



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

MODELLING AND STATISTICAL
INFERENCE OF INFECTIOUS DISEASES

QIANYING LIN

PhD

THE HONG KONG POLYTECHNIC UNIVERSITY

2019

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF APPLIED MATHEMATICS

MODELLING AND STATISTICAL INFERENCE
OF INFECTIOUS DISEASES

QIANYING LIN

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

April 2019

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

 LIN Qianying (Name of student)

Dedicate to Maggie Tsz-Zing Tsui.

Abstract

Emerging infectious diseases (EIDs) have been an enormous burden on public health all around the world. They share the property of rapid spread among countries and continents, which potentially cause enormous harm to human society. Using statistical inferential methods and mathematical models, researchers have been succeeding in exploring the underlying dynamics of EIDs, as well as forecasting their future trends.

With reported cases of Influenza A/H1N1 and A/H3N2 from January 2010 and July 2017 from World Health Organization, we explored and verified the anti-phase temporal pattern of these two subtypes both in Hong Kong and North Temperate Zone, which indicates a high prevalence of A/H1N1 coincides with a low prevalence of A/H3N2 and vice versa. By constructing a statistical metric, i.e., weekly case ratio of influenza B over influenza A, we showed a strong positive correlation with the activity of B/Victoria in China, abnormality of extremely high ratio accompanied by unusual low vaccine effective in the United States, as well as revealed the global pattern of influenza activity by computing the statistical correlation of the ratios between countries and regions.

We investigated heterogeneities in demographic information of tuberculosis (TB) notifications and estimate the effects of age and period (historical trend) on TB and drug-resistant (DR)-TB across cities in Shandong from 2006–2017 and

from 2004–2017, respectively, by age-period-cohort (APC) models. We observed a drastic shift from transmitted TB to reactivated TB, especially among women. We forecast the future trend of TB from 2018–2027 and of DR-TB from 2018–2023 and predicted 1.5–3.5 multi-drug-resistant (MDR)-TB cases per population of 100 thousand by the end of 2023.

Furthermore, using mathematical models (a plug-and-play likelihood-based inference framework), we fitted a susceptible-exposed-infectious-recovered-susceptible (SEIRS) model of camels to the reported human cases with a constant proportion of human cases from camels (i.e., either 25% or 12%). We considered two scenarios: (i) the transmission rate among camels is time-varying with a constant spill-over rate from camels to human or (ii) the spill-over rate is time-varying with a constant transmission rate among camels. Our estimated loss-of-immunity rate and prevalence of Middle East respiratory syndrome coronavirus infections among camels mostly matched with previous serological or virological studies, shedding light on this issue. We recommended including dromedary camels in animal surveillance and control of Middle East respiratory syndrome coronavirus in Saudi Arabia which could help reduce their sporadic introductions to humans.

“Phylodynamics” indicates methods or frameworks that make the statistical inferences from both epidemiological and phylogenetic data. By Coalescent Theory, phylodynamics has become a hot topic recently, while theoretical studies that clarify its feasibility and practical studies that explore the inconsistency and its impact on inference are lacking. We utilized the integrated framework via iterated filtering method within the plug-and-play simulation based scheme and estimated the parameters of interest and their 95% confidence intervals through a global search for the maximum likelihood. We found failures in estimating, which is

probably caused by inappropriate likelihood expression and is consistent with previous studies. We then discussed the inconsistency between coalescent theory and stochastic models, as well as its theoretical proof.

Acknowledgements

Carrying out research is always a fascinatingly non-isolated activity. Here, I express my sincere gratitude and regards to several individuals who have supported me in various aspects during my Ph.D. study.

First and foremost, I would like to thank my Chief Supervisor, Dr. Daihai He, for his patience and encouragement. He helped me all the time of research and writing of this thesis. This research would not be possible without his selfless support.

I would also like to express my sincere gratitude to my advisor, Professor Aaron A. King, during my visit in the University of Michigan, Ann Arbor, for his enlightening guidance and kind assistance in conducting this project.

Furthermore, the essential support of my Ph.D. experience has been the guidance and kindness of my Co-Supervisor, Dr. Cedric Yiu, who has been a source of inspiration and encouragement.

I would also like to thank my friends, research fellows at the Hong Kong Polytechnic University and my parents, for their love and support.

Contents

| | |
|---|----------|
| Certificate of Originality | i |
| Acknowledgements | ix |
| List of Tables | xv |
| List of Figures | xvii |
| Notation | xxvii |
| 1 Introduction | 1 |
| 1.1 Statistical inference and mathematical modeling for infectious diseases | 1 |
| 1.2 Phylodyanmics | 3 |
| 1.3 Objectives, methods, findings and conclusions | 4 |
| 1.4 Outline of this thesis | 4 |
| 2 Influenza | 7 |
| 2.1 Introduction on influenza | 7 |
| 2.2 Data and methods | 9 |
| 2.2.1 Data | 9 |
| 2.2.2 Methods | 9 |
| 2.3 Results | 10 |
| 2.4 Discussion and conclusions | 20 |

| | | |
|----------|---|-----------|
| 3 | Tuberculosis (TB) | 23 |
| 3.1 | TB in China | 23 |
| 3.2 | Data and methods | 25 |
| 3.2.1 | Data | 25 |
| 3.2.2 | Statistical analysis and models | 26 |
| 3.3 | Heterogeneities in TB | 27 |
| 3.3.1 | Heterogeneities in TB incidence between cities. | 27 |
| 3.3.2 | Heterogeneities in historical and future trends. | 32 |
| 3.3.3 | Current situation and forecast trends for DR-TB. | 34 |
| 3.4 | Conclusion and discussion | 37 |
| 3.5 | Future works | 41 |
| 4 | Middle East Respiratory Syndrome (MERS) | 43 |
| 4.1 | Introduction on MERS-CoV | 43 |
| 4.2 | Data and mathematical modeling | 47 |
| 4.2.1 | Data | 47 |
| 4.2.2 | Mathematical modeling | 47 |
| 4.3 | Results | 52 |
| 4.4 | Conclusion | 57 |
| 5 | Phylodynamics | 63 |
| 5.1 | Moran model, coalescent theory, state space model | 63 |
| 5.2 | Statistical framework in iterated filtering | 66 |
| 5.2.1 | An integrated framework | 67 |
| 5.2.2 | Iterated filtering | 68 |
| 5.3 | Simulation and results | 69 |

| | | |
|----------|--|-----------|
| 5.4 | Discussion and future works | 74 |
| 5.4.1 | Stationarity of Moran model and coalescent theory | 75 |
| 5.4.2 | Asynchronous sampling and simulation | 76 |
| 5.4.3 | Improved likelihood | 76 |
| 6 | Conclusion and Discussion | 79 |
| | Appendices | 81 |
| A | | 81 |
| A.1 | Influenza | 81 |
| A.1.1 | Time series decomposition analyses | 81 |
| A.1.2 | Wavelet and reconstruction analyses | 83 |
| A.1.3 | Time series prediction rule | 85 |
| A.2 | Tuberculosis in Shandong, China | 86 |
| A.2.1 | Data Preprocessing | 86 |
| A.2.2 | Heterogeneity in TB incidence in other three cities. | 88 |
| A.2.3 | Further investigations in heterogeneities across seven cities. | 91 |
| A.2.4 | Description of DR-TB notification in Shandong | 95 |
| A.2.5 | Estimation and forecast of DR-TB cases | 97 |
| A.2.6 | Drug correlation | 99 |
| A.3 | MERS-CoV | 99 |
| A.3.1 | Assuming Smaller Proportion of Infections Reservoir | 99 |
| A.3.2 | Extension to a Two-Host Model | 103 |
| A.3.3 | Climatic Factors | 104 |
| A.3.4 | Equilibrium | 108 |
| A.4 | Phylodynamics | 113 |

| | |
|---|------------|
| A.4.1 Implementation of the model | 113 |
| Bibliography | 121 |

List of Tables

| | | |
|-----|---|-----|
| 3.1 | Summary of TB notifications across four cities in Shandong. | 28 |
| 3.2 | Summary of drug sensitivity and resistance to four antituberculosis drugs in Shandong. | 35 |
| 4.1 | Summary of parameter settings and estimates in Figs 3-5. The mean basic reproductive number is defined as $\mathcal{R}_0 = \langle \beta \rangle / \gamma$, where $\langle \cdot \rangle$ is the time average. $S.0$, $E.0$, $I.0$, and $R.0$ are initial conditions; In Figs 3 and 4, ρ is assumed constant and $\beta(t)$ is time-dependent, whereas these are reverse in Fig 5. Camel population size $N=270,000$. $\sigma^{-1} = 2$ days. $\gamma^{-1} = 4$ days. | 59 |
| 5.1 | Summary of the parameters and 95% confidence intervals. | 73 |
| A.1 | Summary of TB notifications across three cities in Shandong | 88 |
| A.2 | Summary of parameter settings and estimates in Figure S2-S4. Same model input parameters are used as in Table 1, except that the proportion of infections from camels is 12%. | 102 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | Temporal patterns of A/H1N1 and A/H3N2 in (a) Hong Kong, and (b) the North Temperate Zone. (c) Spatial patterns of the annual totals of A/H1N1 and A/H3N2 in Hong Kong and 24 countries and regions in the North Temperate Zone, which are ranked according to their longitudes. It is evident that A/H1N1 and A/H3N2 alternately dominated in the recent 4years. A/H1N1 skipped (<i>i.e.</i> , very low numbers of cases) in most countries or regions in 2011/2012 and 2016/2017. Since A/H1N1 resurged in 2012/2013, a resurgence could happen in 2017/2018. | 11 |
| 2.2 | Temporal patterns of reported influenza A/H1N1 and A/H3N2 cases and their reconstructed time-series in (a) Hong Kong, and (b) the North Temperate Zone. Arrows indicate the remarkable anti-phase patterns from 2015 to 2017 in Hong Kong, and from 2014 to 2017 in the North Temperate Zone. | 12 |
| 2.3 | The logarithmic ratios of influenza B over influenza A in mainland China and Hong Kong (a) and the intensities of influenza B lineages in mainland China (b). | 13 |
| 2.4 | Cross wavelet analysis of the ratio with Victoria lineage intensity in China. (a) phase of the ratio and Victoria lineage intensity; (b) period power spectrum; (c) cross wavelet power spectrum. | 14 |
| 2.5 | Pearson's correlations of the ratios between countries or regions. The horizontal axis is the correlation between different countries/regions and Hong Kong. The vertical axis is the correlation between different countries/regions and the United Kingdom (UK). | 16 |

| | | |
|-----|---|----|
| 2.6 | The ratio of influenza B over influenza A in 54 states/regions in the US from 2010/11 to 2017/18. The states/regions are listed in descending order according to the number of total influenza A confirmations from top to bottom. The color key represents the logarithmic ratios. | 17 |
| 2.7 | The adjusted overall VE (a) and weekly confirmations and ratios (a) in the United States, 2010/11-2017/18. Bars in (a) show the estimated adjusted overall VE. Black curve with dots in (b) shows weekly confirmations of Influenza A, while red curve with circles in (b) shows influenza B and the green curve shows the weekly ratio of influenza B over influenza A confirmations. The low VE in 2014/15 season coincided with the high peak of B/A ratio. The influenza B epidemic was severe in 2017-18 season, but the B/A was not high. | 19 |
| 3.1 | Changes in the notification rates for males and females from 2006–2017 in (a) Jinan, (b) Yantai, (c) Linyi, and (d) Liaocheng. In each panel, the blue line and red line represent notification rate trends; the population size for males and females in 2010 is shown in the top-right corner. | 30 |
| 3.2 | Case density changes in 2006 and 2017 for males (upper subpanels) and females (lower subpanels) in (a) Jinan, (b) Yantai, (c) Linyi, and (d) Liaocheng. In each panel, red dashed-and-dotted and green solid lines represent densities in 2006 and 2017, respectively. We show the changes in case proportions for young adults (18–35) and seniors (60+) in 2006 and 2017 in the top-right corner for males and in the bottom-right corner for females. | 31 |
| 3.3 | TB notification rates among migrants and migrant population sizes from 2007–2017 in (a) Jinan, (b) Linyi, and (c) Liaocheng. In each panel, bars represent a population with a unit of 100 thousand, and lines represent TB notifications per population of 100 thousand. | 32 |

| | | |
|-----|---|----|
| 3.4 | Trends of TB notification rates from 2006–2017 and forecasts from 2018–2027 in (a) Jinan, (b) Yantai, (c) Linyi, and (d) Liaocheng. In each panel, (i) upper and low subpanels show trends for males and females, respectively; (ii) dots, lines, and shades colored in red, blue, and black indicate reported trends, forecast trends, and 95% confidential intervals for forecasts of annual TB notifications (in the 100 thousand population) for populations aged 18–35 and over 60 years and the total population, respectively; (iii) red, blue, and black dashed lines are smoothed splines that indicate the trends of annual reported TB notifications for the populations aged 18–35 and over 60 years and the total population, respectively. | 34 |
| 3.5 | Reported trends from 2006–2017 and forecasts from 2018–2023 of DR-TB infections for (a) INH-resistant TB, (b) RFP-resistant TB, and (c) MDR-TB, among the total population in Shandong, China. Within each panel, dotted-and-dashed lines and solid lines colored in green/blue/purple indicate reported DR-TB infections and forecast trends, under assumed test ratios of 50%/75%/100%, respectively. We constructed APC models to predict future total DR-TB notifications and smoothed splines to predict the future population in Shandong. | 37 |
| 4.1 | SEIRS model structure | 49 |
| 4.2 | Spatio-temporal patterns of MERS-CoV human cases and percentage of MERS-CoV camels. The top panel summarises the cumulative global cases from September 2012 to May 2016. The bottom panel shows the weekly confirmed cases of MERS-CoV human cases worldwide (red lines with circles) and in Saudi Arabia only (black lines with diamonds). The bold black curve shows a monthly percentage of MERS-CoV infectious camels from May 2014 to April 2015. There were three waves in the MERS-CoV camels which were followed by three waves in the weekly human MERS-CoV cases. We also showed the period of religious events, Hajj (light grey vertical dash line) and Ramadan (shaded bars). The map was made with R programming language, and the country borders were downloaded from Sandvik B., World Borders Dataset, http://thematicmapping.org (2009). | 53 |

| | | |
|-----|---|----|
| 4.3 | Model simulations versus observed MERS-CoV spill-over cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and blue curve (with circles) represents the transmission rate (in units of \mathcal{R}_0). Inset panels on the left show the profile of BIC as a function of the number of nodes in the transmission rate. Inset panels on the right show the profile of MLL as a function of the spill-over rate ρ , while fixing $n_\beta = 10$ | 55 |
| 4.4 | Model simulations versus observed MERS-CoV spill-over cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and blue curve (with circles) represents the transmission rate (in units of \mathcal{R}_0). Inset panels show the profile of maximum log likelihood as a function of the duration of immunity protection. | 57 |
| 4.5 | Model simulations versus observed MERS-CoV spill-over cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and dashed blue curve represents the spill-over rate. Inset panels show the profile of BIC as a function of the number of nodes in the spill-over rate. | 58 |
| 5.1 | Demonstration of the Moran model. (A) Five lineages construct a genealogy. (B) At some moment, lineage 5 (in dashed line) is selected to be eliminated, while lineage 2 (in red) to be duplicated. (C) The new genealogy consists of five lineages, where two of them are duplicated. | 64 |
| 5.2 | Demonstration of the Coalescent theory. (A) The population genetics, (B) the population genealogy, (C) samples $n = 3$ from the population genealogy, (D) the sub-genealogy of the sample, and (E) the version of coalescent theory for the sub-genealogy. In panel (E), T_1 , T_2 , and T_3 represents the coalescent time between coalescent events, in backward time. | 65 |
| 5.3 | State space model. | 66 |
| 5.4 | Schematic of using iterated filtering method to Rasmussen et al. (2014) framework. | 69 |

| | | |
|-----|--|----|
| 5.5 | Number of reported cases (a) and lineages (b) in the simulated data. | 70 |
| 5.6 | Convergence of the estimations by local search, where <code>loglik</code> is the log-likelihood $l(\theta)$, <code>nfail</code> represents the number of particles which failed to converge, <code>R0</code> , <code>gamma</code> , <code>F</code> , <code>alpha</code> , <code>rho</code> , and <code>tau</code> are parameters of interest, and <code>mu</code> is constant birth/death rate, <code>P</code> is the constant total population size, <code>S_0</code> , <code>I_0</code> , and <code>R_0</code> is the fixed initial proportion of the susceptible, the infectious, and the recovered. . . . | 71 |
| 5.7 | Convergence of the estimations by global search. The notation is the same as those in figure 5.6. | 72 |
| 5.8 | Profile log-likelihood for estimations. In each panel, the blue line with grey shades indicates the profile of log-likelihood obtained by smooth spline, red dashed lines indicates the 95% confident log-likelihood and corresponding confidence interval, and black solid line refers to the real value. | 73 |
| 5.9 | Simulation using the best fit estimations. Red line indicates the data, blue line is the simulation from the best estimations. Blue shades indicates the 95% confidence interval of the simulation. . . . | 74 |
| A.1 | Decomposed seasonality and random components of influenza A subtypes in the NTZ. | 82 |
| A.2 | Decomposed seasonality and random components of influenza A subtypes in Hong Kong. | 83 |
| A.3 | Wavelet analyses for influenza A subtype data in the NTZ. | 84 |
| A.4 | Wavelet analyses for influenza A subtype data in the Hong Kong. | 85 |
| A.5 | Population and TB notifications from 2006 to 2017 in three cities in Shandong Province, China. We use a group of four subpanels to show the data for each city. The subpanels on left/right in each group, colored in blue/red, show data for male/female, respectively. The subpanels on the top show annual total TB notifications (bar) and their change rates (curve) from year to year, while the subpanels on the bottom show annual population size (bar) and TB notification rates (curve, per 100K population). | 88 |

| | | |
|------|---|----|
| A.6 | Case density changes in 2006 and 2017 for males (upper subpanels) and females (lower subpanels) in (a) Weifang, (b) Jining, and (c) Dezhou. In each panel, red dashed-and-dotted and green solid lines represent densities in 2006 and 2017, respectively. We show the changes in case proportions for young adults (18–35) and seniors (60+) in 2006 and 2017 in the top-right corner for males and in the bottom-right corner for females. | 89 |
| A.7 | The proportion of immigrant population from 2007 to 2017 in seven cities in Shandong, China. | 89 |
| A.8 | Trends of TB notification rates from 2006–2017 and forecasts from 2018 to 2027 in (a) Weifang, (b) Jining, and (c) Dezhou. In each panel, (i) upper and low subpanels show trends for males and females, respectively; (ii) dots, lines, and shades colored in red, blue, and black indicate reported trends, forecast trends, and 95% confidential intervals for forecasts of annual TB notifications (in the 100 thousand population) for populations aged 18–35 and over 60 years and the total population, respectively; (iii) red, blue, and black dashed lines are smoothed splines that indicate the trends of annual reported TB notifications for the populations aged 18–35 and over 60 years and the total population, respectively. | 90 |
| A.9 | The components of migrant TB notifications. | 91 |
| A.10 | The ratio of the elderly over the young from 2005 to 2017 in seven cities in Shandong, China. | 92 |
| A.11 | TB notifications against age from 2006 to 2017 for top five occupations in seven cities, from 2006 to 2017: student, worker, migrant worker, farmer and household or unemployed. Dashed-black lines divide the heatmaps into four sub-age groups: 0–17, 18–35, 35–60, and over 60. | 93 |

| | | |
|------|---|----|
| A.12 | Estimated effects of age, calendar period and birth cohort on TB notification rate from 2006 to 2017 in seven cities in Shandong, China. From top to bottom, each row show (a) age effects, (b) period effects, and (c) cohort effects from left to right, for each city (<i>i.e.</i> , Jinan, Yantai, Weifang, Jining, Linyi, and Dezhou), respectively. Within each panels, dots, lines, and transparent lines colored in blue/red show reported effects, estimated effects, and bootstrap simulations for estimated effects on annual TB notifications (per 100K population), for male/female, respectively. | 94 |
| A.13 | Numbers of (a) total collected notifications, (b) INH-resistant notifications, (c) RFP-resistant notifications and (d) MDR notifications from 2004 to 2017 in Shandong. The sudden increasing in 2011 was due to enhanced surveillance. | 96 |
| A.14 | Number of DR-TB notifications across 38 units in 37 regions in Shandong, from 2004 to 2017. | 96 |
| A.15 | Estimated effects of (a) age, (b) calendar period and (c) birth cohort on DR-TB tested notification rate from 2004 to 2017 in Shandong, China. Within each panel, dots, lines, and transparent lines colored in blue/brown/orange show reported effects, estimated effects, and bootstrap simulations for estimated effects on annual TB notifications (per 100K population), for INH-resistant TB/RFP-resistant TB/MDR-TB, respectively. | 97 |
| A.16 | Reported trends from 2004 to 2017 and forecasts from 2018 to 2023 of DR-TB tested notification rate in Shandong, China for INH-resistant TB, RFP-resistant TB, and MDR-TB, from top to bottom, respectively. Dots, lines, and shaded area colored in blue/brown/orange show reported tested notification rates (per 100 tested notifications), estimated tested notification rates, and 95% confidence interval for estimation, for INH-resistant TB/RFP-resistant TB/MDR-TB, respectively. Dashed lines colored in blue/brown/orange are smoothed splines for indicating the trends of reported DR-TB tested notification rate for INH-resistant TB/RFP-resistant TB/MDR-TB, respectively. Future total collected DR-TB records are predicted by APC models. | 98 |
| A.17 | Correlation between four types of commonly used drugs from 2004 to 2017. Darker color indicates higher correlation. | 99 |

| | | |
|------|--|-----|
| A.18 | Model simulations versus observed MERS-CoV spill-over cases. Same model input parameters are used as in Figure 3, except that the proportion of infections from camels is 12%. | 100 |
| A.19 | Model simulations versus observed MERS-CoV spill-over cases. Same model input parameters are used as in Figure 4, except that the proportion of infections from camels is 12%. | 101 |
| A.20 | Model simulations versus observed MERS-CoV spill-over cases. Same model input parameters are used as in Figure 5, except that the proportion of infections from camels is 12%. | 102 |
| A.21 | Two-host model simulations versus observed MERS-CoV human cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and blue curve (with circles) represents the transmission rate (in units of \mathcal{R}_0). Inset panels show the profile of MLL as a function of the spill-over rate. $\mu_h^{-1} = 7$ days, and $r = \frac{3}{4}\mu_h$ | 104 |
| A.22 | Wavelet Analysis for MERS-CoV cases and temperature in Saudi Arabia. (a) Time series of weekly reported MERS-CoV cases and weekly average temperature in Saudi Arabia. (b) Left panel is the power spectrum of wavelet coherency analysis for reported MERS-CoV cases and temperature, where red color represents high power and blue for low power; the angle of the arrows indicates the phase and those amidst white line indicates significant phase difference at 5% confidence level; white shades indicates boundary effects. Right panel shows the cross-wavelet average power for different wavelet periods and red dots and blue dots show significance at 5% and 10% levels respectively.(c) Phase and phase difference at wavelet period of 8 to 13 weeks (i.e. 0.15 to 0.25 year); red line and blue line represent temperature and MERS-CoV human cases; black-dashed line indicates the phase difference. | 106 |

A.23 Wavelet Analysis for MERS-CoV cases and absolute humidity in Saudi Arabia. (a) Time series of weekly reported MERS-CoV cases and weekly average absolute humidity in Saudi Arabia. (b) Left panel is the power spectrum of wavelet coherency analysis for reported MERS-CoV cases and absolute humidity, where red color represents high power and blue for low power; the angle of the arrows indicates the phase and those amidst white line indicates significant phase difference at 5% confidence level; white shades indicates boundary effects. Right panel shows the cross-wavelet average power for different wavelet periods and red dots and blue dots show significance at 5% and 10% levels respectively.(c) Phase and phase difference at wavelet period of 1 year; red line and blue line represent absolute humidity and MERS-CoV human cases; black-dashed line indicates the phase difference. 107

Notation

| | |
|-----------------|---|
| S | Susceptible. |
| E | Exposed. |
| I | Infectious. |
| R | Recovered. |
| N | Population Size. |
| i | Number of lineages in the sample. |
| \mathcal{R}_0 | The basic reproduction number. |
| μ | Birth/Death Rate. |
| β | Transmission Rate. |
| σ^{-1} | Mean Latent Period. |
| γ^{-1} | Mean Infectious Period. |
| λ^{-1} | Mean Lifetime in SEIRS model. |
| λ_t | Coalescent rate in phylodynamics framework. |

Chapter 1

Introduction

Emerging infectious diseases (EIDs) have been an enormous burden on public health all around the world. They share the property of rapid spread among countries and continents, which potentially cause enormous harm to human society. In order to study these EIDs and take effective measures and strategies to control and prevent their epidemics, statistical inference methods, mathematical modelings, as well as bioinformatic techniques have been widely utilized and help researchers achieve successes.

This chapter will first introduce several background knowledge about statistical inference, mathematical modeling, and phylodynamics, and how researchers apply them to investigate EIDs. Afterward, we will state the objectives and goals of this thesis, highlight the main findings, and share some essential conclusions in this study.

1.1 Statistical inference and mathematical modeling for infectious diseases

We are in a great era of “Big Data,” which refers to data that have the four properties of high volume, velocity, great variety, and veracity (See IBM Big Data

& Analytics Hub (2018)). With the increasing amount and variety of EIDs related data (*i.e.*, epidemic, genetic, geographic data, etc.), the statistical inference has become one of the most useful tools to analyze, hypothesize, test, estimate, infer, and draw conclusions on EIDs.

Unlike descriptive statistics, which explores solely on observed data, inferential statistic makes assumptions about a population, samples data from the population, and expects to draw inferences from these samples, based on selected statistical models.

On the basis of probability theory, statistical inference can be categorized into four common forms: i) point estimation, which concerns the particular value that approximates the concerned parameter best; ii) interval estimation, which would contain the true parameter with probability under multiple samplings; iii) hypothesis testing and iv) clustering or classification (See Johnson (2010)). With statistical inference methods, one can assess the risk of infection of concerned factors in epidemiology (See Longini et al. (1987)).

