aisha O Jumaan, and Melinda Wharton. Varicella disease after introduction of varicella vaccine in the United States, 1995-2000. *Journal of the American Medical Association*, 287(5):606–611, 2002.

[189] Jeffrey Shaman and Alicia Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50):20425–30, 2012.

[190] Jeffrey Shaman, Virginia E Pitzer, Cécile Viboud, Bryan T Grenfell, and Marc Lipsitch. Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States. *PLoS Biology*, 8(2):e1000316, February 2010.

[191] A Siedler and U Arndt. Varicella sentinel in Germany shows success of vaccination. 26th Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID). In *26th Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID)*, 2008.

[192] H E Soper. The Interpretation of Periodicity in Disease Prevalence. *Journal of the Royal Statistical Society*, 92(1):34–73, 1929.

[193] C. C. Spicer. Influence of some meteorological factors in the incidence of poliomyelitis. *British journal of preventive & social medicine*, 13.3:139–144, 1959.

[194] Alfred Spira. Epidemiologic Aspects of the Relationship Between Temperature and Male Reproduction. *Advances in Experimental Medicine and Biology*, 286:49–56, 1991.

[195] C.O. Stallybrass. Season and disease. *Proceedings of the Royal Society of Medicine: Section of Epidemiology and State Medicine*, pages 1185–1210, 1928.

[196] TJ Stevenson, ME Visser, W Arnold, P Barrett, S Biello, A Dawson, DL Denlinger, Davide Dominoni, FJ Ebling, S Elton, et al. Disrupted seasonal biology impacts health, food security and ecosystems. *Proc. R. Soc. B*, 282(1817):20151453, 2015.

[197] Tyler J Stevenson and Brian J Prendergast. Photoperiodic time measurement and seasonal immunological plasticity. *Frontiers in Neuroendocrinology*, 37:76–88, 2015.

[198] J. Stoeckel and A. K. M. Choudhury. Seasonal variation in births in rural east pakistan. *Journal of biosocial science*, 4 (01):107–116, 1972.

[199] Lewi Stone, Ronen Olinky, and Amit Huppert. Seasonal Dynamics of Recurrent Epidemics. *Nature*, 446(7135):533–536, March 2007.

[200] Lamont Sweet, Peggy Gallant, Marie Morris, and Scott A Halperin. Canada's first universal varicella immunization program: Lessons from Prince Edward Island. *The Canadian Journal of Infectious Diseases (Journal Canadien des Maladies Infectieuses)*, 14(1):41–44, 2003.

[201] M J Tauber, C A Tauber, and S Masaki. *Seasonal adaptations of insects*, volume 132. 1986.

[202] Thai-Government. National Disease Surveillance. http://www.boe.moph.go.th/boedb/surdata/index.php. Accessed April 12, 2015, 2015.

[203] The Lancet. The Influence of Schools in Accentuating the Spread of Certain Infectious Diseases - I. *The Lancet*, 153(3934):184–185, 1899.

[204] The Lancet. The Influence of Schools in Accentuating the Spread of Certain Infectious Diseases - II. *The Lancet*, 153(3935):256, 1899.

[205] The Lancet. The Influence of Schools in Accentuating the Spread of Certain Infectious Diseases - III. *The Lancet*, 153(3936):330–331, 1899.

[206] H. J. Theunissen, Nicole A. Lemmens den Toom, A. Burggraaf, E. Stolz, and M. F. Michel. Influence of temperature and relative humidity on the survival of chlamydia pneumoniae in aerosols. *Applied and environmental microbiology*, 59(8):2589–2593, 1993.

[207] Sara L Thomas, Jeremy G Wheeler, and Andrew J Hall. Contacts with varicella or with children and protection against herpes zoster in adults: a case-control study. *The Lancet*, 360(9334):678–682, 2002.

[208] SL Thomas and AJ Hall. What does epidemiology tell us about risk factors for herpes zoster? *The Lancet*, 4:26–33, 2001.

[209] TJ Török, PE Kilgore, MJ Clarke, RC Holman, JS Bresee, and RI Glass. Visualizing Geographic and Temporal Trends in Rotavirus Activity in the United States , 1991 to 1996. *The Pediatric Infectious Disease Journal*, 16(10):941–946, 1997.

[210] Christopher Torrence and GP Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.

[211] Hung Fu Tseng, Maragart Chi, Ning Smith, Stephen M Marcy, Lina S. Sy, and Steven J. Jacobsen. Herpes zoster vaccine and the incidence of recurrent herpes zoster in an immunocompetent elderly population. *The Journal of Infectious Diseases*, 206:190–196, 2012.

[212] A. Uberoi, Yoshida S., Frazer I.H., and Pitot H.C.and P.F. Lambert. Role of ultraviolet radiation in papillomavirus-induced disease. *PLoS Pathogens*, 12(5):0, 2016.

[213] UN. http://esa.un.org/unpd/wpp/ Accessed June 18, 2015, 2015.

[214] United Nations. Worldwide Births. http://data.un.org, 2012. Accessed: May 1, 2012.

[215] United States Census Bureau. United States Population Estimates: Historical Estimates. http://www.census.gov/popest/data/historical/index.html, 2010. Accessed: March 1, 2010.