To understand the underlying mechanism of the spread of infectious diseases, deterministic and stochastic models are of great importance (See Anderson and May (1991); Andersson and Britton (2000); Keeling and Rohani (2008)). In particular, epidemiological compartmental models, such as Susceptible-Infectious-Recovered (SIR) model (See Kermack and McKendrick (1927)), divide the whole population into multiple compartments and study the dynamics of the disease between compartments via differential equation system. These models not only simplifies the complicated interaction in a population but helps to estimate key parameters, which describe the properties of the diseases, as well as the behavioral pattern of the population. Specifically, the basic reproduction number (\mathcal{R}_0),

which is the expected number of cases that one infection produces in a completely susceptible population. This key parameter assesses the severity of the epidemic, given that $\mathcal{R}_0 < 1$ indicates the infection will die out in the long run, and $\mathcal{R}_0 > 1$ refers to pandemic.

To combine the observed epidemic data and underlying compartmental models and estimate parameters of interest, State Space Models (SSMs), as known as Hidden Markov Model (HMM) or Partially Observed Markov Process (POMP) are widely used (See Baum and Petrie (1966); Baum and Eagon (1967); Baum and Sell (1968); Baum et al. (1970); Baum (1972)). This kind of models consider the dynamics of the diseases, modeled by compartmental models, as a serial of hidden states which correlate to each other, while the reported data as observed states, which are solely linked to the corresponding hidden states.

1.2 Phylodyanmics

With the great success in bioinformatic techniques, “phylodynamics” has become a hot topic in the field of Epidemiology. The term “phylodynamics” consist of two parts: phylogenetic and epidemic dynamics, in which the former one refers to the evolution of the diseases in the aspect of genes, while the latter one concerns the spreading of the diseases within a population. The general assumption for phylodynamics is, the phylogenetic procedure could completely reflect the transmission of diseases. Therefore, research can regard the phylogenetic tree as a sample of the transmission tree in the population (See Grenfell et al. (2004)). Among efforts to link phylogenies and epidemiology, the “coalescent theory” (See Kingman (1982)) helps to track the branch lengths between branchings (coalescent events) of the phylogenies, and shed light to infer epidemic dynamics with genetic information

using statistical methods (See Pybus et al. (2000); Volz et al. (2009); Rasmussen et al. (2014)).

1.3 Objectives, methods, findings and conclusions

This study aims to investigate the relationships between types of diseases, between hosts of diseases, between cities that diseases are spreading, and between data sources of diseases. We try to use various statistical methods and mathematical models to study the factors that make an impact in epidemics, as well as the underlying transmission mechanism. We also intend to infer disease dynamics by aggregating different types of data. Eventually, we hope to provide the public health departments with helpful suggestions in disease controlling and prevention strategies.

1.4 Outline of this thesis

This thesis is organized by the variety of the data and the complexity of the methods implemented to analyze the epidemic-related data. It is easy to think about applying the same method to all kind of data and conduct inference. However, the variety of epidemic data enrich the number of fields of exploration, from the simple pattern of changes, to impacts of some individual factors, to the trends of key epidemic parameters, and eventually, to the genetic evolution and mutations. The selection of methods not only depend on the variety of data, but on the complexity (*i.e.*, computation resources and time) we can afford. That is the vision of data science in the era of “Big Data”, to fully employ all accessible data and explore the underlying mechanism as much as we can afford.

In Chapter 2, we use simple statistical inferential methods to explore the tem-

poral antiphase pattern between subtypes of influenza A in Hong Kong and North Temperate Zone, as well as construct a simple metric to indicate the abnormal relationship between vaccine implementation and influenza epidemics all around the world.

In Chapter 3, We investigate heterogeneities in demographic information and TB notifications and estimate the effects of age and period (historical trend) on tuberculosis (TB) and drug-resistant (DR)-TB across cities in Shandong from 2006–2017 and from 2004–2017, respectively, by age-period-cohort (APC) models. This chapter shows how an uncomplicated statistical model help researchers discover underlying connections between disease dynamics and patient age, infected calendar year, and patient birth cohort. Through statistical inference, we can further inspect the fundamental reasons that drive heterogeneities of epidemic between regions.

Chapter 4 demonstrates an example of applying mathematical stochastic models to infectious diseases. Using ordinary differential equation (ODE) system, we not only model the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) dynamics among one host (*i.e.*, camel), but also the interaction between two hosts (*i.e.*, camel and human). With statistical inference method, iterated filtering, within a plug-and-play framework, which utilizes one simulation results to proceed estimating process, we disclose the mechanism of MERS transmission among camel, as well as the spillover from camel to human society.

Chapter 5 introduce the concept of “phylogenetics” and “Coalescent Theory” and its practical frameworks to aggregate both epidemiological methods and phylogenetic tools, with a numerical simulation. We demonstrate the fundamental structure of the framework and achieve it via iterated filtering method in a

plug-and-play framework. Furthermore, we discuss the underlying inconsistency.

Chapter 2

Influenza

In this chapter, we focus on the epidemiology of influenza in Hong Kong, and compare it with other regions and countries in the world.

2.1 Introduction on influenza

Influenza affects 10% of the global population annually (See Gerdil (2003)). Seasonal influenza has been a widespread threat to public health around the world, causing illness and leading to a great amount of hospitalizations and mortalities. Influenza A and B are two common types of influenza viruses, where A/H1N1, A/H3N2 are two major subtypes of influenza A, while influenza B is broken down into two lineages: Victoria and Yamagata. A/H3N2 strains typically have a higher case-fatality ratio among the elderly and evolve at higher rate globally. While A/H1N1 strains have been stable since 2009 and showed a large-scale. A/H3N2 typically has a higher case-fatality ratio among the elderly than A/H1N1, and it have evolved at a higher rate globally (See Paules and Subbarao (2017)). In contrast, A/H1N1 strains have been stable since 2009 (See Klein et al. (2014)) and have shown a large-scale skip-and-resurgence phenomenon (See He et al. (2015)), *i.e.*, minor or no outbreaks in 2011/2012 and large outbreaks in 2012/2013 in

Europe and Asia. In Hong Kong (HK) Special Administrative Region China, influenza-associated mortalities in recent four years are 199 (2014), 580 (2015), 271(2016) and 360 (2017, up to July), and the activity of influenza showed a biennial pattern in recent years, namely influenza A/H3N2 was severe in 2013, 2015 and 2017, while influenza A/H1N1 and Influenza B were severe in 2014, 2016 and 2018. Rarely, influenza B dominated in 2018 in Hong Kong. In particular, influenza epidemics in Hong Kong are largely in synchrony with those in Mainland China, as well as those in other countries in the North Temperate Zone(NTZ) (*i.e.*, countries with latitude between 23.5N and 66.5N). The influenza vaccine is designed annually against three strains (Trivalent) or four strains (Quadrivalent) (<https://www.who.int/influenza/vaccines/en/>).The overall vaccine efficacy (VE) is above 40% generally (<http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>). Influenza vaccines are widely used to control the outbreaks and 151 million to 166 million doses are estimated to be available for the 2017-2018 season worldwide. By prognosis from research for the most popular strains of three flu viruses (*i.e.*, influenza A/H1N1, A/H3N2 and influenza B), the trivalent seasonal flu vaccines will be manufactured to protect against these viruses. However, concerns about the vaccine effectiveness (VE) were raised in recent studies. The VE was estimated as low as 17% against A(H3N2) and 55% against influenza B in Canada (See Skowronski et al. (2018)), while only 10% against A(H3N2) in Australia (See Paules et al. (2018)). These estimations have raised alarm about VE and made it crucial to study the relationship between influenza outbreaks and VE.

In this study, we first investigated the anti-phase synchronization in Hong Kong and the NTZ, defined as the alternative dominance of A/H1N1 and A/H3N2 (See He and Stone (2003)). Afterwards, we calculated the ratio of influenza B over

influenza A and estimate the correlation this ratio in different places to reveal the global pattern of influenza activity.

2.2 Data and methods

2.2.1 Data

We included various streams of data from public domains:

1. country and region-level influenza laboratory confirmation for 158 countries and regions from FluNet (https://www.who.int/influenza/gisrs_laboratory/flunet/en/, accessed on Dec 1, 2018), maintained by the World Health Organization;
2. influenza laboratory confirmations for Hong Kong from the website of the Center for Health Protection (<https://www.chp.gov.hk/en/resources/29/304.html>, accessed on Dec 1, 2018),
3. the mean intensities of two influenza B lineages in China from Yang et al. (2018b); and
4. weekly outpatient illness and viral surveillance data, seasonal vaccine effectiveness and the index of adjusted overall VE in the United States from January 2011 to April 2018 from FluView (<http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>, accessed on Dec 1, 2018), where the vaccination coverage among general population is the highest among all countries.

2.2.2 Methods

Cross-wavelet analysis is a tool for the frequency spectra comparison between two time series and revealing the synchronicity in certain frequency-time domain between two times series (See Veleda et al. (2012); Rösch and Schmidbauer (2018)).

To explore the anti-phase pattern of A/H1N1 and A/H3N2 in Hong Kong and countries in NTZ, we selected the countries/regions with more than cumulative 2000 confirmed cases for both A/H1N1 and A/H3N2. Based on the results of spectrum, we chose a suitable period to gain the decomposition then reconstruct the time series to study the temporal trends in the remained countries and regions (See Rösch and Schmidbauer (2018)).

For global pattern of influenza activity, we calculated the ratio of confirmations of influenza B over A, after adding a constant (*e.g.*, 25, note that our results are not sensitively dependent on this choice) to reduce the effects of small number. For instance, A_i and B_i denote the confirmations of influenza A and influenza B in week i , respectively. A direct comparison is inappropriate between countries due to inconsistent testing efforts and flu surveillance policies. The ratio of B_i/A_i is relatively comparable between countries, since the number of flu A on the denominator (A_i) of the ratio is to adjust the heterogeneity of flu testing and surveillance situation in different places. To reduce the effects of small number, we study the ratio of $(B_i + 25)/(A_i + 25)$ thereafter. This ratio would capture unusual influenza B activities, in particular, when influenza A and B are out of phase. We calculate the weekly ratios of influenza B over influenza A confirmations for 159 countries and regions. We calculate Pearson's correlation on the ratios between countries and regions. We use cross-wavelet to analyze the time series of the ratio we defined and the intensity of influenza B Victoria lineage in China.

2.3 Results

The temporal patterns of the weekly reported cases of A/H1N1 and A/H3N2 from Hong Kong and 71 of 150 countries in the North Temperate Zone are presented

in figure 2.1(a) and (b). It can be seen that anti-phase patterns formed between A/H1N1 and A/H3N2 after a transient period of several years. Figure 2.1(c) shows the spatio-temporal patterns of the square roots of the annual totals of A/H1N1 and A/H3N2 for Hong Kong and 24 NTZ countries ordered according to their longitude. The square root transformation was used to make the small-scale epidemic more visible graphically. It can be seen that individual countries/regions exhibited the same anti-phase patterns.

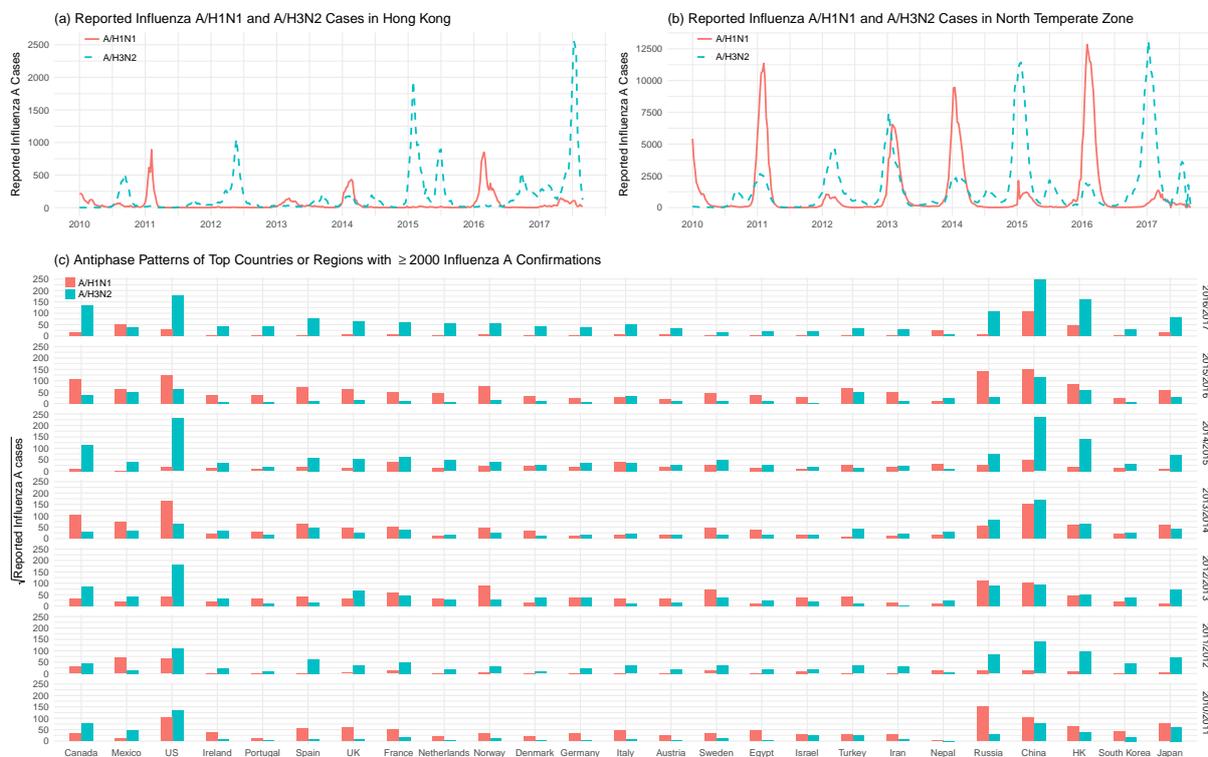


Figure 2.1: Temporal patterns of A/H1N1 and A/H3N2 in (a) Hong Kong, and (b) the North Temperate Zone. (c) Spatial patterns of the annual totals of A/H1N1 and A/H3N2 in Hong Kong and 24 countries and regions in the North Temperate Zone, which are ranked according to their longitudes. It is evident that A/H1N1 and A/H3N2 alternately dominated in the recent 4 years. A/H1N1 skipped (*i.e.*, very low numbers of cases) in most countries or regions in 2011/2012 and 2016/2017. Since A/H1N1 resurged in 2012/2013, a resurgence could happen in 2017/2018.

Time-series decomposition analyses and wavelet analyses were performed on the reported cases of A/H1N1 and A/H3N2 from January 2010 to July 2017 for both Hong Kong and the NTZ (see Appendix). Figure 2.2(a) and (b) shows the reconstructed time series of A/H1N1 and A/H3N2 for both Hong Kong and the NTZ. Again, anti-phase synchronization is observed in these time series.

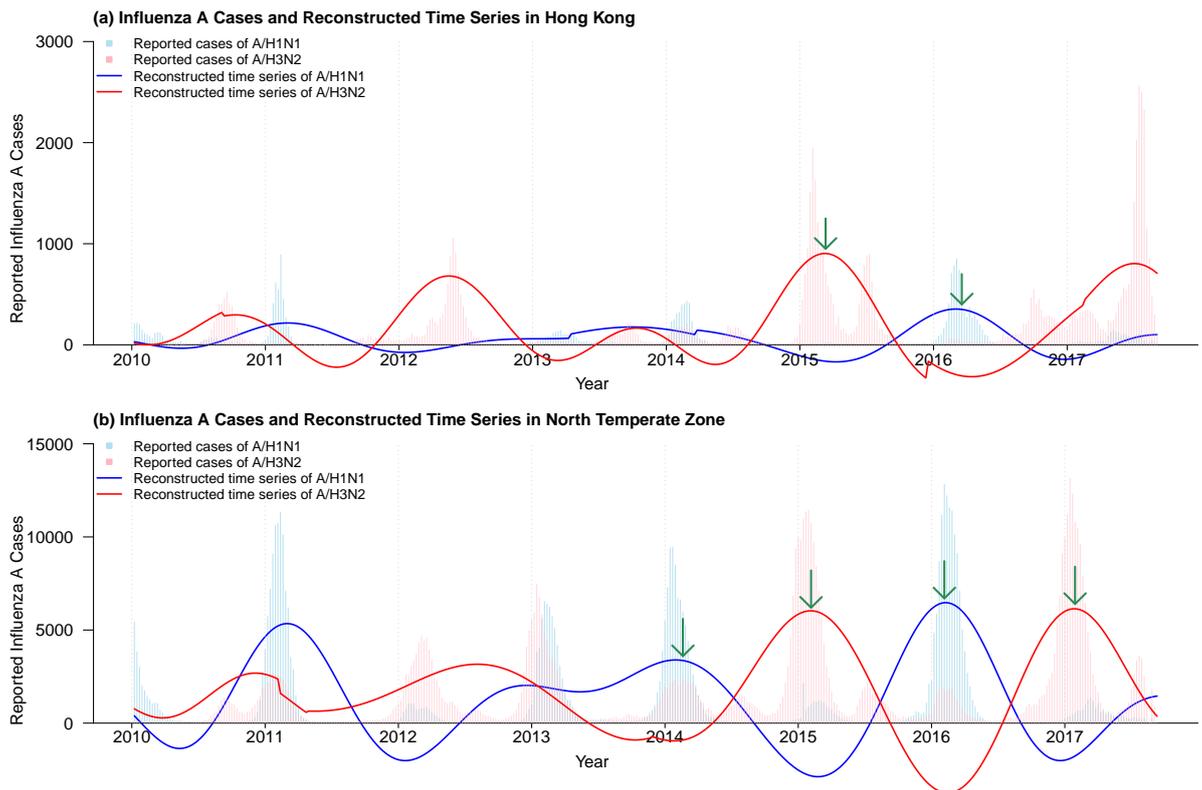


Figure 2.2: Temporal patterns of reported influenza A/H1N1 and A/H3N2 cases and their reconstructed time-series in (a) Hong Kong, and (b) the North Temperate Zone. Arrows indicate the remarkable anti-phase patterns from 2015 to 2017 in Hong Kong, and from 2014 to 2017 in the North Temperate Zone.

Figure 2.3 shows the ratios in China mainland and Hong Kong in panel (a), and the intensities of two influenza B lineages in mainland China in panel (b). The ratios in mainland China and HK show strong correlation ($r = 0.5813$, p-

value $< 1e^{-15}$). The ratio in mainland China and the intensity of influenza B Victoria lineage have a significant correlation ($r = 0.4537$, p-value $< 1e^{-15}$). The correlation between the ratio and Yamagata Lineage activities is not significant ($r = 0.07786$, p-value = 0.1582).

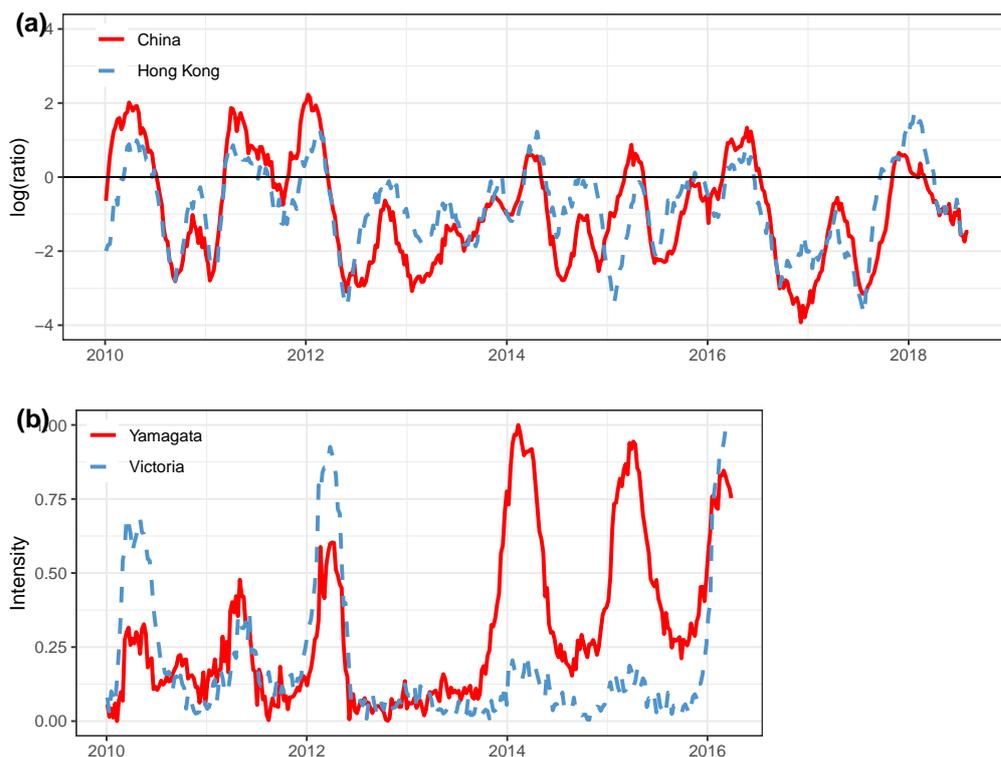


Figure 2.3: The logarithmic ratios of influenza B over influenza A in mainland China and Hong Kong (a) and the intensities of influenza B lineages in mainland China (b).

By adopting the cross wavelet analysis (See Tang et al. (2018)), we find the association between B/A ratio and Victoria lineage intensity is significant, and dominant with the period of one year. The phase dynamics (with wavelet period of one year) indicates the association is positive and stationary over the study period. We show the wavelet analysis in Figure 2.4.

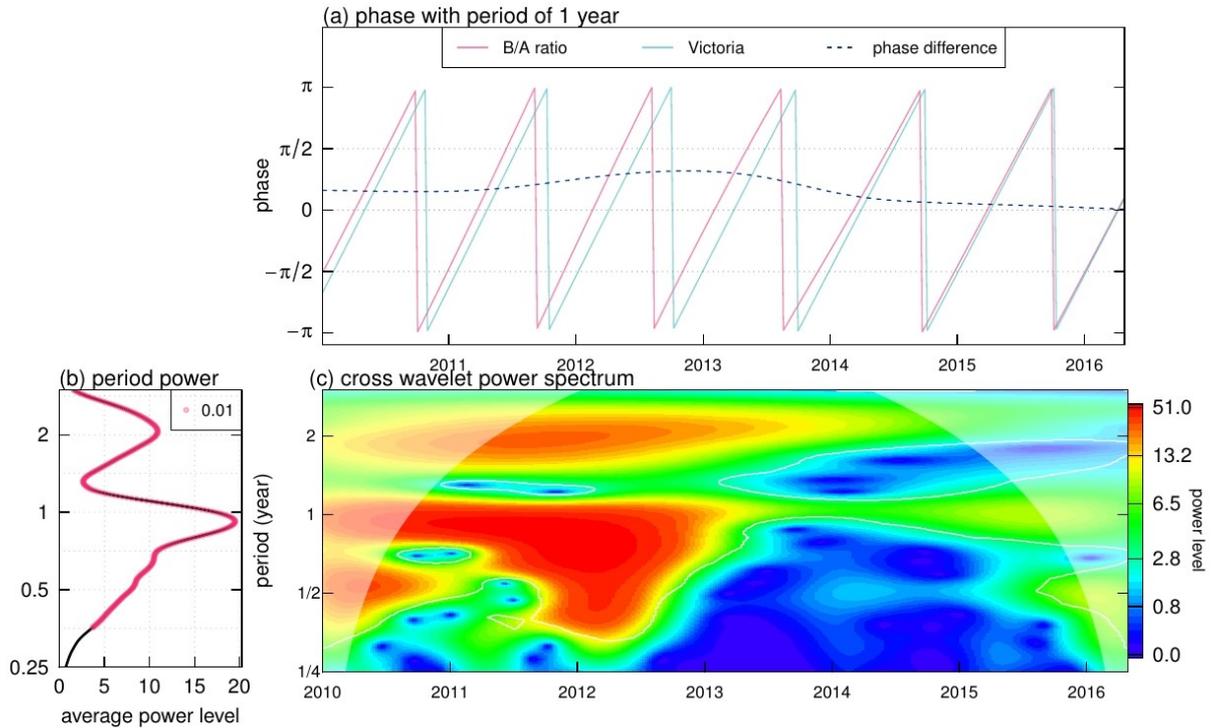


Figure 2.4: Cross wavelet analysis of the ratio with Victoria lineage intensity in China. (a) phase of the ratio and Victoria lineage intensity; (b) period power spectrum; (c) cross wavelet power spectrum.

We calculate the ratios for 159 countries and regions. However, data in many of them are low which lead to a ratio around 1. To resolve this issue, we focus on those where the maximum of the ratio is larger than 3, which means that reported influenza B cases are more than 50 in at least one week. Thus we end up a list of 34 countries and regions, which covers most of the Group of Twenty (G20). We calculate the Pearson's correlation of the ratios between the 34 countries and regions. Figure 2.5 shows the correlations between the ratios of different places with United Kingdom (UK) versus those correlations with Hong Kong. The UK and HK are used as reference, since we find that the correlation between the UK and HK is close to zero ($r = 0.000796$). From figure 2.5, we notice that UK,

the US and Canada are in one cluster and share similar patterns, while HK and mainland China are in the other. Indeed, the unusually severe influenza B epidemic occurred in HK in 2018 but not in the Northern America. The US and Canada have a higher influenza vaccination coverage among the general population than other places, which could partly explain the different patterns between Northern America and other regions (He et al 2015). Figure 3 provides a new metric to reveal the global patterns of influenza activity.