[216] United States Centers for Disease Control and Prevention. United States Vital Statistics. http://www.cdc.gov/nchs/products/vsus.htm, March 2010. Accessed: March 1, 2010.

[217] A.J. van Hoek, A. Melegaro, E. Zagheni, W.J. Edmunds, and N. Gay. Modelling the impact of a combined varicella and zoster vaccination programme on the epidemiology of varicella zoster virus in england. *Vaccine*, 29:2411–2420, 2011.

[218] Willem G van Panhuis, John Grefenstette, Su Yon Jung, Nian Shong Chok, Anne Cross, Heather Eng, Bruce Y Lee, Vladimir Zadorozhny, Shawn Brown, Derek Cummings, and Donald S Burke. Contagious diseases in the United States from 1888 to the present. *New England Journal of Medicine*, 369(22):2152–2158, 2013.

[219] Arely Vergara-Castañeda, Alejandro Escobar-Gutiérrez, Karina Ruiz-Tovar, Julio Sotelo, Graciela Ordoñez, Mayra Yolanda Cruz-Rivera, Salvador Fonseca-Coronado, Armando Martinez-Guarneros, Juan C Carpio-Pedroza, and Gilberto Vaughan. Epidemiology of varicella in mexico. *Journal of Clinical Virology*, 55(1):51–57, 2012.

[220] P Villerme. De la Distribution par Mois des Conceptions. *Annales d'Hygiène Publique, Industrielle et Sociale*, 5:55–155, 1831.

[221] D.J. Vugia, C.L. Peterson, H.B. Meyers, K.S. Kim, A. Arrieta, P.M. Schlievert, E.L. Kaplan, S.B. Werner, and L. Mascola. Invasive group a streptococcal infections in children with varicella in southern california. *The Pediatric infectious disease journal*, 15(2):146–150, 1996.

[222] ME Wells. The seasonal patterns of measles and chicken pox. *American Journal of Epidemiology*, pages 279–317, 1944.

[223] WHO. Combined immunization of infants with oral and inactivated poliovirus vaccines: Results of a randomized trial in the gambia, oman, and thailand. *WHO*, 1996.

[224] WHO. WHO Model List of Essential Medicines. http://www.who.int/medicines/publications/essentialmedicines/en/. Accessed Sept 15, 2015., April 2013.

[225] WHO. Background Paper on Varicella Vaccine: SAGE Working Group on Varicella and Herpes Zoster Vaccines, April 2014.

[226] WHO. World Health Organization, 2014.

[227] WHO. Immunization surveillance, assessment and monitoring/Vaccine-preventable diseases. http://apps.who.int/immunization_monitoring/diseases/en/. Accessed Sept 15, 2015, September 2015.

[228] WHO. WHO Webpage of Country Level Annual Vaccine Uptake Estimates. http://who.int/immunization. Accessed July 7, 2017., May 2017.

[229] A.C. Wilson and I. Mohr. A cultured affair: Hsv latency and reactivation in neurons. *Trends in Microbiology*, 20:604–611, 2012.

[230] J.B. Wilson. Thirty one years of herpes zoster in a rural practice. *British Medical Journal*, 293:1349–1351, 1986.

[231] J C Wingfield and D S Farner. Control of seasonal reproduction in temperate-zone birds. *Seasonal reproduction in higher vertebrates*, pages 62–101, 1980.

[232] M. E. Wise. Poliomyelitis and weather during 10 years in england and wales. *International journal of biometeorology*, 10.1:77–90, 1966.

[233] G. Worrall. Herpes labialis. *Clinical Evidence*, 9:1704, 2009.

[234] Barbara P. Yawn and Don Gilden. The global epidemiology of herpes zoster. *Neurology*, 81:928–930, 2013.

[235] W.K. Yih, D.R. Brooks, S.M. Lett, A.O. Jumaan, Z. Zhang, K.M. Clements, and J.F. Seward. The incidence of varicella and herpes zoster in massachusetts as measured by the behavioral risk factor surveillance system (brfss) during a period of increasing varicella vaccine coverage, 19982003. *BMC Public Health*, 5(1):68, 2005.

[236] Allen M. Young and John H. Thomason. The demography of a confined population of the butterfly morpho peleides during a tropical dry season. *Studies on Neotropical Fauna and Environment*, 9 (1):1–34, 1974.

[237] F. Yu, J.N. Harada, H.J. Brown, H. Deng, M.J. Song, T.T. Wu, J. Kato-Stankiewicz, C.G. Nelson, J. Vieira, F. Tamanoi, and S.K. Chanda. Systematic identification of cellular signals reactivating kaposi sarcoma associated herpesvirus. *PloS pathogens*, 3(3):e44, 2007.

[238] M Zak-Prelich, JL Borkowski, F Alexander, and M Norval. The role of solar ultraviolet irradiation in zoster. *Epidemiological Infection*, 129(3):593–597, 2002.

[239] J. Zhong, M. Rist, L. Cooper, C. Smith, and R. Khanna. Induction of pluripotent protective immunity following immunisation with a chimeric vaccine against human cytomegalovirus. *PloS one*, 3:e3256, 2008.