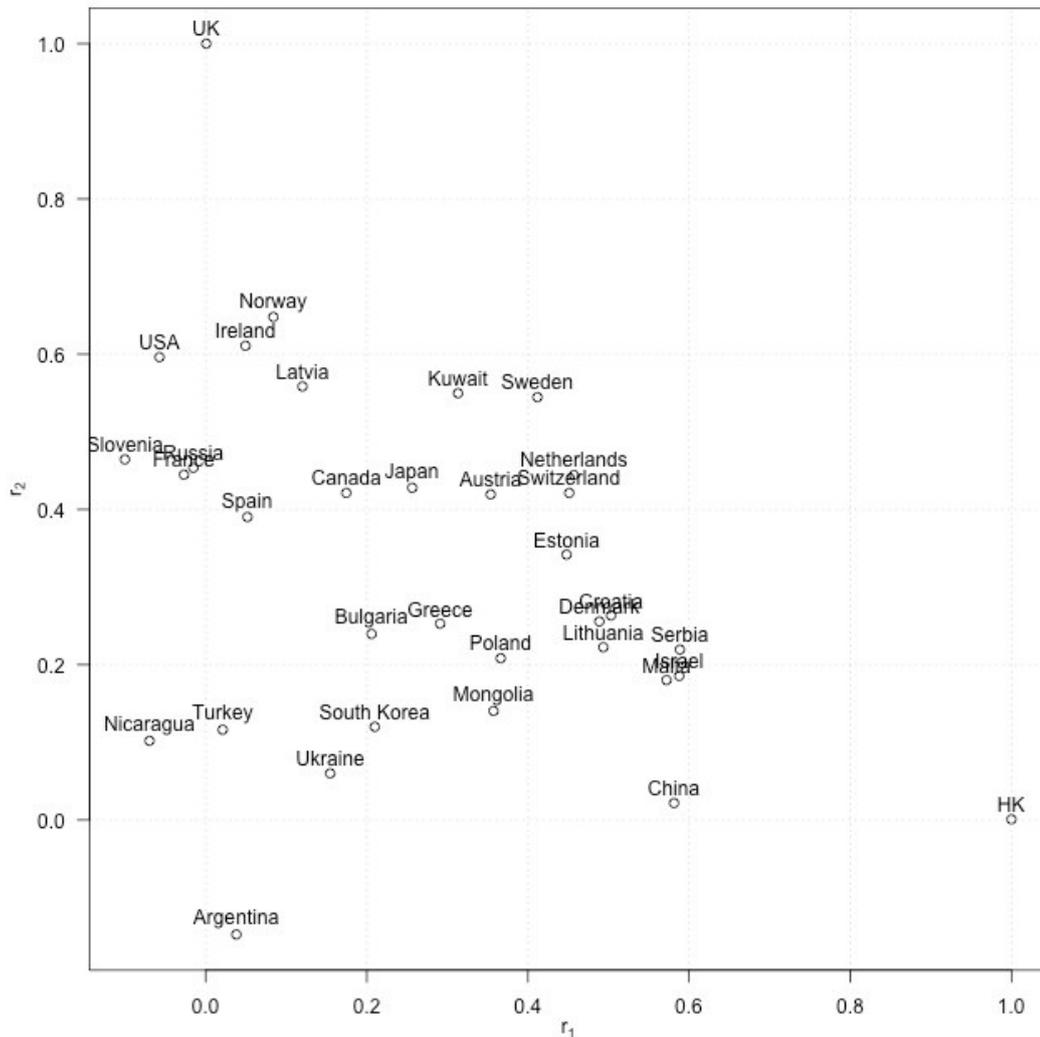


Figure 2.5: Pearson's correlations of the ratios between countries or regions. The horizontal axis is the correlation between different countries/regions and Hong Kong. The vertical axis is the correlation between different countries/regions and the United Kingdom (UK).

Figure 2.6 shows the spatial-temporal patterns of the weekly ratios of influenza B over A in 54 states or regions in the US from 2010/11 to 2017/18. Annual "peaks and valleys" are observed in most of regions or states. However, spikes particularly stood out in 2015/2016 season, when abnormally high B/A ratio occurred in Ohio,

Michigan, Utah, Colorado, Illinois, etc.

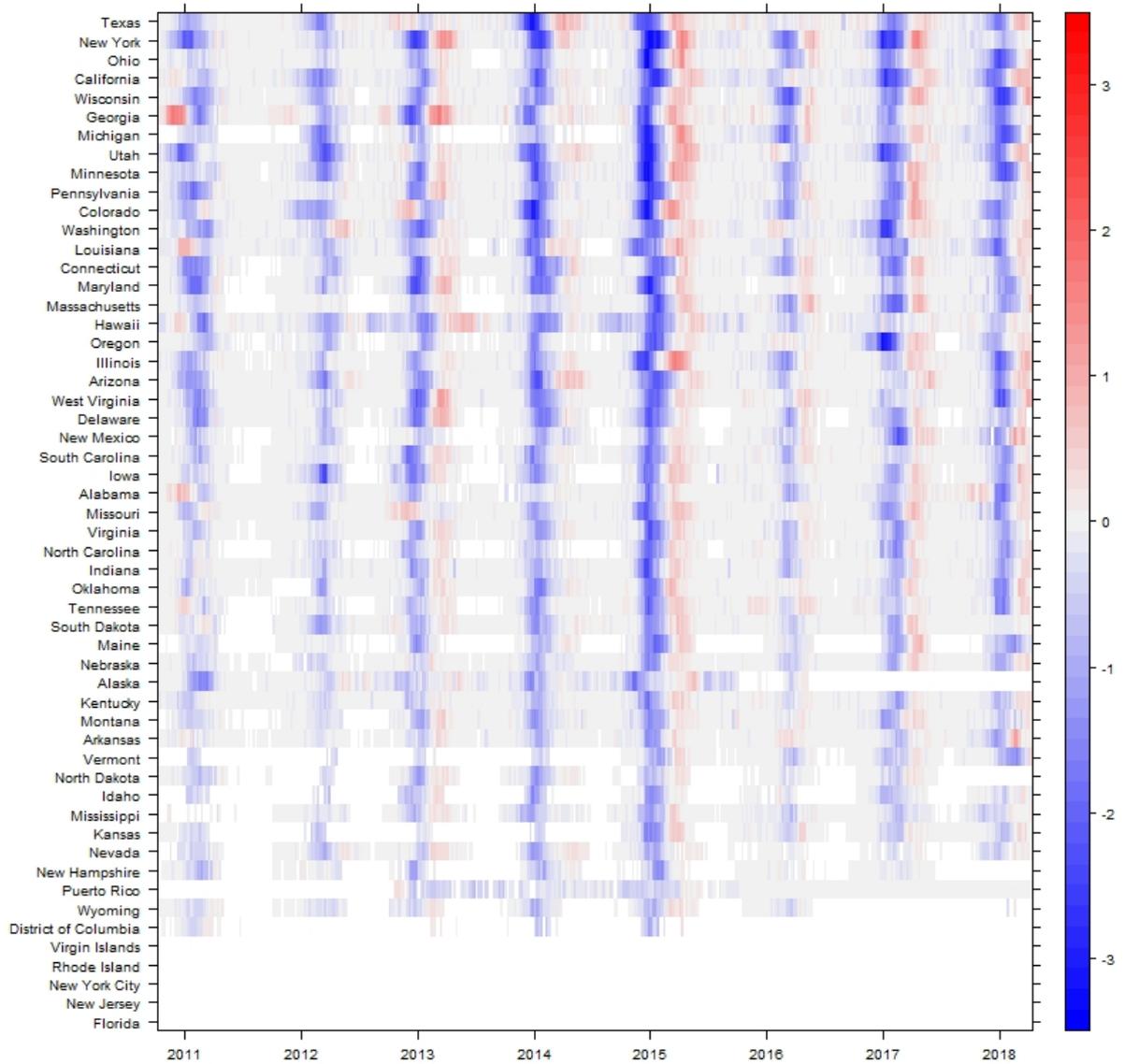


Figure 2.6: The ratio of influenza B over influenza A in 54 states/regions in the US from 2010/11 to 2017/18. The states/regions are listed in descending order according to the number of total influenza A confirmations from top to bottom. The color key represents the logarithmic ratios.

Figure 2.7 shows the adjusted overall VE is presented in panel (a) and the

weekly confirmations of influenza A and B, and the weekly ratio is shown in panel (b). In 2014/15, the epidemic onset and peak of influenza A arrived slightly (but notably) earlier than other years, followed by a moderate outbreak of influenza B. Unlike other years, the A and B epidemics staggered in 2015, which led to a dramatic increase of weekly B/A ratio. Surprisingly, the influenza vaccine attained the lowest VE of the study period in 2015.

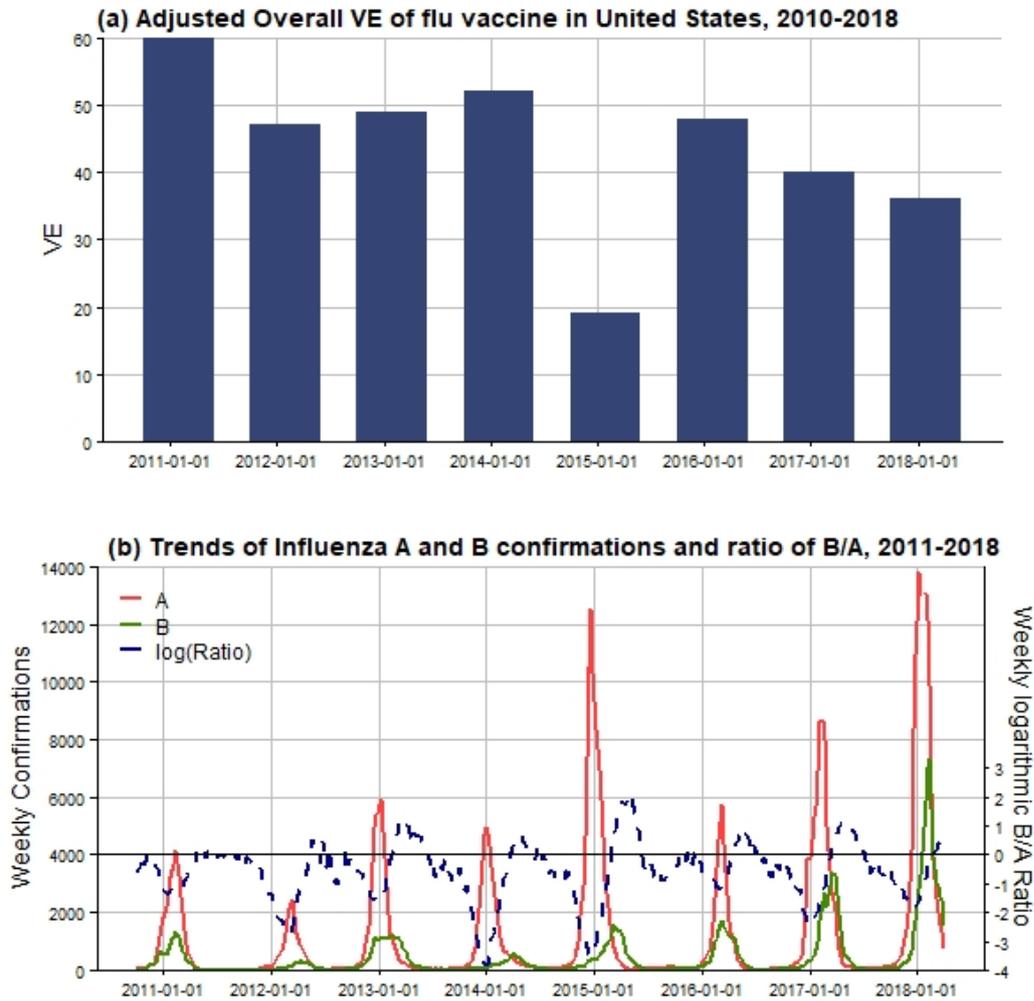


Figure 2.7: The adjusted overall VE (a) and weekly confirmations and ratios (a) in the United States, 2010/11-2017/18. Bars in (a) show the estimated adjusted overall VE. Black curve with dots in (b) shows weekly confirmations of Influenza A, while red curve with circles in (b) shows influenza B and the green curve shows the weekly ratio of influenza B over influenza A confirmations. The low VE in 2014/15 season coincided with the high peak of B/A ratio. The influenza B epidemic was severe in 2017-18 season, but the B/A was not high.

2.4 Discussion and conclusions

The results of the present study appear to be novel in identifying anti-phase synchronization patterns between influenza A/H1N1 and A/H3N2 in Hong Kong and the NTZ for the years 2014 to 2017. Influenza epidemics in Hong Kong are largely in synchrony with those in the NTZ. The study results are in line with those of a previous study by Finkelman et al. (2007), who identified interhemispheric synchrony for A/H3N2. The present exploratory analyses suggest that these patterns could be due to the following factors: (1) the two subtypes competing on susceptibles (See He et al. (2015)); (2) partial cross-immunity between the two subtypes (See Tamerius et al. (2011)); (3) viral evolution (See Lemey et al. (2014)); and (4) herd immunity and waning of immunity (See Piedra et al. (2005)). The finding that many countries displayed similar patterns across multiple years suggests that the anti-phase pattern is unlikely to be a stochastic event.

This study is strengthened by the use of data from FluNet, a global tool for influenza surveillance that has recorded influenza data from more than 100 countries over an extended period of time. However, the study has some limitations. FluNet data, while comprehensive, may be subject to some biases, due to country variations in testing policies, or surveillance data coming from hospitals rather than being representative of the general population (See He et al. (2015)).

In this study, the temporal patterns of influenza A/H1N1 and A/H3N2 in Hong Kong and the NTZ were examined. An anti-phase synchronization pattern between A/H1N1 and A/H3N2 was found. If this pattern persists in the next few years in Hong Kong, A/H3N2 will likely dominate in 2018/2019 and 2020/2021. It is therefore recommended that the Hong Kong government initiates public health preparedness against this potential future influenza epidemic. Fur-

thermore, hospitals should make plans to prepare for larger numbers of patients requiring hospitalization than expected for an A/H1N1 epidemic. This anti-phase synchronization also implies that the patterns of A/H1N1 and A/H3N2 epidemics will repeat biennially. Namely, the epidemic in 2018 could be a repeat of that in 2016 (*i.e.*, 2 years ago), or a simple averaging of the epidemics in 2014 and 2016. A simple prediction formula is provided in the Appendix. More rigorous statistical model fitting can be done in the future. In summary, an anti-phase synchronization pattern between A/H1N1 and A/H3N2 was found for both Hong Kong and the NTZ in recent years.

We find a strong correlation between the B/A ratios in mainland China and HK ($r = 0.5813$, $p\text{-value} < 1e^{-15}$), which is not surprising since the two regions have close tie. The reporting practices are different between the two regions, which could contribute to the difference in the ratios. We also notice that the ratio in mainland China showed a significant correlation ($r = 0.4537$, $p\text{-value} < 1e^{-15}$) with the activities of influenza B Victoria lineage between 2010 and 2016. This relationship is unexpected.

We report that the weekly ratio of influenza B over A in both state-level and country level showed a peak in 2015 and the VE is low in 2015 than other years. We argue that these two events could have some associations. Statistically speaking, the probability that these two events occur in the same year (out of eight years) is 1 out of 8. This is equivalent to toss two fair dices (assume eight sides for each) and obtain the same side for both. Although this is not statistically significant ($1/8$), this finding deserves further investigation.

He et al. (2015) argued that the difference in vaccination coverage could possibly cause some large-scale patterns, e.g. a skip-and-resurgence of influenza

A/H1N1 in mainland China, HK and some European countries. The abnormality in weekly ratio of B over A in 2014/2015 season could be related to the competition between influenza A and B. When the VE is low, the competition of being infected by A or B for the susceptible pool becomes strong. Therefore, the time delay between influenza A and B epidemics is enlarged. Thus, low VE (or low vaccination coverage) favors the "antiphase pattern" between influenza A and B. One of other reasons could be the emergence of a novel strain or clade of influenza B, which was very transmissible and virulent in 2015 (See Bedford and Neher (2016)).

In conclusion, the ratio of influenza B over influenza A is a useful metric to detect hidden patterns in the very complicated dynamics of influenza activities.

Chapter 3

Tuberculosis (TB)

3.1 TB in China

Tuberculosis (TB) remains one of the most severe threats to public health worldwide. As one of the most urgent epidemic goals, the World Health Organization (WHO) End TB Strategy aims to reduce the TB incidence worldwide in 2035 to 90% of the 2015 levels and to eradicate TB by 2050 (See WHO (2015)). With the second largest TB burden after India (See WHO (2018a)), China constitutes approximately 0.9 million TB infections every year; therefore, sustaining declines in TB in China is essential to achieve global goals. Despite considerable success in reducing the growth of TB in China, the midterm goal of an annual 10% reduction in TB incidence by 2025 was not fulfilled (See WHO (2015, 2018a)).

Demographic heterogeneity in TB epidemiology is one of the focuses to achieve further reductions in China. Among reports of demographic heterogeneities between age groups, sexes, identities, regions, occupations, races, and ethnicities (See Dodd et al. (2016); Chan et al. (2017); Horton et al. (2018); Wu et al. (2008, 2017); Iqbal et al. (2018)), some researchers have highlighted the “aging” problem and urbanization process (*i.e.*, migration) and noted that these two phenomena

are crucial for future declines in TB incidence (See Harris et al. (2019); Ge et al. (2015)). Therefore, a thorough understanding of these heterogeneities is critical for governmental TB screening and prevention strategies and vaccine implementation priorities and thus sustainable reductions in future TB incidence rates.

Another barrier to achieving the TB reduction goal in China is the prevalence of drug-resistant (DR)- and multi-drug-resistant (MDR)-TB, which is closely associated with the reactivation of TB that mostly occurs among the older population. The WHO defines MDR-TB as TB strains that fail to respond to at least isoniazid (INH) and rifampicin (RFP), which are the two most potent and widely used antimicrobial drugs for curing TB. There are 63,000 estimated new MDR-TB notifications in China, which constitute one-third of cases worldwide (See WHO (2018b)). Mismanagement of TB treatment and transmission of TB are two major reasons for the prevalence of MDR-TB and have led to a continuous increase in MDR-TB notifications in the recent decade in China (See Wu et al. (2017)).

To investigate the heterogeneities in TB incidence in China, we explored TB and DR-TB notifications in Shandong Province, which is located in eastern China and has the second largest population (*i.e.*, approximately one billion) in China. With 13.2% of the senior population (*i.e.*, those over 65 years old) and approximately 40% of the rural population (See National Bureau of Statistics (2011); Shandong Provincial Bureau of Statistics (2018)), Shandong had an estimated 211,900 TB infections in 2010; 66.7% of infections were residents in rural areas (See Technical Guidance Group of the Fifth National TB Epidemiological Survey (2012)). Through meticulous research on TB incidence heterogeneity across seven cities in Shandong, we hope that our results will provide useful insights for public health policy on TB control to continue the decline in TB incidence in China in

the future.

3.2 Data and methods

3.2.1 Data

We included two sources of TB notifications in this study, namely, TB notifications from 2006–2017 from seven cities in Shandong Province, China, and some DR-TB diagnoses from 2004–2017 across Shandong. Both types of data contained detailed personal information (*e.g.*, sex, age, birthday, diagnosis date). First TB diagnosis notifications were from the Tuberculosis Surveillance System, which included the seven cities in Shandong, Jinan, Yantai, Weifang, Jining, Linyi, Dezhou, and Liaocheng. Additionally, DR-TB laboratory-confirmed records for four commonly used drugs, isoniazid (INH), rifampicin (RFP), ethambutol (EMB), and streptomycin (SM), were from 38 TB monitoring units, including two province-level hospitals, 21 county-level hospitals, and 13 municipal-level local health departments, out of 142 such units across Shandong. Population data were taken from the Statistics Yearbook (See Shandong Provincial Bureau of Statistics (2018)). and The Sixth National Population Census of 2010 (See National Bureau of Statistics (2011)).

Age, birth cohort, and calendar period of diagnosis are three risk factors and time scales that are important in epidemiological risk analyses. However, the sole impacts of these three time scales are unidentifiable due to the linear relation among them, *i.e.*, $\text{age} + \text{cohort} = \text{period}$ (See Yang and Land (2013)). To overcome this difficulty, Kuang et al. (2008) proposed a canonical parameterization approach to map the original variable space into a second-order difference of the original variables space and a linear plane determined by three points of the origi-

nal variable space (to reduce 4 degrees of freedom). In addition to chronic diseases, age-period-cohort (APC) models are also applicable to infectious diseases such as tuberculosis and AIDS (See Wu et al. (2008)), given that these infectious diseases share a long-run infection process.

Throughout this study, we use the following notations and classifications: TB notification rate (*i.e.*, the number of notifications per 100 thousand population), DR-TB tested notification rate (*i.e.*, the number of notifications per 100 tested notifications), drug-sensitive (*i.e.*, sensitive to INH, RFP, EMB, and SM), INH-resistant, and RFP-resistant.

3.2.2 Statistical analysis and models

We fit APC models to the TB notifications and DR-TB notifications separately and estimate the individual effect of age, birth cohort and period on TB notification rate and DR-TB tested notification rate. We conduct the Poisson bootstrap simulation method 100 times to explore model sensitivity (See Efron and Tibshirani (1986)). We predict the future populations for both sexes by multiplying the average change rate all these years and the future number of DR-TB records by APC models. By applying the vector autoregression (VAR) method (See Sims (1980)) to APC models, a 10-year and a 6-year forecast are constructed for the TB notification rate and DR-TB tested notification rate, respectively. By assuming three reporting ratios (*i.e.*, 50%, 75%, and 100%) (See People's Government of Shandong Province (2017)), which indicate the ratio of the number of reported DR-TB laboratory-confirmed results over the total number of TB notifications, and a constant ratio (0.4) (See National Bureau of Statistics (2011)) of the population of the seven cities over that in Shandong Province for all the years, we

estimate and project the DR-TB cases among populations across all of Shandong Province.

All analyses, data visualizations, and models are performed and constructed in R (version 3.4.3). In particular, the R package “`apc`” is implemented for fitting APC models and forecasting future trends (See Nielsen and B (2015)).

3.3 Heterogeneities in TB

Without a loss of generality, we present the results of Jinan, Yantai, Linyi, and Liaocheng as representatives in the main text and the remaining three cities in the Appendix.

3.3.1 Heterogeneities in TB incidence between cities.

As shown in Table 3.1., we identified 22,756 notifications in Jinan, 28,107 in Yantai, 63,089 in Linyi, and 35,376 in Liaocheng. Except for Yantai, which was missing residential information, approximately 90% of TB notifications in the other three cities were from local residents. Young adults (18–35 years old) and seniors (over 60 years old) were the two leading sources for TB notifications among all four cities. Regarding occupation, farmers had the highest TB burden, followed by students, workers and homemakers/unemployed. However, the proportion of farmers varied greatly between Jinan and Yantai (53.48% and 63.55%) and Linyi and Liaocheng (91.91% and 86.98%), as the former two cities had higher urbanization rates (70.53% and 63.66%) (See Shandong Provincial Bureau of Statistics (2018)).

Table 3.1: Summary of TB notifications across four cities in Shandong.

| | Jinan (N=22 756) | | Yantai (N=28 107) | | Linyi (N=63 089) | | Liaocheng (N=35 376) | |
|------------------------|---------------------|---------------------|----------------------|---------------------|---------------------|---------------------|-------------------------|---------------------|
| | no. | % (95% CI) | no. | % (95% CI) | no. | % (95% CI) | no. | % (95% CI) |
| Gender | | | | | | | | |
| Male | 15 154 | 66.59 (65.98–67.21) | 20 675 | 73.56 (73.04–74.07) | 45 939 | 72.82 (72.47–73.16) | 23 295 | 65.85 (65.36–66.34) |
| Female | 7 602 | 33.41 (32.79–34.02) | 7 432 | 26.44 (25.93–26.96) | 17 150 | 27.18 (26.84–27.53) | 12 081 | 34.15 (33.66–34.64) |
| Identity | | | | | | | | |
| Local | 18 894 | 83.03 (82.54–83.52) | – | – | 59 841 | 94.85 (94.68–95.02) | 33 645 | 95.11 (94.88–95.33) |
| Immigrant | 3 862 | 16.97 (16.48–17.46) | – | – | 3 247 | 5.15 (4.97–5.32) | 1 731 | 4.89 (4.67–5.12) |
| Age groups | | | | | | | | |
| 0–17 | 594 | 2.61 (2.4–2.82) | 715 | 2.54 (2.36–2.73) | 1 186 | 1.88 (1.77–1.99) | 647 | 1.83 (1.69–1.97) |
| 18–35 | 8 021 | 35.25 (34.63–35.87) | 10 540 | 37.5 (36.93–38.07) | 15 922 | 25.24 (24.9–25.58) | 9 890 | 27.96 (27.49–28.42) |
| 36–60 | 7 822 | 34.37 (33.76–34.99) | 10 114 | 35.98 (35.42–36.55) | 23 981 | 38.01 (37.63–38.39) | 13 386 | 37.84 (37.33–38.34) |
| over 60 | 6 319 | 27.77 (27.19–28.35) | 6 738 | 23.97 (23.47–24.47) | 22 000 | 34.87 (34.5–35.24) | 11 453 | 32.38 (31.89–32.86) |
| Occupations | | | | | | | | |
| Students | 2 129 | 9.36 (8.98–9.73) | 1 851 | 6.59 (6.3–6.88) | 1 534 | 2.43 (2.31–2.55) | 1 080 | 3.05 (2.87–3.23) |
| Workers | 1 698 | 7.46 (7.12–7.8) | 2 036 | 7.24 (6.94–7.55) | 733 | 1.16 (1.08–1.25) | 1 153 | 3.26 (3.07–3.44) |
| Migrant workers | 285 | 1.25 (1.11–1.4) | 1 002 | 3.56 (3.35–3.78) | 643 | 1.02 (0.94–1.1) | 238 | 0.67 (0.59–0.76) |
| Farmers | 12 171 | 53.48 (52.84–54.13) | 17 863 | 63.55 (62.99–64.12) | 57 983 | 91.91 (91.69–92.12) | 30 770 | 86.98 (86.63–87.33) |
| Household / Unemployed | 2 635 | 11.58 (11.16–12) | 3 418 | 12.16 (11.78–12.54) | 564 | 0.89 (0.82–0.97) | 627 | 1.77 (1.63–1.91) |
| Others & unknown | 3 838 | 16.87 (16.38–17.35) | 1 937 | 6.89 (6.6–7.19) | 1 632 | 2.59 (2.46–2.71) | 1 508 | 4.26 (4.05–4.47) |

Heterogeneity in TB notification rates was observed between males and females in Shandong (Figure 3.1). The notification rates of males were higher than those of females throughout the 12 study years. However, heterogeneities in gaps between male and female cases was observed between cities. TB notification rates in Yantai and Linyi for both males and females significantly declined from 2006–2017, while those in Jinan and Liaocheng remained relatively steady. Although stable, Jinan experienced a small reduction in the TB notification rate for males and a slight increase for females. TB notifications for both males and females in Liaocheng increased from 2006–2017. The ratios of TB notification rates for males to females in Jinan and Liaocheng were approximately two, while the ratios were approximately three in Yantai and Linyi.

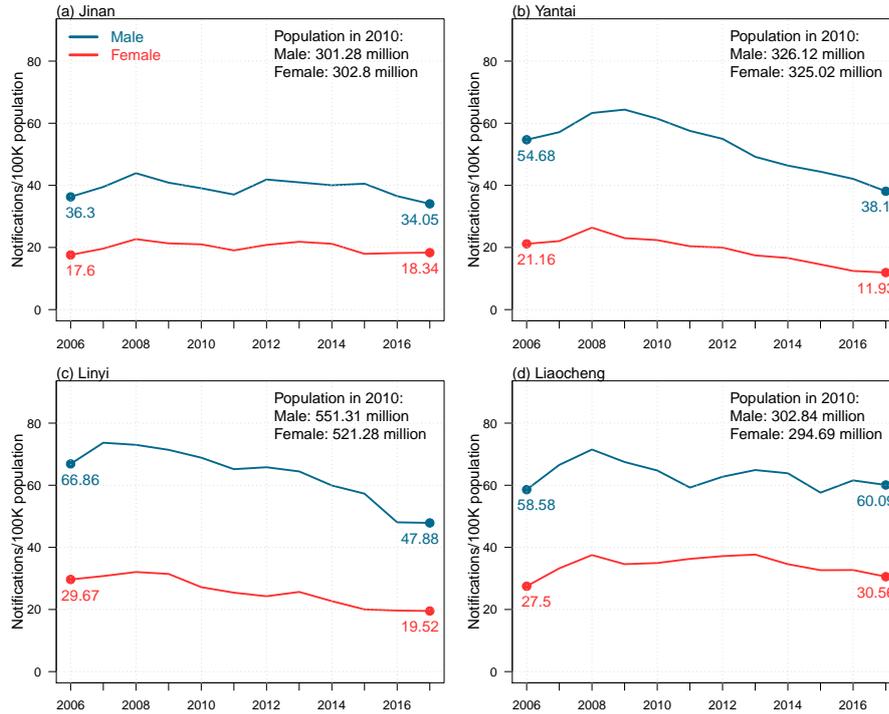


Figure 3.1: Changes in the notification rates for males and females from 2006–2017 in (a) Jinan, (b) Yantai, (c) Linyi, and (d) Liaocheng. In each panel, the blue line and red line represent notification rate trends; the population size for males and females in 2010 is shown in the top-right corner.

In figure 3.2, we present the changes in TB notification density by age between 2006 and 2017 for both sexes across four cities in Shandong. We highlight the density changes for young adults and seniors for a clearer demonstration. Homogeneities were present in 2006 across all four cities. For males, young adults at approximately 18 years old and seniors at approximately 60 years old were the two primary sources for TB notifications. For females, TB notifications from young females occupied a majority of the proportion of cases. However, notification densities by age shifted in 2017, and heterogeneity between cities was observed. For males, notifications from young males mainly remained stable in Jinan and Yantai or decreased slightly in Linyi and Liaocheng from 2006–2017, while those

from seniors increased slightly in all four cities. For females, the proportions in Jinan remained unchanged, while large proportions of TB notifications shifted from young adults to seniors in Yantai, Linyi, and Liaocheng. This drastic “aging” phenomenon presented in TB notifications among females from 2006–2017 implies that the primary source of TB notifications changed from transmitted TB to reactivated TB for females.

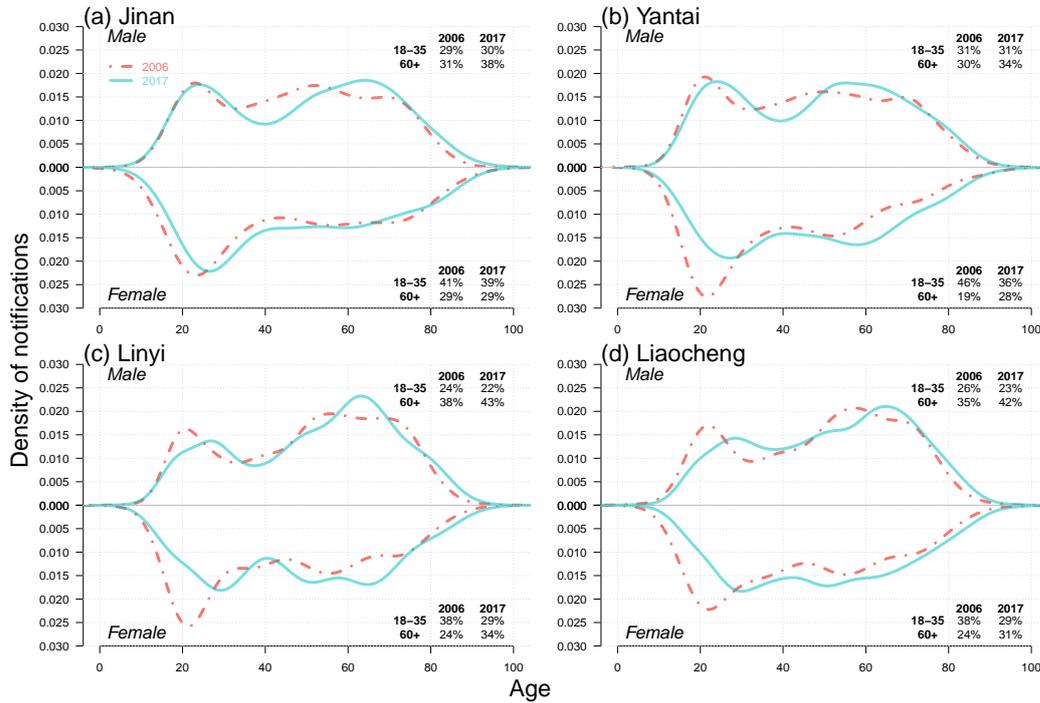


Figure 3.2: Case density changes in 2006 and 2017 for males (upper subpanels) and females (lower subpanels) in (a) Jinan, (b) Yantai, (c) Linyi, and (d) Liaocheng. In each panel, red dashed-and-dotted and green solid lines represent densities in 2006 and 2017, respectively. We show the changes in case proportions for young adults (18–35) and seniors (60+) in 2006 and 2017 in the top-right corner for males and in the bottom-right corner for females.

We present TB notification rates among migrants in Jinan, Linyi, and Liaocheng in figure 3.3, along with migrant population sizes. With a large population of mi-

grants and an increasing trend, TB notification rates in Jinan remained steady and low throughout the studied period. However, dramatic increases in TB notification rates in Linyi and Liaocheng occurred with increasing migrant population sizes. In particular, the TB notification rates in Linyi and Liaocheng reached over 150 cases per 100 thousand population and over approximately 250 cases per 100 thousand population by the end of 2017, respectively. Rapid urbanization rates in 2007–2017 in Linyi (from 38.23%–57.40%) and Liaocheng (from 33.71%–50.34%) (See Shandong Provincial Bureau of Statistics (2018)) may have resulted in TB prevalence among migrants. We further investigated components of migrant TB notifications, shown in figure A.9, and found that most migrant TB cases in Liaocheng were farmers, for both young adults and seniors, and both males and females.

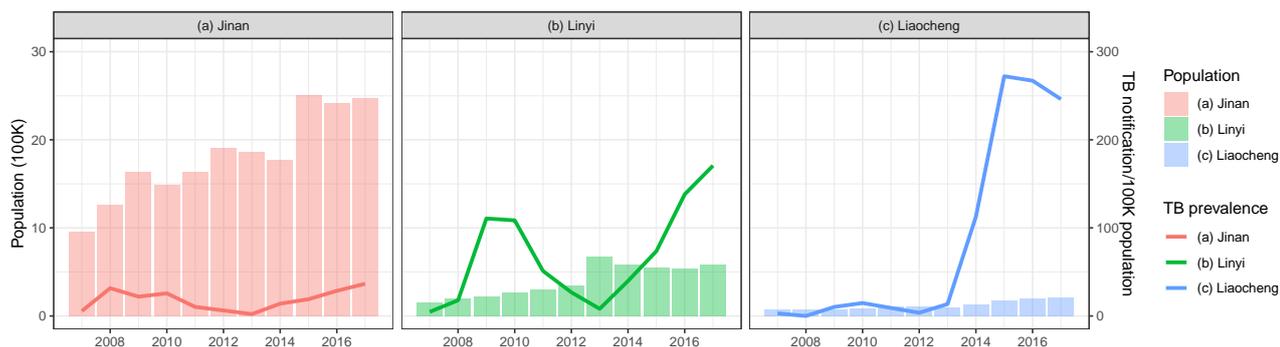


Figure 3.3: TB notification rates among migrants and migrant population sizes from 2007–2017 in (a) Jinan, (b) Linyi, and (c) Liaocheng. In each panel, bars represent a population with a unit of 100 thousand, and lines represent TB notifications per population of 100 thousand.

3.3.2 Heterogeneities in historical and future trends.

In figure 3.4, regarding overall future trends for males, we predict that annual TB notifications per population of 100,000 in 2027 will be approximately 50 in Jinan,

Linyi, and Liaocheng and approximately 30 in Yantai. At the end of 2027, the TB notification rates of senior males will be two times greater than those of young adults in Jinan and Yantai and 3–4 times greater than those of young males in Linyi and Liaocheng. For females, the TB notification rate gaps between young adults and seniors will be negligible in Jinan and Yantai, but seniors' notifications will be approximately 2–3 times greater than those of young adults' notifications in Linyi and Liaocheng. We also observed heterogeneities in the leading age groups between two city categories: in Jinan and Yantai, TB notifications from young adults mainly predominate in early years (*i.e.*, 2006–2023), and those from seniors dominate later on; however, in Linyi and Liaocheng, seniors were and will always be the primary source of TB notifications from 2006–2027. In Liaocheng, unlike in the other three cities, the overall TB notifications per 100 thousand population for males and females remains stable at approximately 50 and 30, respectively, and senior female cases will not decrease in the next ten years.

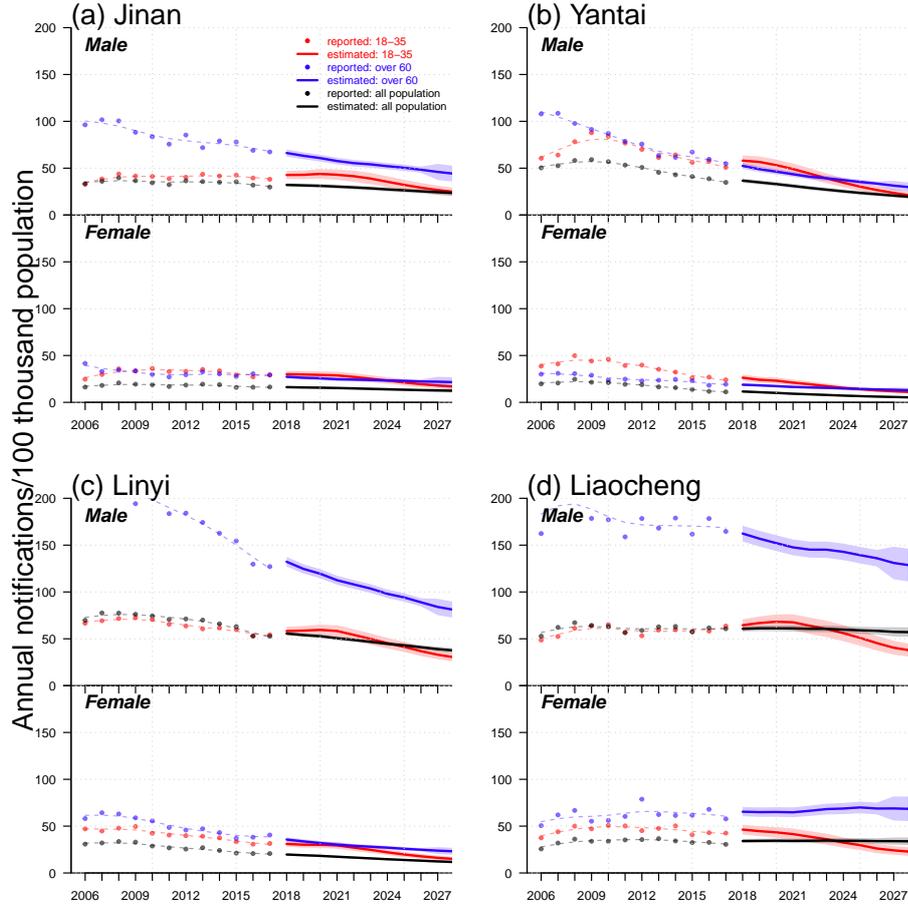


Figure 3.4: Trends of TB notification rates from 2006–2017 and forecasts from 2018–2027 in (a) Jinan, (b) Yantai, (c) Linyi, and (d) Liaocheng. In each panel, (i) upper and low subpanels show trends for males and females, respectively; (ii) dots, lines, and shades colored in red, blue, and black indicate reported trends, forecast trends, and 95% confidential intervals for forecasts of annual TB notifications (in the 100 thousand population) for populations aged 18–35 and over 60 years and the total population, respectively; (iii) red, blue, and black dashed lines are smoothed splines that indicate the trends of annual reported TB notifications for the populations aged 18–35 and over 60 years and the total population, respectively.

3.3.3 Current situation and forecast trends for DR-TB.

We included 12,984 DR-TB laboratory records with complete drug sensitivity information in this study (Table 3.2). A total of 22.27% (21.5%–23.05%) of cases

presented resistance to at least one of the four commonly used drugs among first-diagnosed patients and 30.59% (28.56%–32.63%) of cases among those with previous treatment presented resistance. For all four drugs, the proportion of drug-resistant cases was generally higher among patients with a treatment history than among first-diagnosed patients. Among first-diagnoses, 6.47% (6.01%–6.93%) were MDR-TB, while among cases with previous treatment, 13.8% were MDR-TB. Regarding RFP-resistant, EMB-resistant and MDR cases, the proportions among previously treated cases significantly exceeded those among first-diagnosed cases. However, on the basis of drug-resistant notifications, 80.3% DR-TB cases and 72.4% MDR-TB cases were from first-diagnosed patients, which is consistent with previous studies in Shandong He et al. (2016, 2017); Tao et al. (2017).

Table 3.2: Summary of drug sensitivity and resistance to four antituberculosis drugs in Shandong.

| Sensitivity or Resistance | First-diagnosed TB notifications (N=11 013) | | Previously Treated TB notifications (N=1971) | |
|-------------------------------|---|---------------------|--|---------------------|
| | no. | % (95% CI) | no. | % (95% CI) |
| Sensitivity to all four drugs | 8560 | 77.72 (76.95–78.50) | 1368 | 69.40 (67.37–71.44) |
| Resistance to four drugs | | | | |
| At least one of four drugs | 2453 | 22.27 (21.5–23.05) | 603 | 30.59 (28.56–32.63) |
| INH | 1628 | 14.78 (14.12–15.45) | 450 | 22.83 (20.98–24.68) |
| RFP | 837 | 7.60 (7.11–8.10) | 326 | 16.54 (14.90–18.18) |
| EMB | 365 | 3.31 (2.98–3.65) | 124 | 6.29 (5.22–7.36) |
| SM | 1808 | 16.41 (15.73–17.11) | 438 | 22.22 (20.39–24.06) |
| INH or RFP (but not both) | 1039 | 9.43 (8.89–9.98) | 232 | 11.77 (10.35–13.19) |
| MDR | 713 | 6.47 (6.01–6.93) | 272 | 13.80 (12.28–15.32) |

In figure 3.5, by assuming three different reporting ratios (*i.e.*, 50%, 75%, and 100%), we describe three possible projections for INH-resistant, RFP-resistant and MDR cases, respectively, from 2018–2023. The historical trend of INH-resistant cases experienced slight growth from 2006–2011 and a sudden drop in 2014–2015 before another small increase. We predict a gradual decrease from 2018–2023, and

there will be approximately 2–4 INH-resistant cases per 100 thousand population in Shandong by the end of 2023. For both RFP-resistant and MDR cases, a relatively sharp increase emerged from 2006–2009, followed by a long-term plateau from 2010–2017, where DR-TB cases per 100 thousand population remained between 1–3 annually. However, projections show strong increases after 2018, with RFP-resistant cases per 100 thousand population at approximately 2–4, and MDR cases per 100 thousand population at approximately 1.8–3.5 by the end of 2023. Figure 5 suggests that before 2016–2017, INH-resistant cases were the primary source of MDR-TB cases, but RFP-resistant cases are predicted to show substantial increases in the future and become the dominant component of MDR-TB cases.

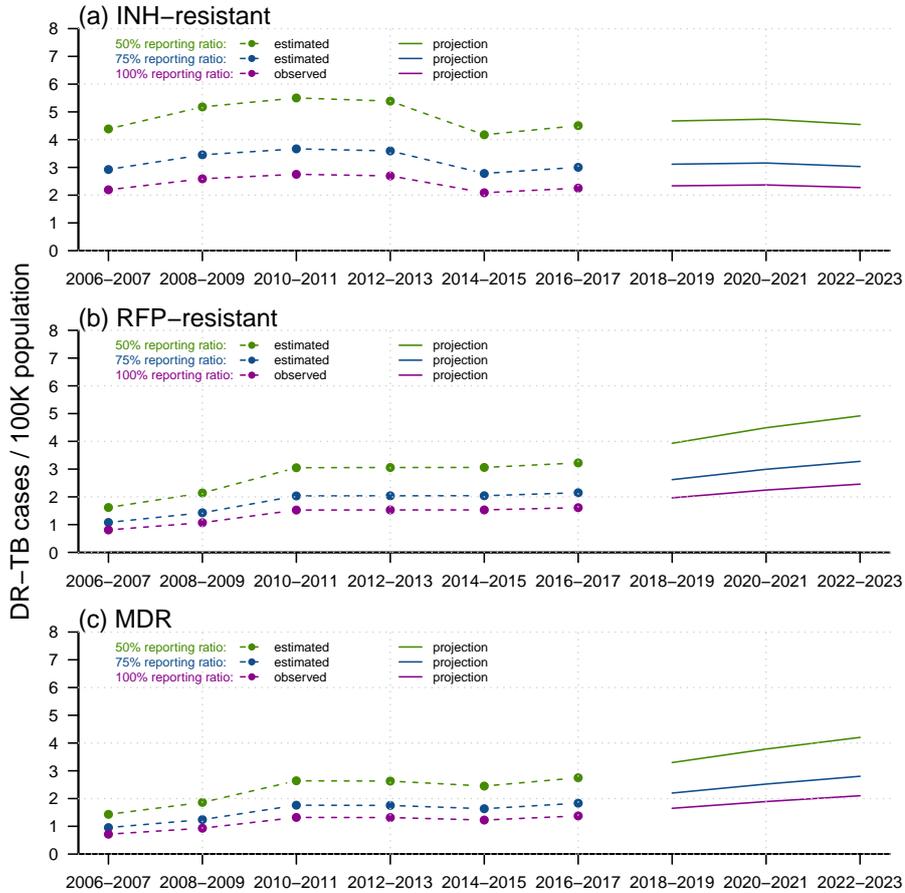


Figure 3.5: Reported trends from 2006–2017 and forecasts from 2018–2023 of DR-TB infections for (a) INH-resistant TB, (b) RFP-resistant TB, and (c) MDR-TB, among the total population in Shandong, China. Within each panel, dotted-and-dashed lines and solid lines colored in green/blue/purple indicate reported DR-TB infections and forecast trends, under assumed test ratios of 50%/75%/100%, respectively. We constructed APC models to predict future total DR-TB notifications and smoothed splines to predict the future population in Shandong.

3.4 Conclusion and discussion

Tuberculosis has been one of the most onerous burdens on public health in Shandong, China. Under the guidance of the “Twelfth Five-Year Plan” for TB control proposed in 2012, province- and city-level governments have improved surveillance,

screening and control strategies with remarkable results (See Center of Autituberculosis in Shandong (2018)). Jinan, Yantai, Linyi, Liaocheng, Weifang, Jining, and Dezhou ranked 1st, 2nd, and 5th–9th in population out of 17 cities in the province, in which Jinan is the capital city, in 2010 (See National Bureau of Statistics (2011)). By studying the heterogeneity in TB incidence rates in these cities, our results not only revealed heterogeneities between cities, sex, and age groups for TB and DR-TB notifications but also provided crude forecasts, insights for the affected population, and TB and DR-TB control and prevention strategies. A large number of immigrants and the aging population are two serious barriers to the further prevention of and reduction in TB incidence (See Chan et al. (2017); Arregui et al. (2018); Wang et al. (2016); Sun et al. (2016); Yang et al. (2018a)). Therefore, our research has shown the underlying relationship between these two factors and the heterogeneity in TB incidence factors.

With significant proportions of rural populations, Jining (42.9%), Linyi (42.6%), Dezhou (44.43%) and Liaocheng (49.66%) had the most TB notifications among the seven cities, whereas Jinan (29.47%), Yantai (36.34%) and Weifang (40.05%) had the least (See Shandong Provincial Bureau of Statistics (2018)). The rapid progression of urbanization provides better medical facilities in these regions, while simultaneously leading to a significant increase in the migrant population from rural areas to urban areas. Due to a lack of proper screening and prevention measures, dramatic increases emerged in TB notification rates among migrants; these lack of measures will make further reductions in TB incidence rates in cities such as Linyi and Liaocheng much more difficult.

Compared to transmitted TB, which usually occurs among young adults, reactivated TB, which is probably MDR-TB, usually occurs among the senior pop-

ulation and requires much more effort for detection, prevention, and treatment. Along with the “aging” population, drastic shifts from transmitted TB to reactivated TB among females from 2006–2017 is undoubtedly responsible for sustaining and speeding up the decline in TB incidence in China. Our forecasts for both TB notifications and DR-TB notifications via APC models further validate this concern.

Due to urbanization across China, many migrant farmers migrated to the cities and thus became the primary source of TB infections, consistent with previous studies in China (See Xu et al. (2005)). Screening and prevention are priorities for migrant farmers and can help reduce the TB notification rates among migrants, further reducing the spread of TB in cities. In addition, measures including prevention education, regular surveillance and diagnostic awareness improvement can be useful as well. College students are the second primary source of TB infections (See Zhang et al. (2010)) in Shandong, especially in Jinan and Yantai; this occurs because there are more colleges and universities in Jinan (46) and Yantai (17) than in Weifang (9), Jining (9), Linyi (6), Dezhou (5), and Liaocheng (4). A high population density on campus and frequent interactions between students increase the prevalence of transmitted TB among young adults. TB control measures and treatment programs in China (See WHO (2018b)) have effectively reduced the number of transmitted TB infections among young adults.

The TB notification rates differed notably between males and females from 2006–2017. In addition to differential behaviors, such as contact patterns and diagnosis delays (See Dodd et al. (2016); Xu et al. (2005)), urbanization and aging, which prolongs life expectancy and increases the risk of TB reactivation, may also play an important role. Consistent with previous studies that suggested a

focus on the population aged over 65 years for vaccine implementations (See Harris et al. (2019)), we further demonstrated that senior females are the top priority to continue the decline in TB incidence in China.

In general, DR-TB notifications in Shandong have experienced a gradual increase over the past decade and are estimated to grow continuously in the coming six years. We propose that RFP-resistant notifications will replace INH-resistant notifications to become the primary source of MDR-TB notifications. Technical improvements in DR-TB detection, increasing reporting ratios and extended transmission of DR-TB strains Sharma et al. (2017); Kendall et al. (2015) may partially explain this change. Furthermore, we suggest that increasing proportions of re-activated TB among senior females and among the aging population should be a primary concern. In addition to improving surveillance systems and treatment techniques, drug choice for the treatment of drug-resistant TB is another crucial factor. Because EMB-resistant TB has maintained a low correlation with both INH and RFP, which makes EMB a better choice than SM for use in treating MDR-infected patients (figure A.17).

There were several limitations in our study: (i) population data and DR-TB data were far from sufficient to elucidate complete and precise inferences and predictions; (ii) age and period effects were not entirely differentiated in our 15-year TB notification data; (iii) the forecasting of future trends relies heavily on the population prediction for TB notification and that of future DR-TB records; therefore, different predicting methods may lead to diverse future trends; and (iv) the rate of DR-TB notifications can be affected by new testing technologies, such as Gene Xpert, which our model did not consider.

In conclusion, our results highlight the heterogeneities among sex, age groups,

regions, and occupations for TB notifications, as well as a continuing increase in MDR-TB notifications in the future. Based on criteria such as rural population size, occupation, and sex, we suggest varied focuses for the future control of TB: (i) raising awareness among college students in urban cities, such as Jinan, while improving medical facilities in rural regions, such as Linyi and Liaocheng; (ii) regular surveillance of the senior population, especially senior females; and (iii) strengthening detection techniques for reactivated TB over the next decade.

3.5 Future works

APC models estimate the relative risks for different age group, periods, and birth cohorts, which have close relation with “force of infection” in mathematical models. Through estimated risk ratios from APC models, we can easily construct age-structured compartmental models with deterministic force of infection and time-dependent reactivated rates to further investigate the epidemic dynamics of TB between different age groups all these years. Via this framework, we can gain better understanding in the heterogeneities between men and women, cities, in terms of concerned parameters, such as reactivated rates.

Chapter 4

Middle East Respiratory Syndrome (MERS)

4.1 Introduction on MERS-CoV

Since the first human case of Middle East respiratory syndrome coronavirus (MERS-CoV) infection, identified in Jeddah, Saudi Arabia in September 2012, 2,040 laboratory-confirmed cases, including 712 deaths, have been reported in 27 countries as of July 12, 2017 (See WHO (2016)). The virus circulated in the Arabian Peninsula, particularly in Saudi Arabia. In May 2015, the Republic of Korea reported the largest outbreak outside the Middle East region.

Many external factors could possibly impact the transmission of MERS-CoV among camels and human. In particular, Hajj and Umrah are two annual religious festivals that attract more than 10 million pilgrims from 184 countries to Saudi Arabia. Hajj is considered a major pilgrimage and involves mass gathering that lasts for six days during the 12th month of the Islamic calendar each year, while Umrah, a minor pilgrimage which could be highly individualized, is made mostly during Ramadan, a period of 29 to 30 days during the 9th month of the Islamic calendar (See Memish et al. (2014)). Several studies have demonstrated

that these mass gatherings, which attracted pilgrims worldwide, were potential for transmission of infectious diseases (See Khan et al. (2013)). Lessler et al. (2014) estimated the potential incidence of MERS-CoV among pilgrims attending Hajj in Saudi Arabia in 2014 to be 6.2, 11.7 and 47.6 pilgrims under the expected, high and very high incidence scenarios, respectively. Furthermore, Poletto et al. (2014) suggested that population movement and human mixing, which were common in Hajj and Umrah, played an important role in MERS-CoV transmission. Besides population movement and human mixing during religion events, camels racing, close and re-opening of camel market, as well as climatic factors could impact of the transmission of MERS-CoV among camels, spill-over from camels to human and among human.

Reproduction number (\mathcal{R}), defined as the average number of cases generated by a typical case, quantifies the intensity of transmission (See Diekmann and Heesterbeek (2000)). Past research showed substantial variations on the estimate of reproduction number on MERS-CoV epidemic, depending on a number of factors: the specific location (i.e. Middle East region or Republic of Korea), the time period considered, mode of transmission, and the choice of mathematical modeling techniques (See Poletto et al. (2014); Kucharski and Althaus (2015); Hsieh (2015); Chowell et al. (2014); Poletto et al. (2016); Breban et al. (2013)).

Chowell et al. (2014) developed a dynamic transmission model that distinguished between the different routes of MERS-CoV transmission. They found that zoonotic transmission was much more effective than human-to-human transmission, with their \mathcal{R}_0 being 0.84 and 0.36, respectively. Breban et al. (2013) used Bayesian analysis to estimate the pandemic risk for MERS-CoV. \mathcal{R}_0 generated for human-to-human transmission under the most pessimistic scenario and

the most optimistic scenario were 0.69 and 0.60 respectively. Their study also highlighted the importance of reducing the rate of zoonotic introductions to the human populations. Poletto et al. (2014) used a novel maximum likelihood approach to jointly estimate \mathcal{R}_0 and the rate of sporadic introduction of MERS-CoV in the Middle East region. They found that \mathcal{R}_0 was 0.50 and the rate of sporadic introduction was 0.28 case per day. More recently, Poletto et al. (2016) used a stochastic modeling approach combined with analysis of imported cases out of the Middle East region. Their model, which assumed partial information on transmission scenario, was a much better fit than the model that assumed complete information on transmission. Their combined modeling approach was able to show that about 75% of the MERS-CoV cases were human-to-human transmission. In other words, primary human cases that are acquired via camels or camel-related products, accounted for 25% of all human cases. Cauchemez et al. (2016) developed a comprehensive statistical framework to analyze the transmission patterns of MERS-CoV among human in Saudi Arabia between January 2013 and July 2014. They found that the proportion of human cases from the reservoir was 12% (95% CI: 9%, 15%), and they also noted that the ratio could be 17% (95% CI: 13%, 20%) if only restricted to passive surveillance cases only. Kucharski and Althaus (2015) modeled on MERS-CoV transmission in South Korea. Their study showed substantial over-dispersion in transmission, which indicated there was a potential for superspreading of MERS-CoV. Superspreading cases are primary cases which infect disproportionately more secondary contacts than other primary cases of the same disease. Hsieh (2015) studied the South Korean MERS-CoV outbreak using a Richard's model and found that the turning point, which is the peak timing of an epidemic, took place at 23 to 24 days after the onset of the index case.

The estimated \mathcal{R}_0 ranged between 7.0 to 19.3. All of these modeling studies on MERS-CoV had primarily focused on human cases.

In a review article by Zumla et al. (2015), it was pointed out that camel-to-human transmission could be due to indirect exposure, e.g. patients might be exposed to MERS-CoV virus via consumption of unpasteurized camel milk, a practice which is not uncommon in Saudi Arabia. In a sero-prevalence study of MERS-CoV by Müller et al. (2015), the percentages of MERS-CoV-antibody positive in the general population, camel shepherds and slaughterhouse workers were found to be 0.15%, 2.3% and 3.6%, respectively. Sero-positive rates exceeding 74%, which indicates previous infections, were found among camels in Saudi Arabia and its neighboring countries. The virus isolation rate was also high among camels, exceeding 7% but they vary across localities and are higher among juvenile camels (See Reusken et al. (2013); Haagmans et al. (2014); Azhar et al. (2014); Reusken et al. (2014); Meyer et al. (2014); Farag et al. (2015)). Re-infections of MERS-CoV among camels were also identified (See Alagaili et al. (2014); Hemida et al. (2014a)).

In this work, we fit the weekly primary human cases of MERS-CoV with a Susceptible-Exposed-Infected-Recovered-Susceptible (SEIRS) model for the dromedary camel population, with a constant proportion of primary human cases from camels (i.e. either 25% or 12%). We considered two scenarios: (i) the transmission rate among camels is time-varying while the spill-over rate is constant; (ii) the spill-over rate is time-varying while the transmission rate is constant. Our aim is to find out whether simple epidemic models with largely biologically plausible parameters can simulate the epidemics of MERS-CoV among camels. If yes, our objectives are to answer the followings: (i) How prevalent is the MERS-CoV infection among

camels? (ii) How fast does the immunity decay among camels? (iii) Do these estimates match those of earlier serological and virological studies?

4.2 Data and mathematical modeling

4.2.1 Data

We obtained weekly MERS-CoV cases from the EMPRES-i Global Animal Disease Information System (See Food and Agriculture Organization of the United Nations (2016)). Monthly percentages of MERS-CoV infected camels from May 2014 to April 2015 in Saudi Arabia are obtained from Sabir et al. (2016).

4.2.2 Mathematical modeling

Our mathematical model is a susceptible-exposed-infectious-recovered-susceptible (SEIRS) framework for camel-to-camel transmission. The model structure is depicted in Figure 4.1. Here, S , E , I and R represent the numbers of susceptible, exposed, infectious and recovered, and C represents the weekly laboratory-based human confirmations. The total population size of dromedary camels, $N = S + E + I + R$, is assumed to be constant at 270,000 (See Kupferschmidt (2014); United Nations (2016)) and the study period went from January 1, 2014 to May 31, 2016.

The maximum lifespan of dromedary camels in captivity could reach 28.4 years (See The Animal Ageing and Longevity Database (2016)). A number of studies showed that calves that are younger than two years old with primary infections were more likely play a stronger epidemiological role in MERS-CoV transmission compared to older camels that are two to four years old. This is because infected calves have frequent viral shedding and demonstrate a lower rate of seroconver-

sion than older camels (See Hemida et al. (2014b); Reusken et al. (2016); Wernery et al. (2015); Meyer et al. (2016a); Hemida et al. (2013); Khalafalla et al. (2015)). Furthermore, a more recent study by Sabir et al. (2016) found that the evolution of MERS-CoV has led to diverse lineages, they suggested that dromedary camels hosted this recombination event. Thus, we suspect the duration of immunity protection induced by MERS-CoV infection could last for many years. The infectious period for camels was in days (See Wernery et al. (2015)), and we assume a latent period of two days and an infectious period of four days.

We focused on weekly human cases in Saudi Arabia from January 1, 2014 to May 31, 2016. In our model (Figure 4.1), ρ is the spill-over rate between camel cases and primary human cases. We assume that the primary human cases are 1/4 of all human cases in any week, and we consider this ratio to be 12% in the Appendix. Thus we convert the reported weekly overall human cases to primary human cases by multiplying 1/4 and round all numbers up to the next integer. Thus, we assume each MERS-CoV camel case leads to roughly 4ρ human cases (i.e. both primary and secondary cases). The assumption of 1/4 is due to the observation that 25% of human cases are likely primary cases. A minor change on this primary case ratio (e.g. from 1/4 \rightarrow 1/2) will change the estimate from $\rho \rightarrow \tilde{\rho}$, but this will keep their product roughly constant, i.e. $4\rho = 2\tilde{\rho}$. Also, it will have a minor impact on the overall fitting and other parameter estimates. However, if the primary case ratio is too small, (e.g. 1/4 \rightarrow 1/10), the estimated primary human cases will become 1 for almost every week due to data transformation, thus we will lose the temporal pattern in the original data. In that case, we need to seriously consider modeling for the human-to-human transmission. In our data, the primary case ratio is not too small, the temporal patterns are largely

kept. Human-to-human transmission is limited to close contacts within a hospital ward or a household setting (See WHO (2014)), all of which justify our simple transformation.

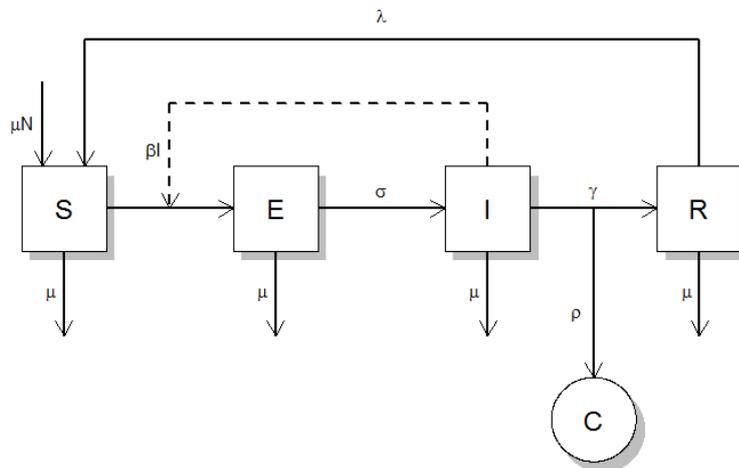


Figure 4.1: SEIRS model structure

The model is as follows:

$$\begin{aligned}
 \dot{S} &= \mu N + \lambda R - \beta(t)SI - \mu S, \\
 \dot{E} &= \beta(t)SI - \sigma E - \mu E, \\
 \dot{I} &= \sigma E - \gamma I - \mu I, \\
 \dot{R} &= \gamma I - \lambda R - \mu R.
 \end{aligned} \tag{4.1}$$

where $\beta(t)$ is the transmission rate, μ are the natural death and birth rates, N is the population size of camels, σ and γ are rates at which infected camels change status from exposed to infectious or from infectious to recovered. Finally, λ is the rate at which recovered camels lose immunity protection due to multiple factors, such as evolution of the virus, natural loss-of-immunity (See Sabir et al. (2016); Meyer et al. (2016b)), imports of susceptible juvenile camels and exports or consumption of recovered adult camels (See Knoema (2016)). Susceptible camels

could also be exported and consumed for meat. We assumed that the replenishment of susceptible camels to be immediate, i.e. the export and consumption of susceptible camels will cancel out with the import of them, and these terms will not appear in the model equations. We also assumed that there is no export or consumption of infectious camels, which is acceptable if the prevalence is less than 10% during the study period.

Thus, the model simulated weekly number of primary human cases due to spill-over are:

$$Z_i = \int_{\text{week}_i} \rho\gamma I dt \quad (4.2)$$

where I is the number of MERS-CoV camel cases at any time. We denoted C_i as the observed number of primary human cases in i th week, and we assume it follows Negative-Binomial (NB) distribution (R version 2.15.2)

$$C_i \sim \text{NB}(\text{mean} = Z_i, \text{variance} = Z_i(1 + \tau Z_i)) \quad (4.3)$$

where τ is an over-dispersion parameter which will be estimated. When $\tau = 0$, the NB distribution is reduced to Poisson distribution.

The log likelihood function is

$$\mathfrak{l}(\theta) = \sum_{i=1}^n \log f(C_i | C_{1:i-1}, \theta) \quad (4.4)$$

where θ is the set of unknown parameters and $f(C_i | C_{1:i-1}, \theta)$ are the conditional densities for C_i given $C_{1:i-1}$, which will be numerically calculated via Sequential Monte Carlo (see He et al. (2010)).

We use Partially Observed Markov Process (POMP) iterated filtering method (See Ionides et al. (2006); King et al. (2008); Ionides et al. (2011); King et al. (2016);

Ionides et al. (2015)) within a plug-and-play likelihood-based inference framework He et al. (2010) to generate the maximum likelihood estimates for θ (R Package ‘pomp’ is available at King et al. (2019)). A tool for maximum likelihood inference on partially observed dynamical system is the iterated filtering algorithm. For parameter space exploration, stochastic perturbations to the unknown parameters are introduced. Sequential Monte Carlo, i.e. the particle filter, is applied to the extended model and this will result in selection of parameter values that are more consistent with the data. If these procedures are well constructed, then iterated filtering with successively diminished perturbations, will result in the convergence of the maximum likelihood estimate. This approach is non-trivial. It is described in more detail in the Appendixs, technical documentation of R package (See King et al. (2019)) and wikipedia (See Ionides (2019)). Infectious disease modeling studies such as Ebola, cholera, malaria, influenza, pertussis, HIV, measles and polio virus have used this method (See Ionides (2019)). The merits of this package are fivefold. (i) Only model simulation is required to compute the likelihood, with only a few lines of C programming codes that describes the model equations; (ii) It can be implemented in the R software package which is freely available, and where C programming codes can be executed; (iii) hidden state variables are calculated and are available in the output; (iv) one can assume a parameter to be time-dependent (such as a cubic-spline function of time) and then estimate the flexibility. Flexibility is reflected by the number of nodes, or degrees of freedom, in a time-dependent cubic spline function. If the number of nodes is very large, then the parameters represented by the spline function could change very rapidly over time, or vice versa. This is an advancement compared to previous models where constant parameters (or a particular function, such as an exponential function)

were frequently assumed; (v) The POMP package can be run on high performance workstation in a parallel or a serial manner, which makes large-scale modeling selection feasible. POMP is described in more detail in Appendix S1.

Bayesian information criterion (BIC) (See Schwarz et al. (1978)) is used for assessing the performance of different models:

$$BIC = -2l(\hat{\theta}) + N_p \ln N_d \quad (4.5)$$

where N_d denotes the number of data points and N_p denotes the number of free parameters. We use BIC rather than Akaike Information Criterion because the former generally penalizes free parameters more strongly.

We studied the following modeling scenarios: (i) we assumed only transmission rate is time-dependent and the flexibility is determined by BIC; (ii) we assumed only transmission rate is time-dependent without restriction on the flexibility; (iii) we assumed only the spill-over rate is time-dependent and the flexibility is determined by BIC.

4.3 Results

Figure 4.2 shows the spatio-temporal patterns of MERS-CoV overall human cases in Saudi Arabia and worldwide, monthly percentages of MERS-CoV infectious camels in Saudi Arabia Sabir et al. (2016), noting Hajj and Ramadan periods. The top panel shows the distribution of confirmed cases of MERS-CoV infection worldwide, which are mostly concentrated in the Middle East, followed by the Republic of Korea, and have spread to 27 countries. The bottom panel shows the temporal patterns of MERS-CoV human cases and percentages of MERS-CoV camels cases. Major MERS-CoV waves peaked in May 2014, February 2015, June

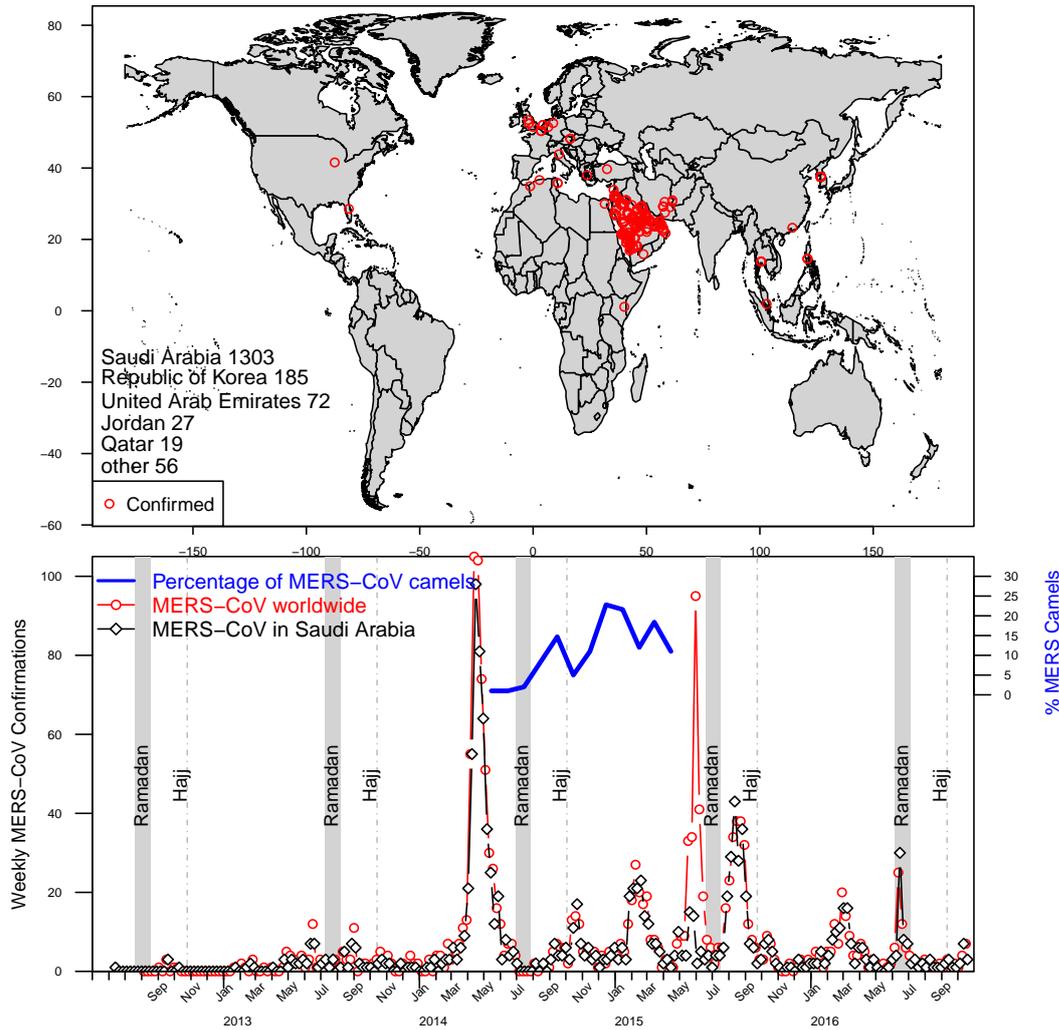


Figure 4.2: Spatio-temporal patterns of MERS-CoV human cases and percentage of MERS-CoV camels. The top panel summarises the cumulative global cases from September 2012 to May 2016. The bottom panel shows the weekly confirmed cases of MERS-CoV human cases worldwide (red lines with circles) and in Saudi Arabia only (black lines with diamonds). The bold black curve shows a monthly percentage of MERS-CoV infectious camels from May 2014 to April 2015. There were three waves in the MERS-CoV camels which were followed by three waves in the weekly human MERS-CoV cases. We also showed the period of religious events, Hajj (light grey vertical dash line) and Ramadan (shaded bars). The map was made with R programming language, and the country borders were downloaded from Sandvik B., World Borders Dataset, <http://thematicmapping.org>(2009).

2015 (South Korea), August 2015 and March 2016. Most of them occurred mainly in Saudi Arabia, except that the third wave occurred mainly in the Republic of Korea. The percentage of MERS-CoV infectious camels showed three waves in a year (May 2014 to April 2015) which were followed by three waves in the weekly human MERS-CoV confirmations. This phenomenon suggests that the waves in the human cases could be driven by the epidemic waves of infections among camels.

The transmission rate, β , can be affected by climatic factors, as well as closing-and-reopening of markets during general holidays or religious festivals. Thus, we used a cubic spline to model the transmission rate. We uniformly distributed n_β nodes over the time period, where n_β is to be estimated. The cubic spline function was a smooth function that passed through the n_β nodes. We compared models with different number of nodes and found the best model which attained the smallest BIC. We fixed the following parameters within biologically reasonable ranges: the lifespan of camels at 14 to 20 years, the latent period at 2 days, and the infectious period at 4 days. The results are shown in Figure 4.3. Panel (a) shows the case of the life span μ^{-1} of 14 years, while panel (b) shows that of 20 years. The inset panels on the left show the profile of the model BIC as a function of n_β . In both cases, the smallest BIC was attained when n_β is 10. Therefore, in the main panels, we display the model fitting results with n_β is 10. The inset panels on the right show the profile of MLL as a function of the spill-over rate ρ . In both simulations, the models were able to capture the four major epidemic waves out of six waves during the period. We found that the duration of immunity is about one year in both cases, or more specifically, 0.86 year in panel (a) and 1.14 year in panel (b). We found that the spill-over rate was about 0.00053 (95% confidence interval (CI) 0.00042, 0.0028) and 0.00065 (95% CI 0.00044, 0.00268),

respectively, which corresponded to the prevalence of MERS-CoV infection among camels to be as high as 9%. This largely matched the observed prevalence of 11% (See Sabir et al. (2016)). A smaller spill-over rate implies a higher prevalence. Through computing the models' profile MLL as a function of λ^{-1} , we estimated the 95% CI of the duration of immunity to be (0.34 year, 5.82 years) in panel (a) and (0.35 year, 7.09 years) in panel (b). The estimated mean \mathcal{R}_0^{camels} were 2.71 and 3.34.

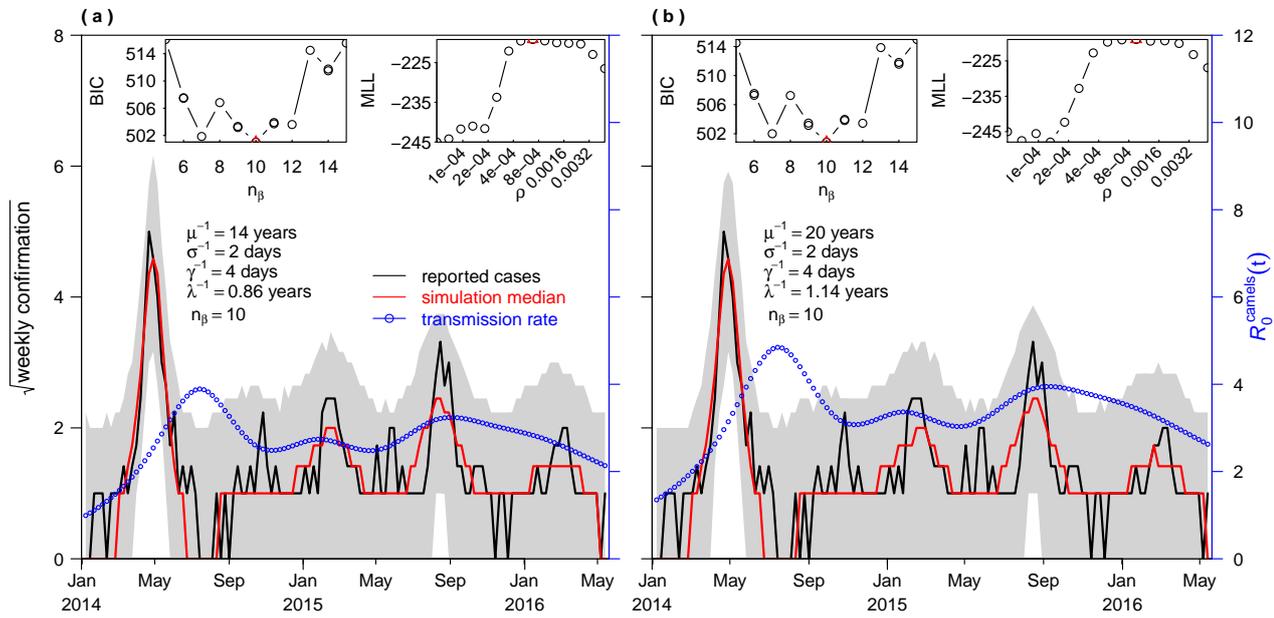


Figure 4.3: Model simulations versus observed MERS-CoV spill-over cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and blue curve (with circles) represents the transmission rate (in units of \mathcal{R}_0). Inset panels on the left show the profile of BIC as a function of the number of nodes in the transmission rate. Inset panels on the right show the profile of MLL as a function of the spill-over rate ρ , while fixing $n_\beta = 10$.

The above results used standard model selection approaches to achieve the optimal flexibility in the transmission rate. Recall that the transmission rate

could be affected by climatic factors and camels-related activities, thus the real flexibility in the transmission rate could be higher than our model estimate. If we choose a large n_β , such as 17 or 19, we could achieve improved fitting as shown in Figure 4.4. Panel (a) shows the fitting results with $n_\beta = 17$ and panel (b) shows that of $n_\beta = 19$. In the inset panel, we show the model's profile MLL as a function of λ^{-1} . The MLL estimates of λ^{-1} are 0.62 year (95% CI: 0.27 year, 3.07 years) and 0.44 year (95% CI: 0.08 year, 3.23 years). Again, we found that the duration of immunity protection was short. More importantly, the estimated maximum prevalence of MERS-CoV among camels was about 15% which is consistent with a previous study (See Sabir et al. (2016)). The estimated spill-over rates were about 0.00039 and 0.00033 respectively, which were lower than those in Figure 4.3. The estimated mean \mathcal{R}_0^{camels} were 3.31 and 2.71.

So far, we have limited knowledge about the spill-over rate ρ , and we have assumed it to be constant. It was possible that both transmission rate and spill-over rate could vary over time. However, if we assume both to be flexible and then modelled them as cubic spline functions, the model would be too complex. Alternatively, we could assume the transmission rate to be constant and then allowed the spill-over rate to vary. We repeated the fitting in Figure 4.3, but assumed constant β and time-varying ρ . We showed the results in Figure 4.5. The BIC is worsened by 4.13, and the difference is marginally significant. The estimated λ^{-1} is still short, and the results were consistent. The prevalence of MERS-CoV among camels is now lower than 0.04. Thus we rejected this scenario. In reality, the observed patterns in the spill-over cases could be due to variations of other factors. the estimated spill-over rate was about 0.00053 and 0.00056 respectively, which were comparable to that in Figure 4.3. The estimated \mathcal{R}_0^{camels} were 3.67

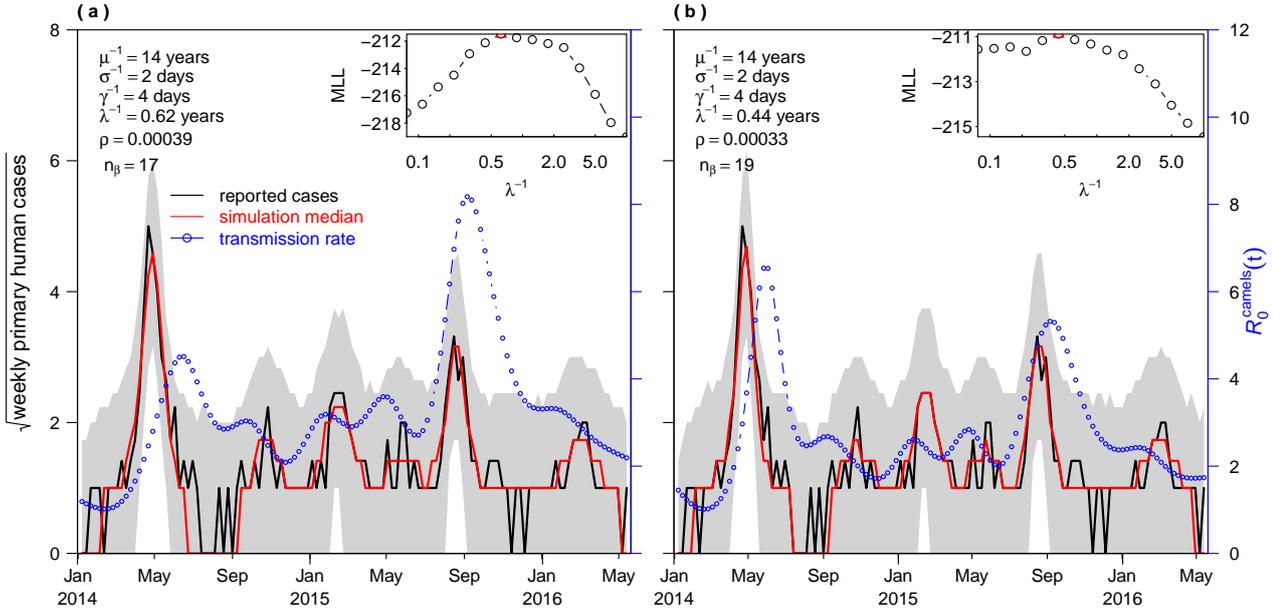


Figure 4.4: Model simulations versus observed MERS-CoV spill-over cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and blue curve (with circles) represents the transmission rate (in units of \mathcal{R}_0). Inset panels show the profile of maximum log likelihood as a function of the duration of immunity protection.

and 3.88 respectively.

4.4 Conclusion

We used a mechanistic model and likelihood-based inference framework to investigate the transmission pattern of MERS-CoV among dromedary camels in Saudi Arabia from January 1, 2014 to May 31, 2016. To our knowledge, our methodology in studying MERS-CoV is novel. Previous mathematical modeling studies on MERS-CoV had primarily focused on human-to-human transmission (See Poletto et al. (2014); Kucharski and Althaus (2015); Hsieh (2015); Chowell et al. (2014); Poletto et al. (2016); Breban et al. (2013); Cauchemez et al. (2016)). Our work was

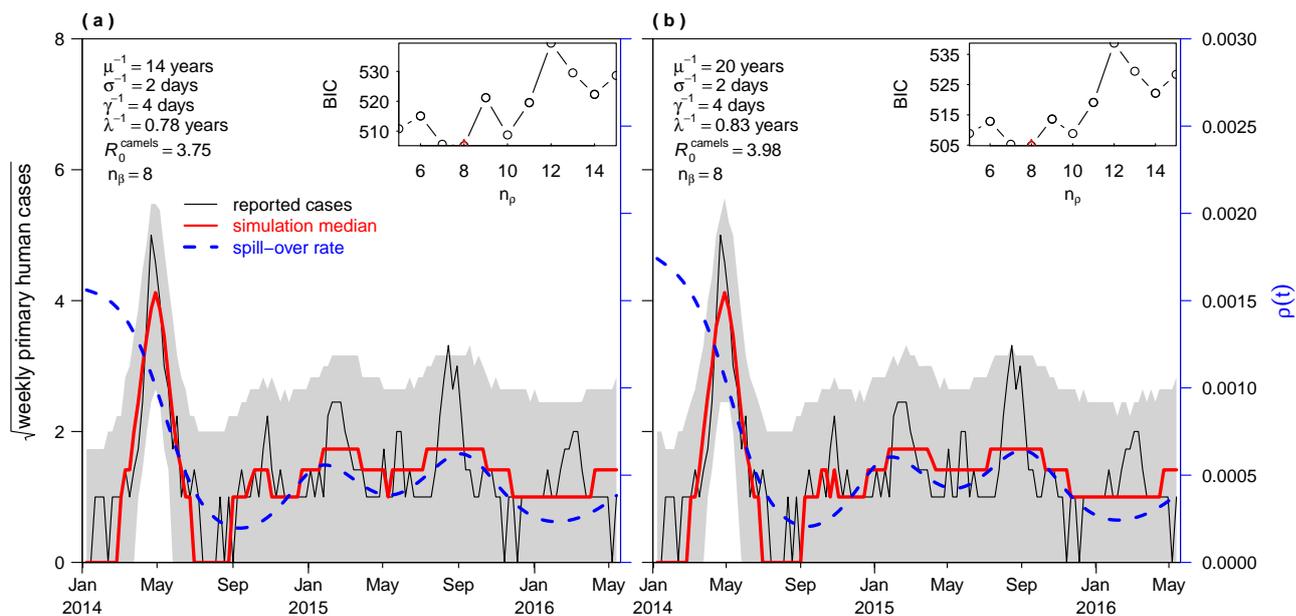


Figure 4.5: Model simulations versus observed MERS-CoV spill-over cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and dashed blue curve represents the spill-over rate. Inset panels show the profile of BIC as a function of the number of nodes in the spill-over rate.

largely motivated by the serological findings in Sabir et al. (2016) which showed that there were high and fluctuating prevalence of MERS-CoV among camels. We noted that the fluctuation of the MERS-CoV camels matched the waves in the weekly human confirmation, although this comparison was done for only one year (May 2014 to April 2015) (See Sabir et al. (2016)). For the best-fitting model, the changes in transmission rate compared to the other models were not very dramatic. We considered three modeling scenarios. We found short immunity duration or high replenishment rate of camels. This is in-line with previous serologically studies. For instance, Meyer et al. (2016b) found that serum from 6 of 11 calves had completely lost their neutralizing activity after 56 months. We also found high prevalence of MERS-CoV infections among camels which matches the results from

Table 4.1: Summary of parameter settings and estimates in Figs 3-5. The mean basic reproductive number is defined as $\mathcal{R}_0 = \langle \beta \rangle / \gamma$, where $\langle \cdot \rangle$ is the time average. $S.0$, $E.0$, $I.0$, and $R.0$ are initial conditions; In Figs 3 and 4, ρ is assumed constant and $\beta(t)$ is time-dependent, whereas these are reverse in Fig 5. Camel population size $N=270,000$. $\sigma^{-1} = 2$ days. $\gamma^{-1} = 4$ days.

| Parameter | Fig.3a | Fig.3b | Fig.4a | Fig.4b | Fig.5a | Fig.5b |
|--------------------------|----------|----------|----------|----------|----------|----------|
| μ^{-1} (years) | 14 | 20 | 14 | 14 | 14 | 20 |
| λ^{-1} (years) | 0.86 | 1.14 | 0.62 | 0.44 | 0.78 | 0.83 |
| n_β | 10 | 10 | 17 | 19 | 8 | 8 |
| $\langle \beta \rangle$ | 247.01 | 305.05 | 301.95 | 247.65 | 342.55 | 363.63 |
| \mathcal{R}_0^{camels} | 2.707 | 3.343 | 3.309 | 2.714 | 3.754 | 3.985 |
| $\langle \rho \rangle$ | 0.000531 | 0.000653 | 0.000387 | 0.000331 | 0.000542 | 0.000576 |
| τ | 0.003375 | 0.003784 | 0.002881 | 0.004018 | 0.007507 | 0.007338 |
| $S.0$ | 0.79508 | 0.593464 | 0.971608 | 0.980595 | 0.196026 | 0.184083 |
| $E.0$ | 0.003552 | 0.002307 | 0.001719 | 0.001047 | 0.000521 | 0.000499 |
| $I.0$ | 0.003552 | 0.002307 | 0.001719 | 0.001047 | 0.000521 | 0.000499 |
| $R.0$ | 0.197815 | 0.401921 | 0.024954 | 0.017312 | 0.802933 | 0.814919 |
| MLL | -219.18 | -219.1 | -211.49 | -210.89 | -226.06 | -225.89 |
| BIC | 501.03 | 500.86 | 514.57 | 523 | 505.14 | 504.81 |

previous virological studies. Finally, $\mathcal{R}_0^{camels} \approx 3$ or 4 is biologically reasonable and in-line with that of Severe Acute Respiratory Syndrome (SARS), which also belonged to the same virus family as MERS-CoV.

Many external factors could impact the transmission of MERS-CoV among camels, from camels to human, and among humans. The transmission rate of MERS-CoV could vary over time, so did the spill-over rate from camels to human, and transmission rate among human. In this work, we attempt to address the transmission of MERS-CoV among camels and also from camels to human (spill-over). In the future, we could include human-to-human transmission into our framework. The reason we did not include this part in current work is due to lack of detailed data. Since the \mathcal{R}_0 among human is substantially less than 1, the transmission among human was limited in hospital setting and in household

transmission, we believe that the transmission among camels played the key role to form the observed temporal pattern in human cases.

In the main text, we assume that the infection from reservoir was 1/4 due that this number was used in WHO reports (See WHO (2014)). In the Supplementary, we show the results when this ratio is 12% (See Cauchemez et al. (2016)). Our main conclusions largely hold. In particular, we observed high rate of loss-of-immunity (and replenishment of susceptible pools), low spill-over rate (ie, high prevalence among camels) and reasonable \mathcal{R}_0^{camels} .

We proposed a one-host (camel) model in the main text. In the Appendixs, we extended it into a two-host (i.e. camels and human) model and presented some preliminary findings. These results have led to similar conclusions as in a crude one-host model. We considered possible impact of climatic factors (air temperature and absolute humidity) and showed preliminary analyses in the Appendixs. There were no evident correlation between MERS-CoV outbreaks and the timing of Hajj.

We compared the monthly prevalence of MERS-CoV infections among camels with weekly human cases for a year in Figure 1. The waves in camels cases were followed by the waves in human, which implies that camel cases have led to human cases. Longer surveillance among camels and phylogenetic work are needed to further clarify this issue. Nevertheless, our results will motivate modeling work on the transmission of MERS-CoV among camels. In the Appendixs, we illustrated how a simple two-host model could be realized. Our modeling work could be improved if data on primary and secondary human cases are available. Our framework could be extended to incorporate two time series. We assumed constant camels population size, which could be varying indeed. We expected that our fitting could still be robust if the population size changed mildly during a year or over the study

period. The changes in the population size in a susceptible pool can be translated into changes in the transmission rate (See Earn et al. (2000)). It is challenging but worth trying to model the estimated transmission rate using actual factors such as climate, closing and reopening of markets, holidays or evolution of the virus and come up a mechanistic model of these driving forces.

The major strengths of our study are that we had used a simple epidemic model to study the prevalence of MERS-CoV among camels, and that our estimates are biologically reasonable. However, our study was still subject to limitations. First, since current animal health surveillance did not routinely report MERS-CoV cases among dromedary camels, we could only assume that primary human cases were at a ratio to camel cases (i.e. constant or time-varying spill-over rate). These assumptions were appropriate when cases were predominantly zoonotic and that the propensity for MERS-CoV to cross species boundaries did not change over time. In the future, more detailed epidemiological information on human cases could improve the accuracy of estimating weekly primary cases. Animal surveillance and control of MERS-CoV in Saudi Arabia should also target dromedary camels. This could help reduce sporadic introductions to humans. Effective measures to identify, diagnose and isolate infected dromedary camels were much needed. Preventive measures should be undertaken at the animal/human interface to reduce zoonotic transmission of MERS-CoV.

Chapter 5

Phylodynamics

The term “phylodynamics” was first proposed by Grenfell et al. (2004), which indicates the diversity of both epidemiological and phylogenetic patterns and their correlations in the virus. In this chapter, we introduced the fundamental theory of phylodynamics and the frameworks that utilize these theory to practical works. Furthermore, on the basis of existent framework, we applied new algorithms to produce a more flexible framework. At the end, we discuss the existing problems and potential threats of these frameworks.

5.1 Moran model, coalescent theory, state space model

In population genetics, the Moran model (See Ewens (1979)) describe the dynamics of genes in a population with a constant size n . A “death-and-birth” event occurs in each step, indicating that one randomly-selected lineage gets eliminated and another gets duplicated, as shown in figure 5.1. In terms of phylodynamics, we are interested in the genealogy of the existing lineages at a given time, as well as how they integrate and construct the historical trace of evolution. One can study the Moran model by investigating its two independent components: the topology

of the genealogy, and the branch lengths. Regarding the branch lengths, one can derive the coalescent model from the Moran model (See Wakeley (2001)).

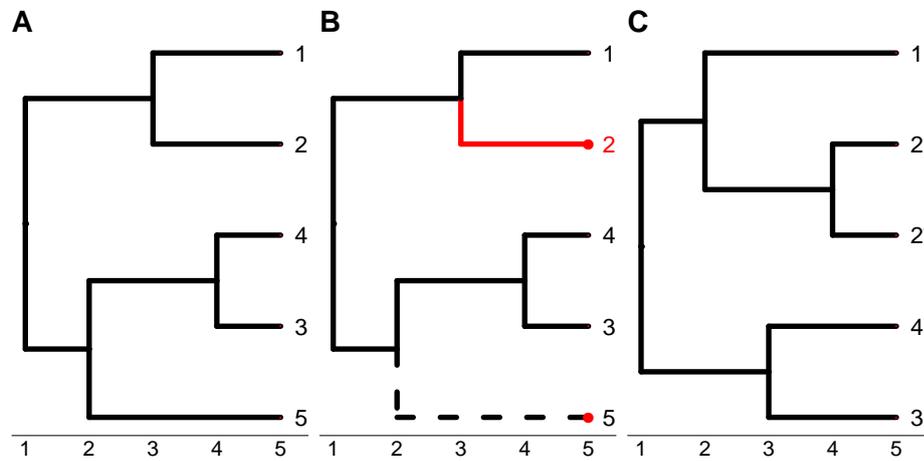


Figure 5.1: Demonstration of the Moran model. (A) Five lineages construct a genealogy. (B) At some moment, lineage 5 (in dashed line) is selected to be eliminated, while lineage 2 (in red) to be duplicated. (C) The new genealogy consists of five lineages, where two of them are duplicated.

The well-known “Coalescent Theory” (Branching Theory) (See Kingman (1982)) is widely used in phylogenetic studies. Meanwhile, compartmental models (See Kermack and McKendrick (1927)), such as Susceptible-Infectious-Recovered (SIR) model, are used extensively as epidemiological processes. However, seldom research involved both of them and theories for combining these two kinds of biological data are undeveloped until recently.

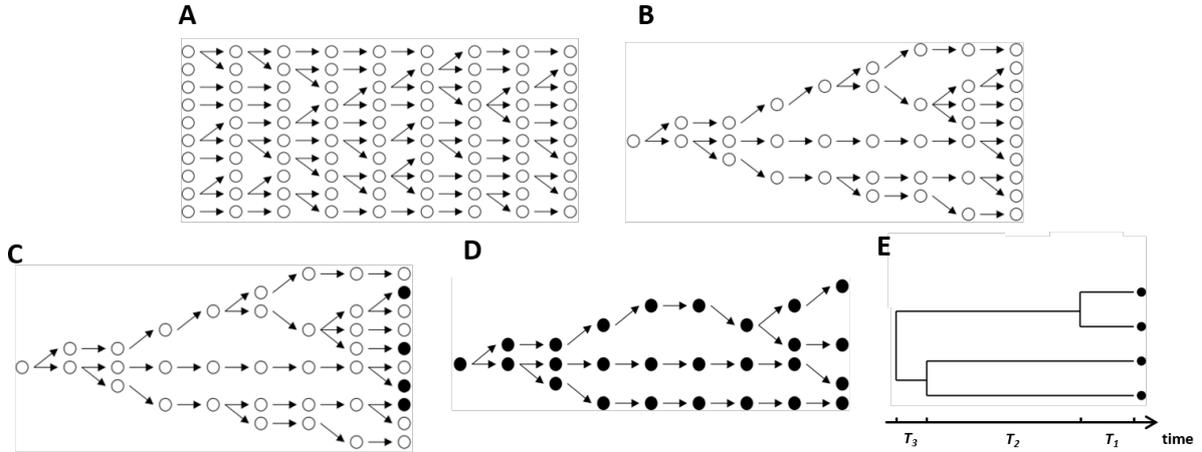


Figure 5.2: Demonstration of the Coalescent theory. (A) The population genetics, (B) the population genealogy, (C) samples $n = 3$ from the population genealogy, (D) the sub-genealogy of the sample, and (E) the version of coalescent theory for the sub-genealogy. In panel (E), T_1 , T_2 , and T_3 represents the coalescent time between coalescent events, in backward time.

Coalescent theory is a model of how gene variants sampled from a population may have originated from a common ancestor. In the simplest case, coalescent theory assumes no recombination, no natural selection, and no gene flow or population structure (See Kingman (1982)), meaning that each variant is equally likely to have been passed from one generation to the next. The model looks backward in time, merging alleles into a single ancestral copy according to a random process in coalescence events. In a population with i lineages and N_e effective number of individuals, the coalescent rate is:

$$p = \binom{i}{2} \frac{1}{2N_e} \quad (5.1)$$

State Space Model (SSM), is also known as Hidden Markov Model (HMM) or Partially Observed Markov Process (POMP). This model contains two stochastic

processes: hidden states X_t and observed states y_t . The hidden states X_t are assumed to fit for a Markov process, in which the current state X_t only depends on its previous state X_{t-1} , with transition probability $p(X_t|X_{t-1})$. The observed probability is $p(y_t|X_t)$. In this report, X_t refers to the actual infected cases and y_t could be either reported cases or coalescent events.

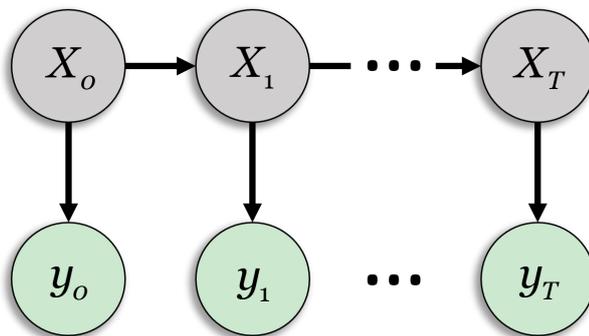


Figure 5.3: State space model.

5.2 Statistical framework in iterated filtering

Pybus et al. (2000) proposed an integrated framework for inferring epidemic history by reconstructing phylogenetic trees, which set a milestone for phylodynamics. Based on Pybus et al. (2000), Volz et al. (2009) connected the numbers of ancestral lineages and the compartmental models and built up a practical framework for phylodynamics. By linking the coalescent time and transmission model and applying particle Markov Chain Monte Carlo (pMCMC) algorithm (See Andrieu et al. (2010)), Rasmussen et al. (2011) made analyses for phylodynamics feasible. However, frameworks based on coalescent models can suffer from conceptual inconsistency (See Smith et al. (2017)) between proposed compartmental models for

disease dynamics and calculation of likelihood using genetic data.

On the other hand, Lau et al. (2011) proposed an integrated system using Bayesian methods for inference both phylogenetics and phylodynamics, which required extra efforts towards specifying MCMC samplers. Smith et al. (2017) developed a methodology by constructing a transmission forest among hosts and linking it to the genealogy. This method circumnavigates the inconsistency mentioned above, but neither allow multiple infections for hosts nor scale well for larger problems, which makes it inapplicable to diseases like Dengue and MERS among camels.

5.2.1 An integrated framework

Rasmussen et al. (2011) proposed a framework to incorporate both genetic data and time series of reported cases based on the coalescent duration. Due to different time intervals may be used in genetic data (e.g. daily) and time series data (e.g. weekly or monthly), a unified time interval named observation time will be used and both data will be adjusted according to new time interval ω_t . Coalescent event is assumed to follow an exponential distribution with the mean time-dependent coalescent rate λ . Therefore, the probability of observing a coalescent event after time t is:

$$f(t) = \lambda_t e^{-\lambda_t t} \tag{5.2}$$

and the coalescent rate is

$$\lambda_t = \frac{\binom{i}{2}}{\binom{I}{2}} \beta(t) \frac{SI}{N} \quad (5.3)$$

where i is the number of lineages in the genealogy.

In observation time interval from $t - 1$ to t (i.e. ω_t), it's divided into k sub-intervals ω_{t_j} ($j = 1, \dots, k$). Then the observation probability in ω_t is $p(\omega_t|x_t) = \prod_{j=1}^k p(\omega_{t_j}|x_{t_j})$. Each sub-interval ends in either coalescent event or no coalescent event, with probability equal to $\lambda_{t_j} e^{-\lambda_{t_j} \omega_{t_j}}$ and $1 - \int_{t=0}^{\omega_{t_j}} \lambda_{t_j} e^{-\lambda_{t_j} t} dt = e^{-\lambda_{t_j} \omega_{t_j}}$, respectively.

Eventually, the joint likelihood of time series and a genealogy at time t is:

$$p(y_t, \omega_t | \theta, x_t) = p(y_t | \theta, x_t) p(\omega_t | \theta, x_t) \quad (5.4)$$

However, calculation for the likelihood of genealogy at current time t involve epidemics in previous k sub-intervals ω_{t_j} ($j = 1, \dots, k$), which is inconsistent with the assumption for the state space model.

5.2.2 Iterated filtering

Iterated filtering is used for estimating parameters and calculating the likelihoods. The improved iterated filtering methods was proposed by Ionides et al. (2015). It implements the particle filtering method (also known as Sequential Monte Carlo (SMC)) (See Del Moral (1996)) iteratively with perturbations to generate parameter samples and calculating the likelihoods. It has been widely used in inference for dynamical systems (See Ionides et al. (2006); Breto et al. (2009)).

Using this method, we build up a customized schematic to include both time series and genetic data to infer the phylodynamics (Figure 5.4).

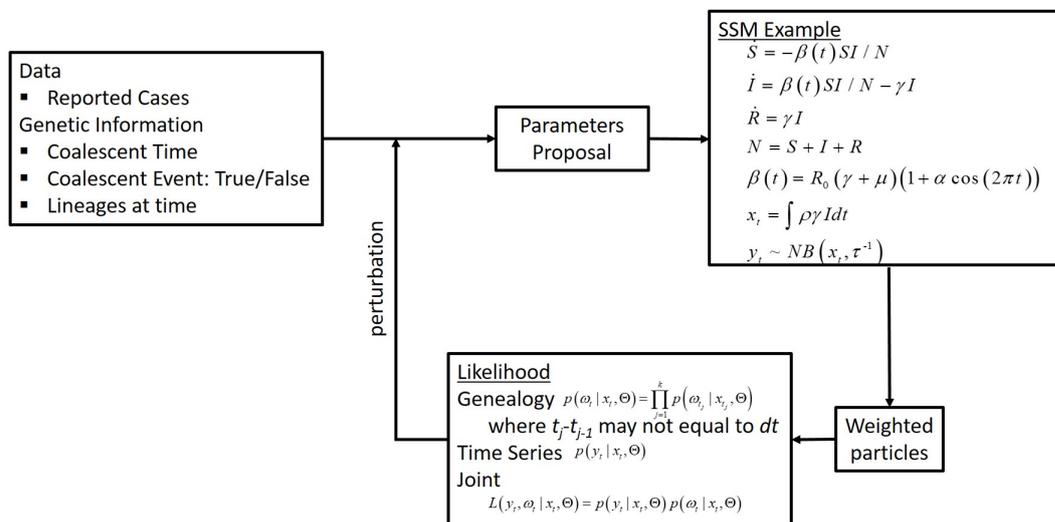


Figure 5.4: Schematic of using iterated filtering method to Rasmussen et al. (2014) framework.

R package ‘pomp2’ is available for implementation (<http://kingaa.github.io/pomp/>).

5.3 Simulation and results

In this simulation, we use the same simulated data from Rasmussen et al. (2014) and the model is as that in Figure 5.4. Within the studied period (*i.e.*, 48 months), Figure 5.5 shows four waves of epidemics of reported cases, and in the meanwhile, roughly consistent four peaks in the corresponding number of lineages. The similar pattern demonstrated in the reported cases and the number of lineages suggest a good quality of data, and the feasibility of our framework.

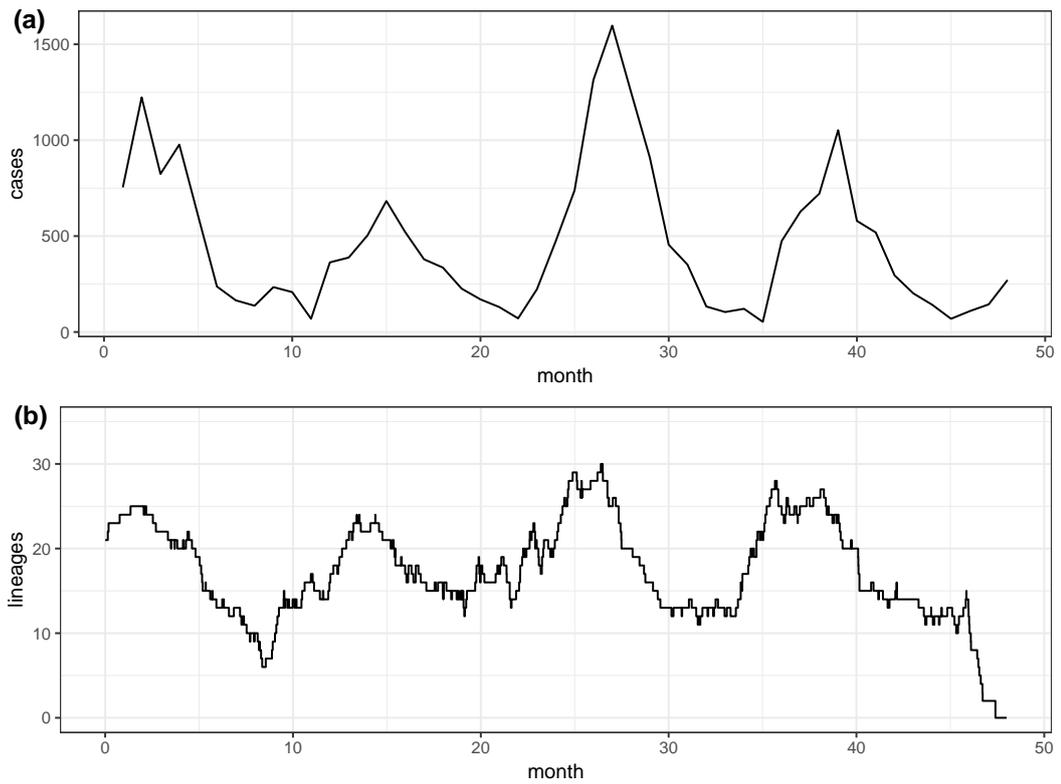


Figure 5.5: Number of reported cases (a) and lineages (b) in the simulated data.

We first applied the framework in a local search for maximum likelihood, to explore the possible range of parameters and the convergence of the algorithm. Figure 5.6 indicates fast convergence to the maximum likelihood without significant failure of converging and the true value of the parameters.

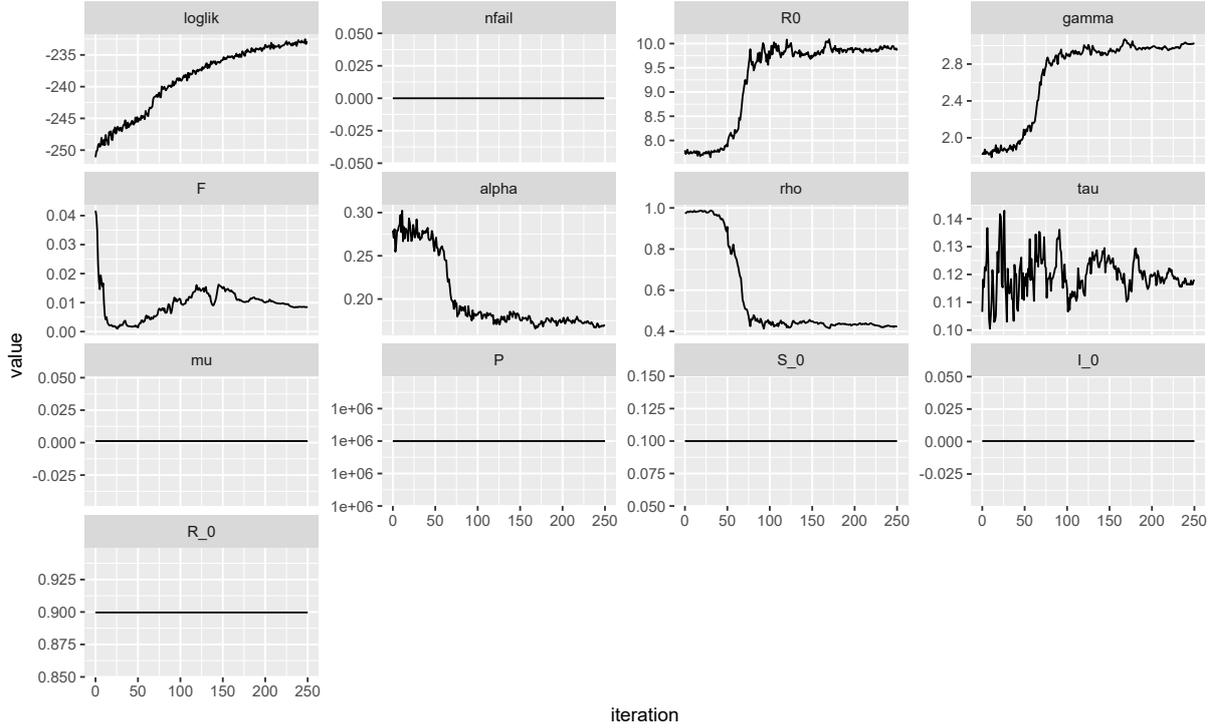


Figure 5.6: Convergence of the estimations by local search, where `loglik` is the log-likelihood $l(\theta)$, `nfail` represents the number of particles which failed to converge, `R0`, `gamma`, `F`, `alpha`, `rho`, and `tau` are parameters of interest, and `mu` is constant birth/death rate, `P` is the constant total population size, `S_0`, `I_0`, and `R_0` is the fixed initial proportion of the susceptible, the infectious, and the recovered.

However, since the iterated filtering method proceeds based on the previous parameter proposal with small perturbations, the local search for the maximum likelihood would not be as reliable as what we want. Therefore, we ran a wide range of initial proposal for searching for the global maximum likelihood and the results are shown in Figure 5.7. It is obvious that initial proposal of parameters that are far from the true value converge to the neighborhood of the true value after a few iterations. Some estimates may diverge, for example, ρ , γ and α , but most of them shrink to the neighborhood fast.

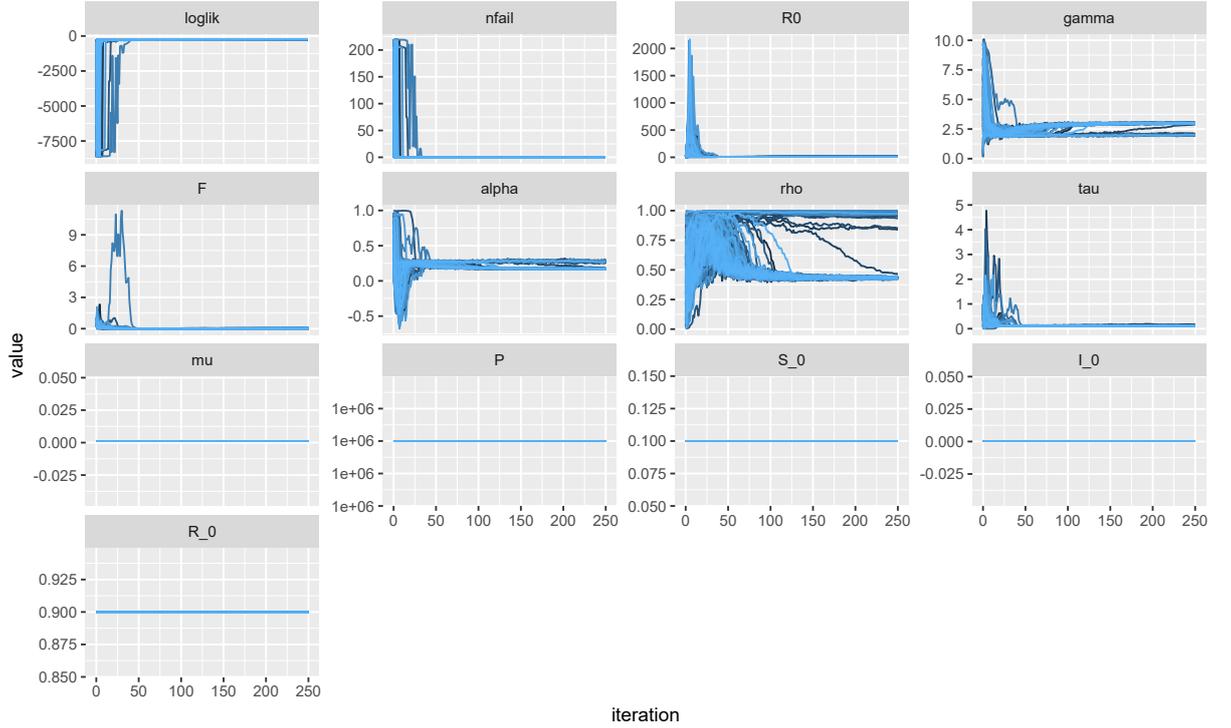


Figure 5.7: Convergence of the estimations by global search. The notation is the same as those in figure 5.6.

Through iterated filtering, we can also construct the profile of likelihood for specified parameter and estimate its confidence interval Ionides et al. (2017), as shown in Figure 5.8 and summarized in Table 5.1. The framework and the algorithm precisely estimate two out of six parameters (*i.e.*, \mathcal{R}_0 and γ), and two estimates (*i.e.*, α and ρ) are close to the real value. However, the aggregated framework via iterated filtering method failed to estimate τ and F , which particle-MCMC method also fails to estimate Rasmussen et al. (2014).

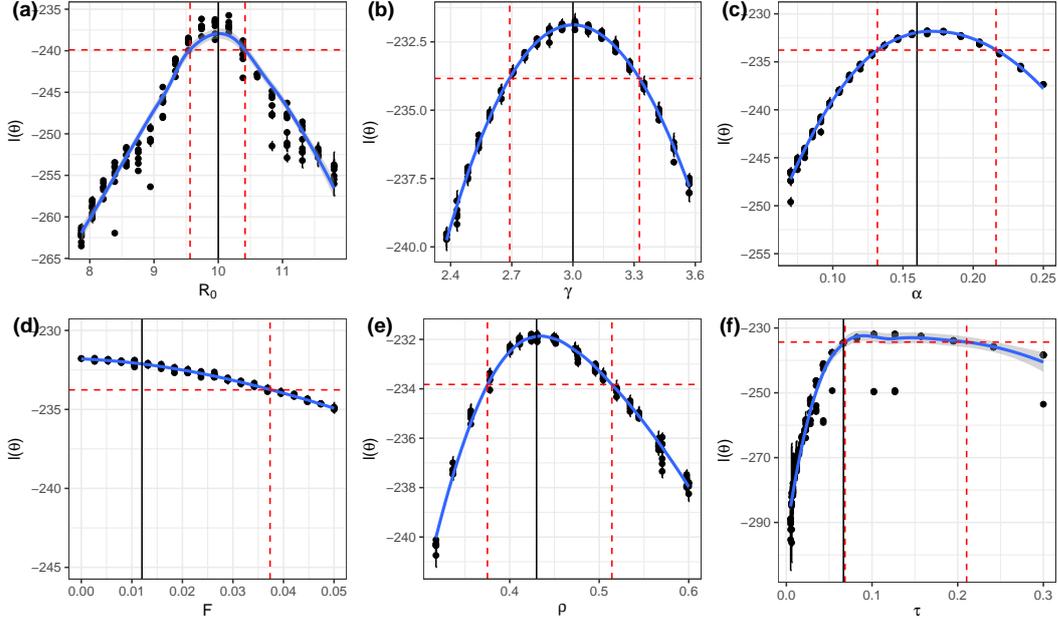


Figure 5.8: Profile log-likelihood for estimations. In each panel, the blue line with grey shades indicates the profile of log-likelihood obtained by smooth spline, red dashed lines indicates the 95% confident log-likelihood and corresponding confidence interval, and black solid line refers to the real value.

Table 5.1: Summary of the parameters and 95% confidence intervals.

| Parameter | Real value | Global search range | Estimation | 95% CI |
|----------------------|-------------------------|---------------------|------------|------------------|
| Estimated parameters | | | | |
| \mathcal{R}_0 | 10 | [0, 30] | 10.00917 | (9.561, 10.417) |
| γ | 3 | [0, 10] | 2.997762 | (2.690, 3.326) |
| α | 0.16 | [0, 1] | 0.1702532 | (0.1318, 0.2163) |
| F | 0.012 | [0, 1] | 0.00001169 | (0, 0.0373) |
| ρ | 0.43 | [0, 1] | 0.4355827 | (0.3751, 0.5142) |
| τ | 0.06667 | [0, 1] | 0.09088608 | (0.0685, 0.2104) |
| Known parameters | | | | |
| μ | 0.0012 | - | - | - |
| P | 1,000,000 | - | - | - |
| S_0 | 0.1 | - | - | - |
| I_0 | $\frac{357}{1,000,000}$ | - | - | - |

Using the maximum likelihood estimates (MLEs), we can simulate the reported cases (Figure 5.9).

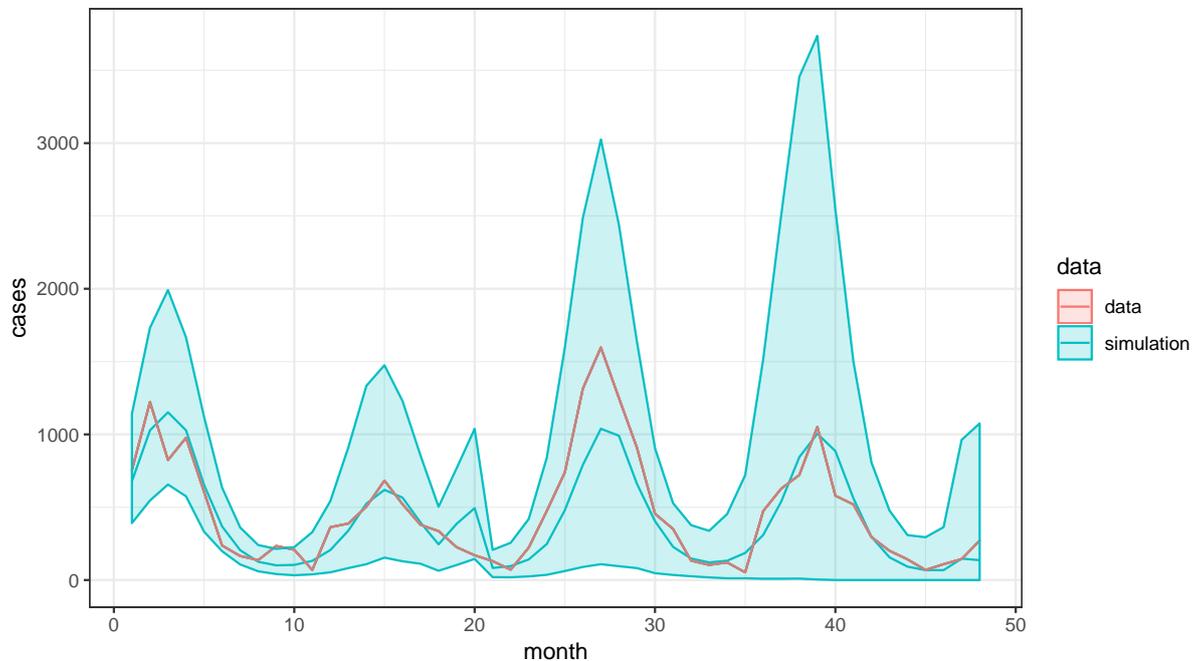


Figure 5.9: Simulation using the best fit estimations. Red line indicates the data, blue line is the simulation from the best estimations. Blue shades indicates the 95% confidence interval of the simulation.

5.4 Discussion and future works

The advantages of current framework include convenient implementation, acceptable computational time, and great precision in estimations. It is easy to collect epidemic data and genetic data, conduct pre-processing, and apply this framework to compute useful parameter estimations. However, inconsistency between backward-time “coalescent theory” and forward-time stochastic compartment model cannot be neglected. Moreover, the term “likelihood” would be inappropriate and need to be improved. Last but not least, Rasmussen et al. (2011) regarded ge-

nealogy as a one-time data that corporate to estimate the parameters. However, genealogy is also a “observed state” and a sample from the underlying population genealogy, and can be sampled asynchronously in multiple times as well.

In Volz et al. (2009), the author explained the rate of observed coalescent within a deterministic process, where the rates derived from the forward time is fully consistent with those the from backward time. However, in the stochastic processes, the consistency of the rates is not necessarily valid, because one can infer various outputs from past to present, as well as from future to present. Unfortunately, to the best of our knowledge, neither Volz et al. (2009) nor Rasmussen et al. (2011) verified the feasibility of their frameworks in a stochastic model, theoretically. This problem is closely related to the relationship between Moran model and Kingman coalescent model, while few studies demonstrated and discussed it theoretically and precisely (See Wakeley and Aliacar (2001); Durrett (2008); Hein et al. (2005); Etheridge (2011)), in terms of the topology, coalescent times, and the number of lineages.

5.4.1 Stationarity of Moran model and coalescent theory

Within a constant population setting, Moran genealogy process is a continuous-time Markov process with an uncountable state space. It can be divided into two parts: the topology of the process and the branch lengths (coalescent times) between events. These two parts belong to different spaces, and are independent to each other. One can prove that the stationarity of Moran genealogy process is the coalescent model by following steps:

1. One can easily prove that the topology of the Moran genealogy is a double stochastic Markov process (See Aldous (1999)), and accompanied with exponential

branch lengths, the Moran genealogy is uniformly ergodic and converges to an invariant limiting distribution (See Meyn and Tweedie (2009); Down et al. (1995)).

2. Then one can compute the infinitesimal generator of the semigroup of the Moran genealogy process (See Feller (1957)) and prove that the unique invariant probability measure of the Moran genealogy process is the coalescent model.

3. Furthermore, with the infinitesimal generator, one can explore the Kolmogorov forward and backward equations.

5.4.2 Asynchronous sampling and simulation

The existing framework helps researchers to estimate key parameters in epidemic models, while it is not practical to simulate the right epidemic dynamics using these estimations. Except for regularly reported data from epidemic surveillance departments, genetic samples can also be collected regularly from medical centers, such as hospitals. However, it will be complicated to consider asynchronous sampling process of Moran genealogy as a hidden Markov model (See Smith and Wakefield (2016)). Therefore, it will be essential to investigate the transition probability between states of samples. One essential issue is that, unlike one-time sampling procedure, which has properties similar to those of the original Moran genealogy process, asynchronous sampling procedure shows some “tracebacks” that retrieve more ancient information on evolution.

5.4.3 Improved likelihood

The “likelihood” implemented in current framework is not the standard likelihood, because Rasmussen et al. (2011) didn’t justify the underlying distribution of this “likelihood”. On the other hands, in the setting of asynchronous, one should

assume and justify the distribution of sampled genealogy, then combine the “genealogy likelihood” with the epidemic likelihood.

Chapter 6

Conclusion and Discussion

In the great era of “Big Data”, we have limitless data with great variety, and at the same time we have invented thousands of methods, algorithms, and frameworks that help use investigate these data.

This thesis shows several scenarios in which researchers apply methods with different complexities to a variety of data, from basic time series data, to the combination of graph/tree data and time series data. With limited data, one can still apply statistical methods, such as Pearson Correlation or construct a simple Metric to explore the changes of cases of infectious diseases. Accompanied by the individual information, such as age, gender, some relatively complex statistical inference tools (*i.e.*, APC models) can help to explore the relationship between infection risks and age, gender, birth cohort and infected period. Under the help of systematic mathematical theories, one can build up a complicated dynamical system to investigate key epidemic parameters, consuming much more computational resources and time. Finally, to incorporate different types of data, more delicate and complicated theories and frameworks are designed and much more computational time are consumed.

Nowadays, the focus of using “Big Data” not only focus on how to use them,

but on why to use them. This thesis presents that fact that, using the suitable methods, simple data can still provide some important insights of epidemics, and methods should be designed based on the nature of data, not on the amount of the data.

Appendix A

A.1 Influenza

We further investigate the anti-phase synchronization of A/H1N1 and A/H3N2 in Hong Kong and NTZ using Seasonal and Trend decomposition using Loess (STL) Time Series Decomposition Analyses (See Cleveland et al. (1990)) and Wavelet and Reconstruction Analyses.

A.1.1 Time series decomposition analyses

The reported cases of influenza A subtypes from January 2010 to July 2017 are further analysed by region using time series decomposition analyses. We decomposed the time series into the multiplication of three components: trends, seasonality and remainder components, where the latter two components are investigated (See Cleveland et al. (1990)).

Figure A.1 shows the decomposed seasonality and random components of influenza A subtypes in the NTZ. In the top panel, seasonality component of A/H1N1 displays annual periodicity whereas that of A/H3N2 displays bi-annual periodicity. In the bottom panel, the remaining components of A/H1N1 and A/H3N2 show remarkable antiphase patterns – a peak in A/H1N1 corresponds to a trough in A/H3N2, and vice versa.

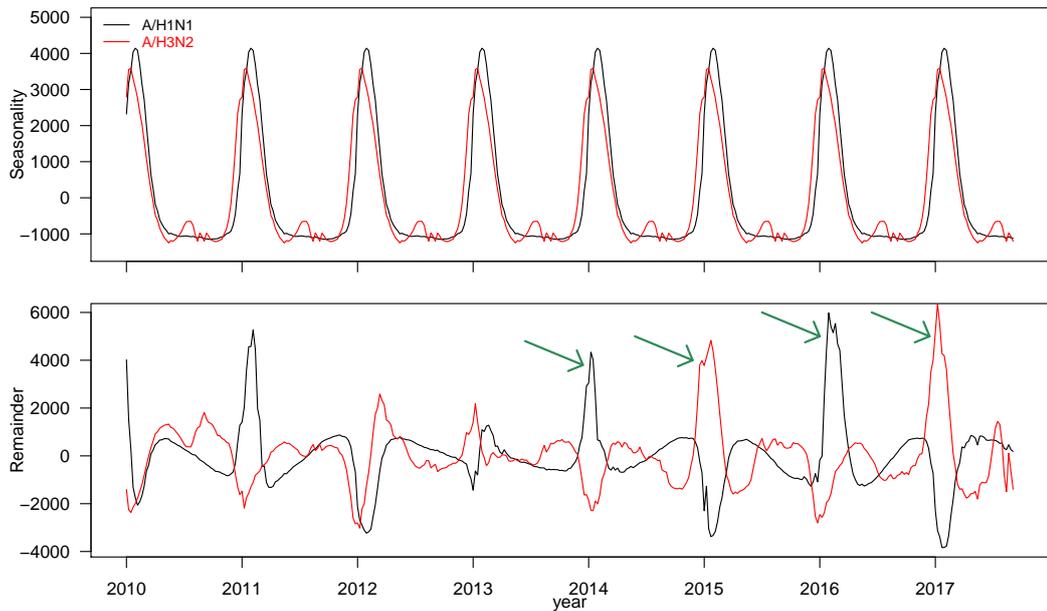


Figure A.1: Decomposed seasonality and random components of influenza A subtypes in the NTZ.

Figure A.2 shows the time series decomposition of that in Hong Kong. In the top panel, seasonality component of A/H1N1 displays annual periodicity while A/H3N2 displays tri-annual periodicity. In the remainder components display more complex patterns than that of NTZ, but the antiphase patterns are still observable.

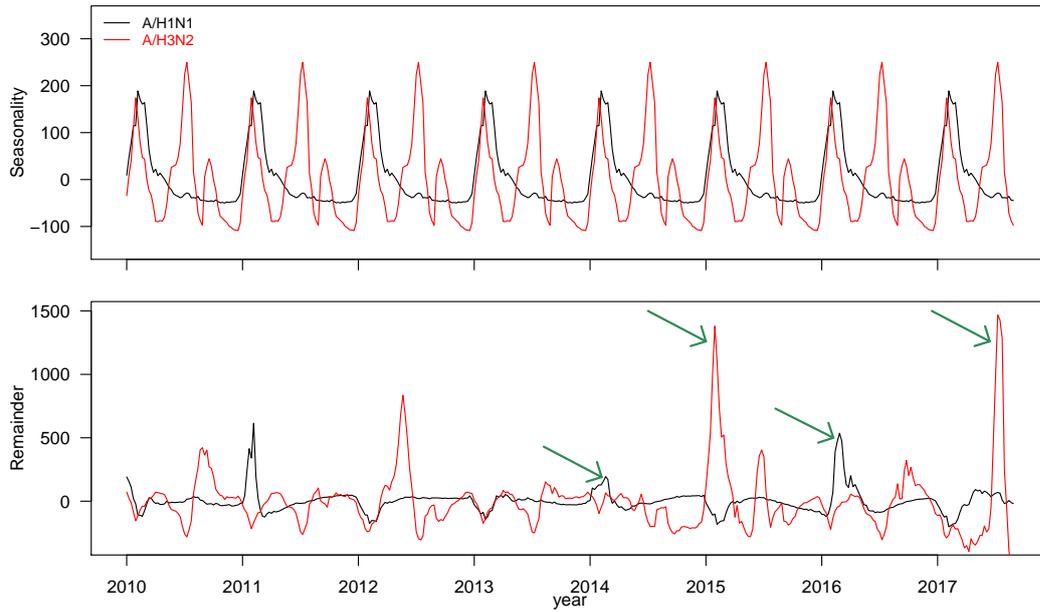


Figure A.2: Decomposed seasonality and random components of influenza A subtypes in Hong Kong.

A.1.2 Wavelet and reconstruction analyses

Wavelet analyses are performed on the influenza A subtypes time series from January 2010 to July 2017. The wavelet periodicities are determined for further Reconstruction Analysis.

Figure A.3 shows the results of wavelet analyses for influenza A subtype data in the NTZ. Wavelet coherence analysis displays both high power and antiphase pattern when the wavelet period is approximately two years (Figure A.3(b)). Phase pattern and phase difference diagram between A/H1N1 and A/H3N2 in NTZ show that after a transient period of several years, the alternative dominance pattern is formed (Figure A.3(c)).

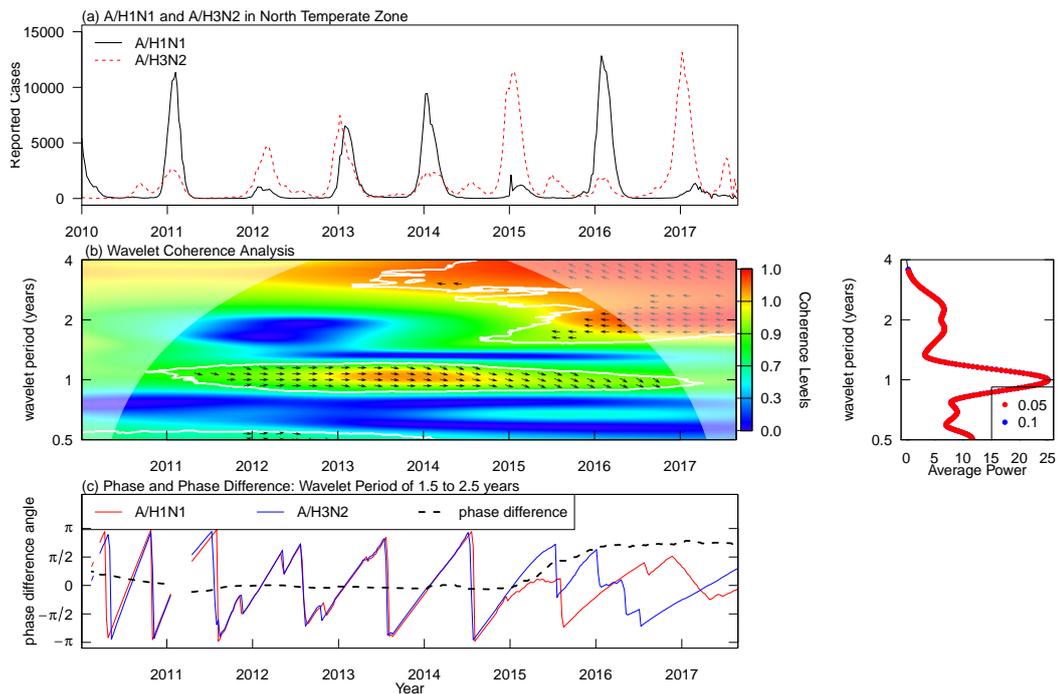


Figure A.3: Wavelet analyses for influenza A subtype data in the NTZ.

Figure A.4 shows the wavelet analyses results of HK. Again, wavelet coherence analysis display both high power and antiphase pattern with a wavelet period of approximately two years (Figure A.4(b)). Phase pattern and phase difference diagram between A/H1N1 and A/H3N2 in HK are more obvious in HK than in the NTZ (Figure A.4(c))

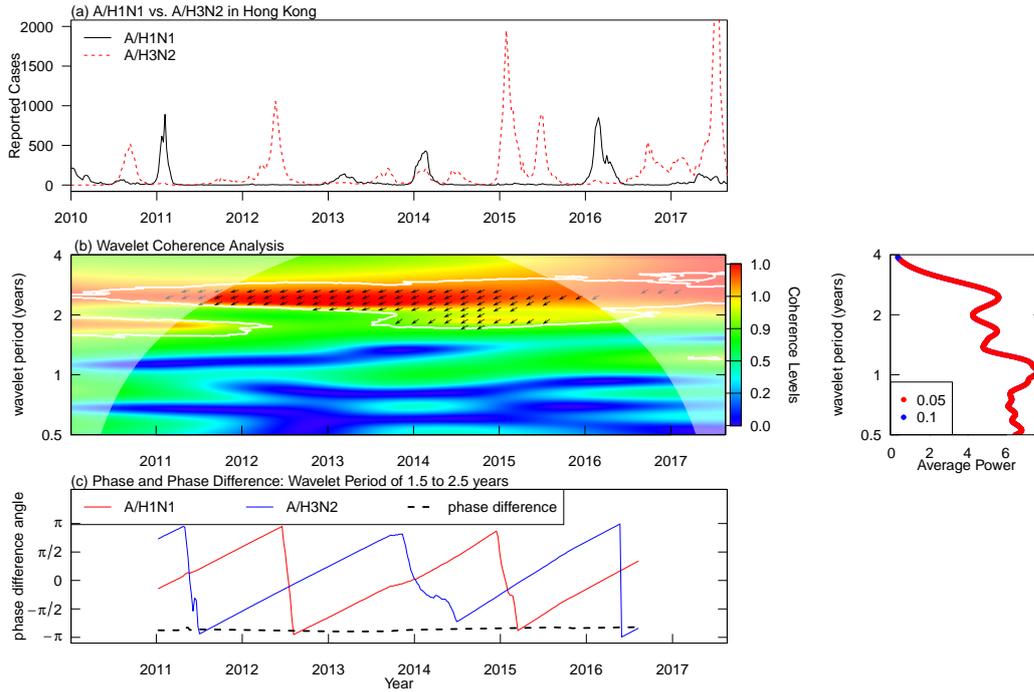


Figure A.4: Wavelet analyses for influenza A subtype data in the Hong Kong.

Based on wavelet analyses, we applied a wavelet period of 1.5 to 2 years in the reconstruction analyses of influenza time series which are shown in Figure 2.2 in Chapter 2.

A.1.3 Time series prediction rule

Based on our exploratory analyses described in the manuscript, we develop a simple time series prediction rule to forecast about A/H1N1 and A/H3N2 epidemics in 2018 and beyond.

The biennial antiphase patterns of A/H1N1 and A/H3N2 described in Chapter 2 and this Appendix are used to develop a simple time series prediction rule. Let $H1(t)$ and $H3(t)$ denote the weekly reported time series of HK or NTZ, where t denotes the time in years, such as $t = 2018.019, 2018.038, 2018.058, \dots$ would

represent the first, second and third week of 2018. We predict that

$$\begin{cases} H1(t) = 0.5H1(t-2) + 0.5H1(t-4), \\ H3(t) = 0.5H3(t-2) + 0.5H3(t-4). \end{cases}$$

where $t > 2018$.

The accuracy of this prediction will depend on how well the anti-phase synchronization patterns persist

A.2 Tuberculosis in Shandong, China

A.2.1 Data Preprocessing

We extracted annual population data from 2006 to 2017 of these seven cities from their Statistics Yearbooks. This dataset contains both total population data, corresponding sex and age structure (i.e., 0–17, 18–35, 35–60, and over 60), and residential population (i.e., people living in the city for over half a year) data. In the meanwhile, we retrieved a detailed age structure (i.e., 0, 1–4, 5–9, 10–14, \dots , 95–99, and over 100) for both sexes for the residential population in the same seven cities in 2010 from The Sixth National Population Census. We combined these two sets of population data by applying the detailed age structure from Census in 2010 to data from Yearbooks from 2006 to 2017, constructing a complete demographic statistic from 2006 to 2017 for all seven cities. We include more details in the Appendix. Eventually, we arrange the TB notifications and demographic data from 2006 to 2017 into 72 age groups, and the DR-TB tested notifications into 34 age groups and seven period groups.

To calculate the exact population for each age for either sex in these cities from 2006 to 2017, we conducted the following steps for either sex for each city:

(i) Computing the residential population for specified sex and every *rough age group* from 2006 to 2017, using the annual ratio of residential population over total population; (ii) Computing the ratios of population in each year over that in the year of 2010 for each *rough age group*; (iii) Applying the above ratios for *rough age group* in annual population data to the *detailed age group* data in 2010 to get the population for each *detailed age group* from 2006 to 2017; (iv) Subdividing each *detailed age group* to get the population of each age, by considering the proportion of each age is the same within the group.

For investigation on TB notification rate, we first eliminate the notifications from individuals who were over 100 years old, due to negligible TB prevalence among the corresponding population (less than 0.03% overall). The first 15 age groups (*i.e.*, 0–14) and the last 16 age groups (*i.e.*, 85–100) are then aggregated in two ensembles: under 15 years old and over 84 years old. We eventually consolidate the notifications and demographic data from 2006 to 2017, which are divided into 72 age groups, for the APC model to explore the effects of age, birth cohort, and calendar period.

For investigation on DR-TB tested notification rate, notifications with age under 17 (0.3% overall) or over 84 (0.18% overall) are first eliminated. To compensate for the shortness of data insufficiency, which leads to 0 records for certain ages in several years, we then combine the notifications every two years. At the end, INH-resistant notifications, RFP-resistant notifications, and MDR notifications are divided into 34 age groups (*i.e.*, 17–18, 18–19, \dots , 81–82, 83–84) and 7 period groups (*i.e.*, 2004–2005, \dots , 2016–2017), respectively.

A.2.2 Heterogeneity in TB incidence in other three cities.

Table A.1: Summary of TB notifications across three cities in Shandong

| | Weifang (N=32 020) | | Jining (N=35 205) | | Dezhou (N=26 642) | |
|------------------------|-----------------------|---------------------|----------------------|---------------------|----------------------|---------------------|
| | no. | % (95% CI) | no. | % (95% CI) | no. | % (95% CI) |
| Gender | | | | | | |
| Male | 22 553 | 70.43 (69.93–70.93) | 25 551 | 72.58 (72.11–73.04) | 17 371 | 65.2 (64.63–65.77) |
| Female | 9467 | 29.57 (29.07–30.07) | 9654 | 27.42 (26.96–27.89) | 9271 | 34.8 (34.23–35.37) |
| Identity | | | | | | |
| Local | 28 065 | 87.65 (87.29–88.01) | 33 753 | 95.88 (95.67–96.08) | 24 615 | 92.39 (92.07–92.71) |
| Immigrant | 3955 | 12.35 (11.99–12.71) | 1452 | 4.12 (3.92–4.33) | 2026 | 7.6 (7.29–7.92) |
| Age groups | | | | | | |
| 0–17 | 1177 | 3.68 (3.47–3.88) | 783 | 2.22 (2.07–2.38) | 511 | 1.92 (1.75–2.08) |
| 18–35 | 11608 | 36.25 (35.73–36.78) | 10441 | 29.66 (29.18–30.13) | 7703 | 28.91 (28.37–29.46) |
| 36–60 | 11177 | 34.91 (34.38–35.43) | 12926 | 36.72 (36.21–37.22) | 9411 | 35.32 (34.75–35.9) |
| over 60 | 8058 | 25.17 (24.69–25.64) | 11055 | 31.4 (30.92–31.89) | 9017 | 33.85 (33.28–34.41) |
| Occupations | | | | | | |
| Students | 2720 | 8.49 (8.19–8.8) | 1382 | 3.93 (3.72–4.13) | 807 | 3.03 (2.82–3.23) |
| Workers | 2511 | 7.84 (7.55–8.14) | 1302 | 3.7 (3.5–3.9) | 1309 | 4.91 (4.65–5.17) |
| Migrant workers | 338 | 1.06 (0.94–1.17) | 224 | 0.64 (0.55–0.72) | 605 | 2.27 (2.09–2.45) |
| Farmers | 22 912 | 71.56 (71.06–72.05) | 29 345 | 83.35 (82.97–83.74) | 21 211 | 79.61 (79.13–80.1) |
| Household / Unemployed | 812 | 2.54 (2.36–2.71) | 781 | 2.22 (2.06–2.37) | 924 | 3.47 (3.25–3.69) |
| Others & unknown | 2727 | 8.52 (8.21–8.82) | 2171 | 6.17 (5.92–6.42) | 1786 | 6.7 (6.4–7) |

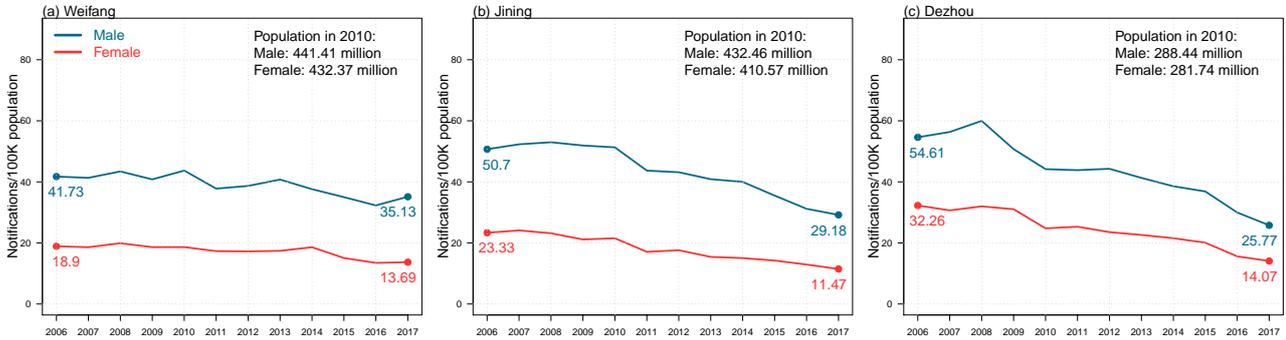


Figure A.5: Population and TB notifications from 2006 to 2017 in three cities in Shandong Province, China. We use a group of four subpanels to show the data for each city. The subpanels on left/right in each group, colored in blue/red, show data for male/female, respectively. The subpanels on the top show annual total TB notifications (bar) and their change rates (curve) from year to year, while the subpanels on the bottom show annual population size (bar) and TB notification rates (curve, per 100K population).

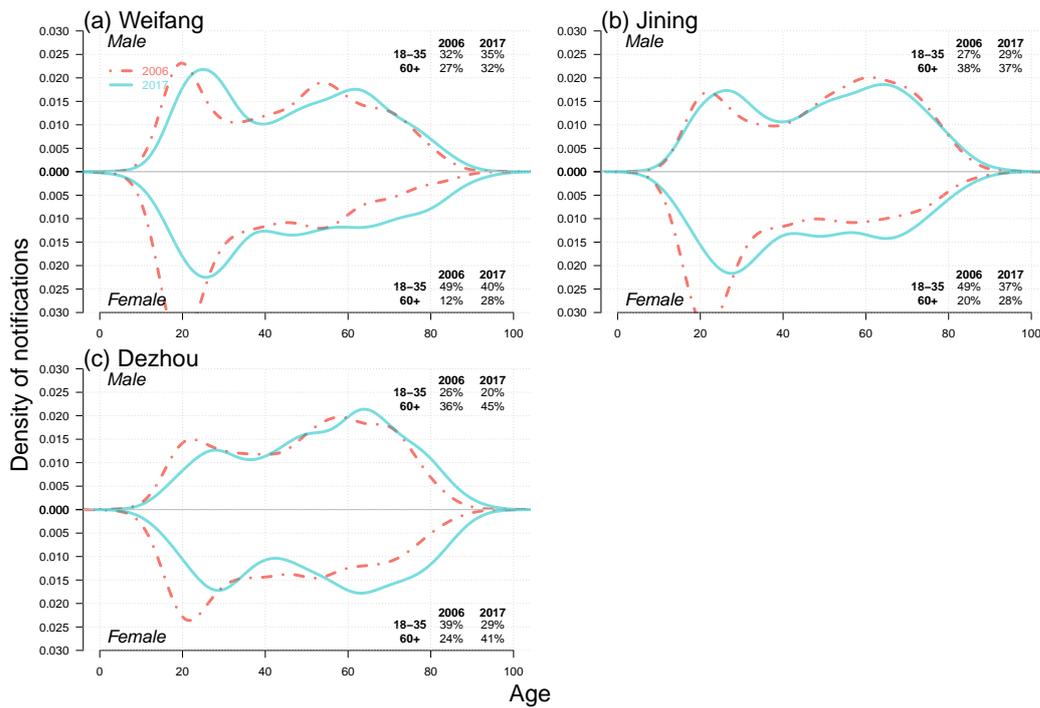


Figure A.6: Case density changes in 2006 and 2017 for males (upper subpanels) and females (lower subpanels) in (a) Weifang, (b) Jining, and (c) Dezhou. In each panel, red dashed-and-dotted and green solid lines represent densities in 2006 and 2017, respectively. We show the changes in case proportions for young adults (18–35) and seniors (60+) in 2006 and 2017 in the top-right corner for males and in the bottom-right corner for females.

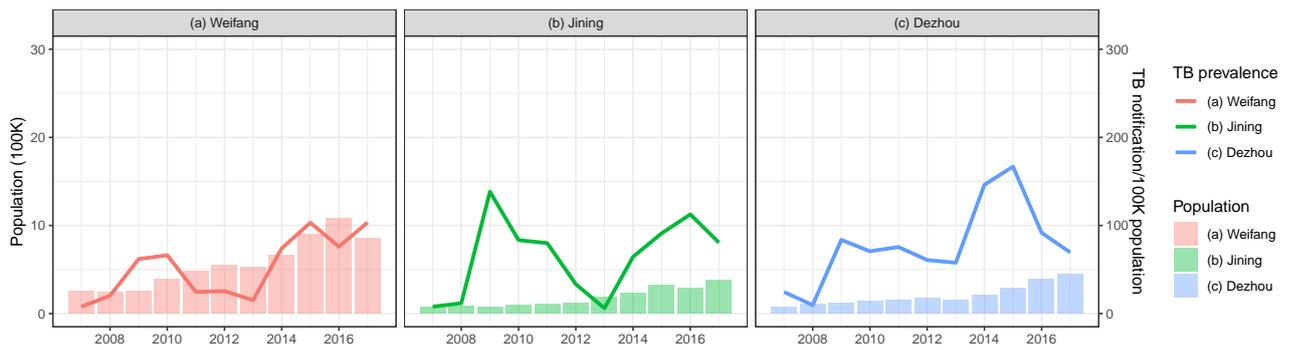


Figure A.7: The proportion of immigrant population from 2007 to 2017 in seven cities in Shandong, China.

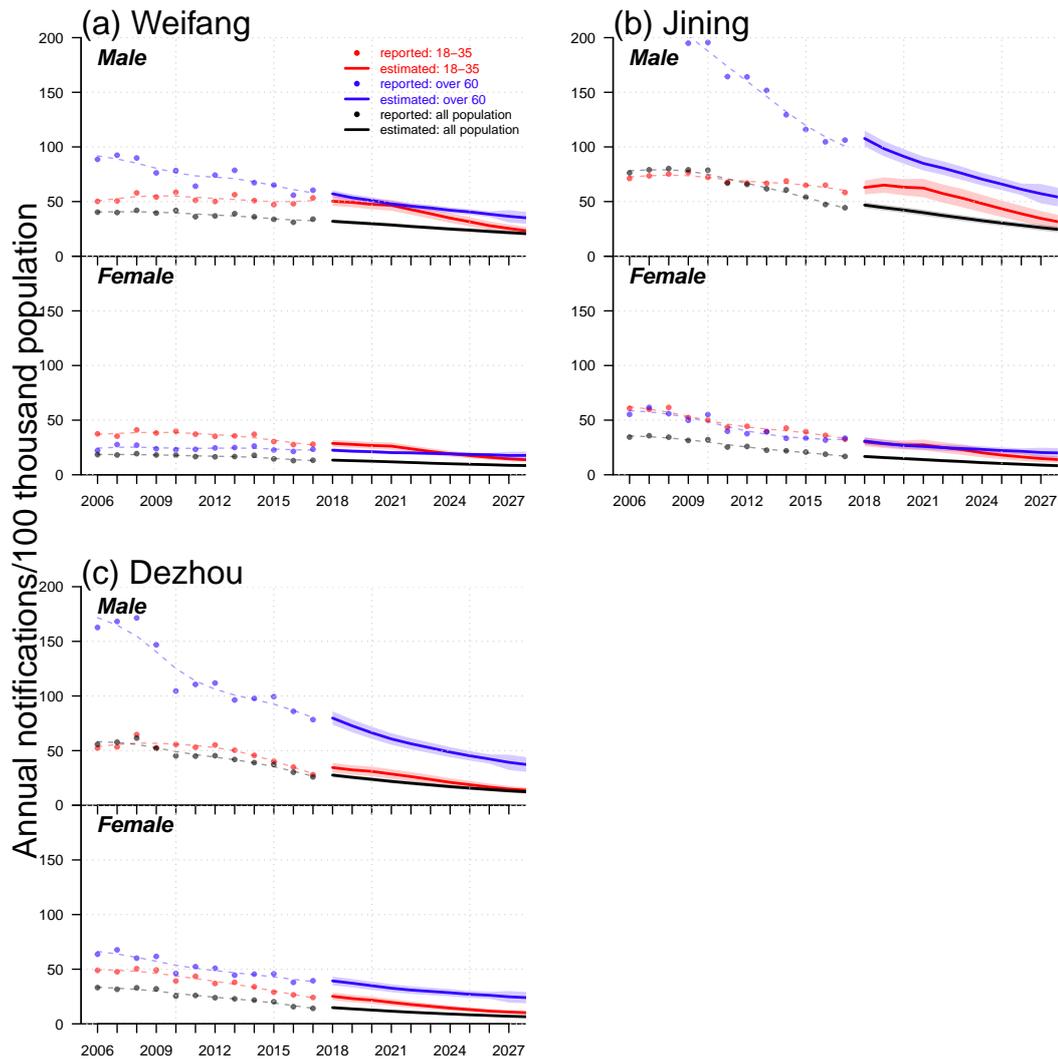


Figure A.8: Trends of TB notification rates from 2006–2017 and forecasts from 2018 to 2027 in (a) Weifang, (b) Jining, and (c) Dezhou. In each panel, (i) upper and low subpanels show trends for males and females, respectively; (ii) dots, lines, and shades colored in red, blue, and black indicate reported trends, forecast trends, and 95% confidential intervals for forecasts of annual TB notifications (in the 100 thousand population) for populations aged 18–35 and over 60 years and the total population, respectively; (iii) red, blue, and black dashed lines are smoothed splines that indicate the trends of annual reported TB notifications for the populations aged 18–35 and over 60 years and the total population, respectively.

A.2.3 Further investigations in heterogeneities across seven cities.

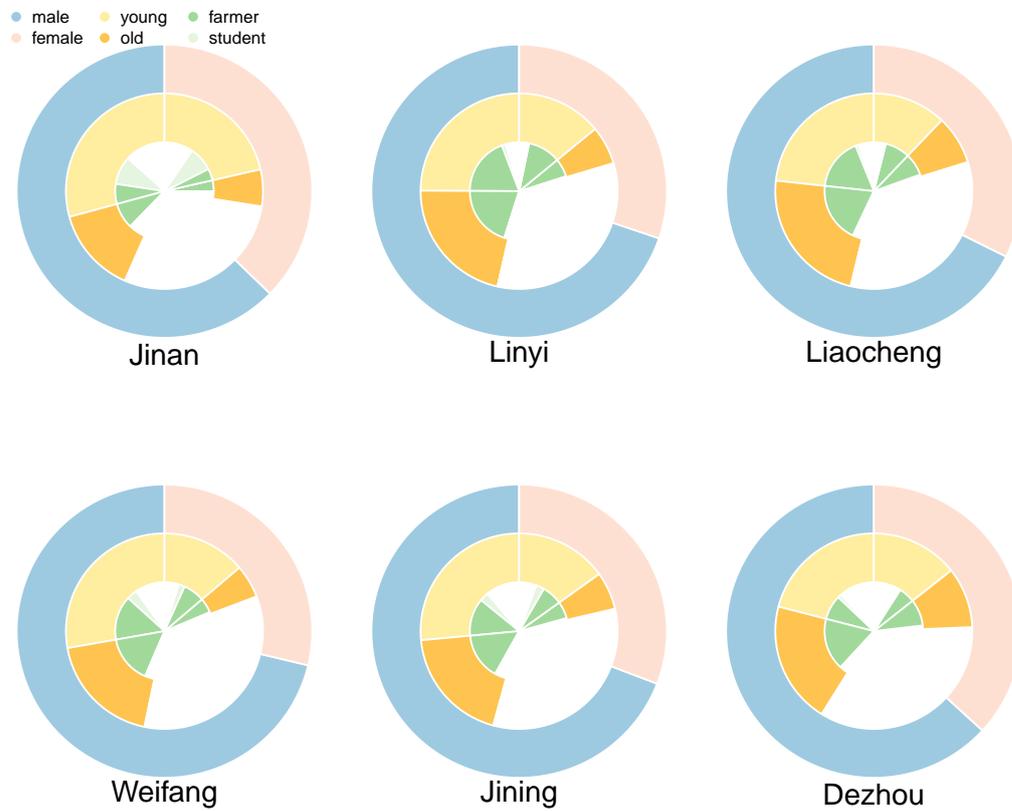


Figure A.9: The components of migrant TB notifications.



Figure A.10: The ratio of the elderly over the young from 2005 to 2017 in seven cities in Shandong, China.

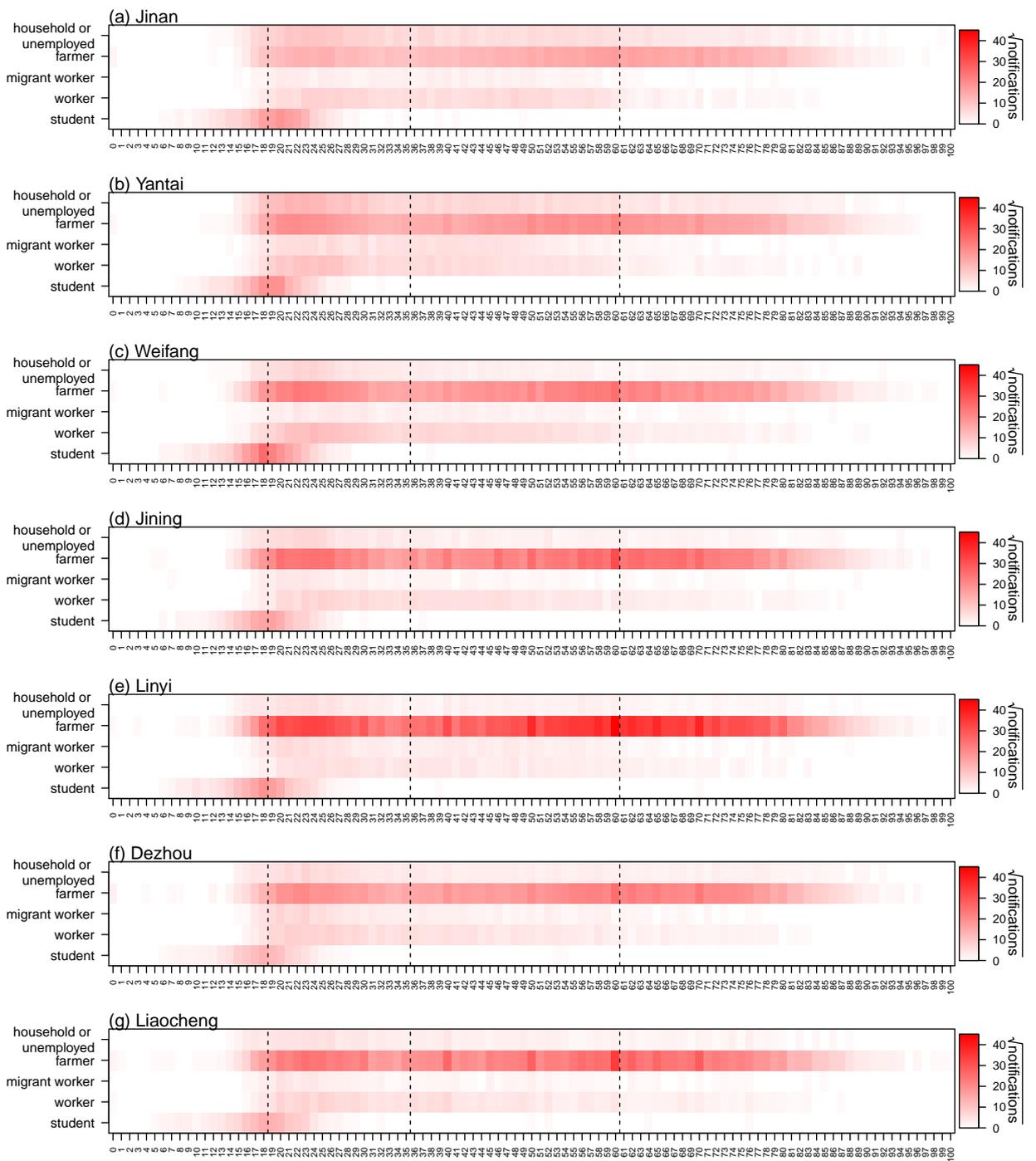


Figure A.11: TB notifications against age from 2006 to 2017 for top five occupations in seven cities, from 2006 to 2017: student, worker, migrant worker, farmer and household or unemployed. Dashed-black lines divide the heatmaps into four sub-age groups: 0–17, 18–35, 35–60, and over 60.

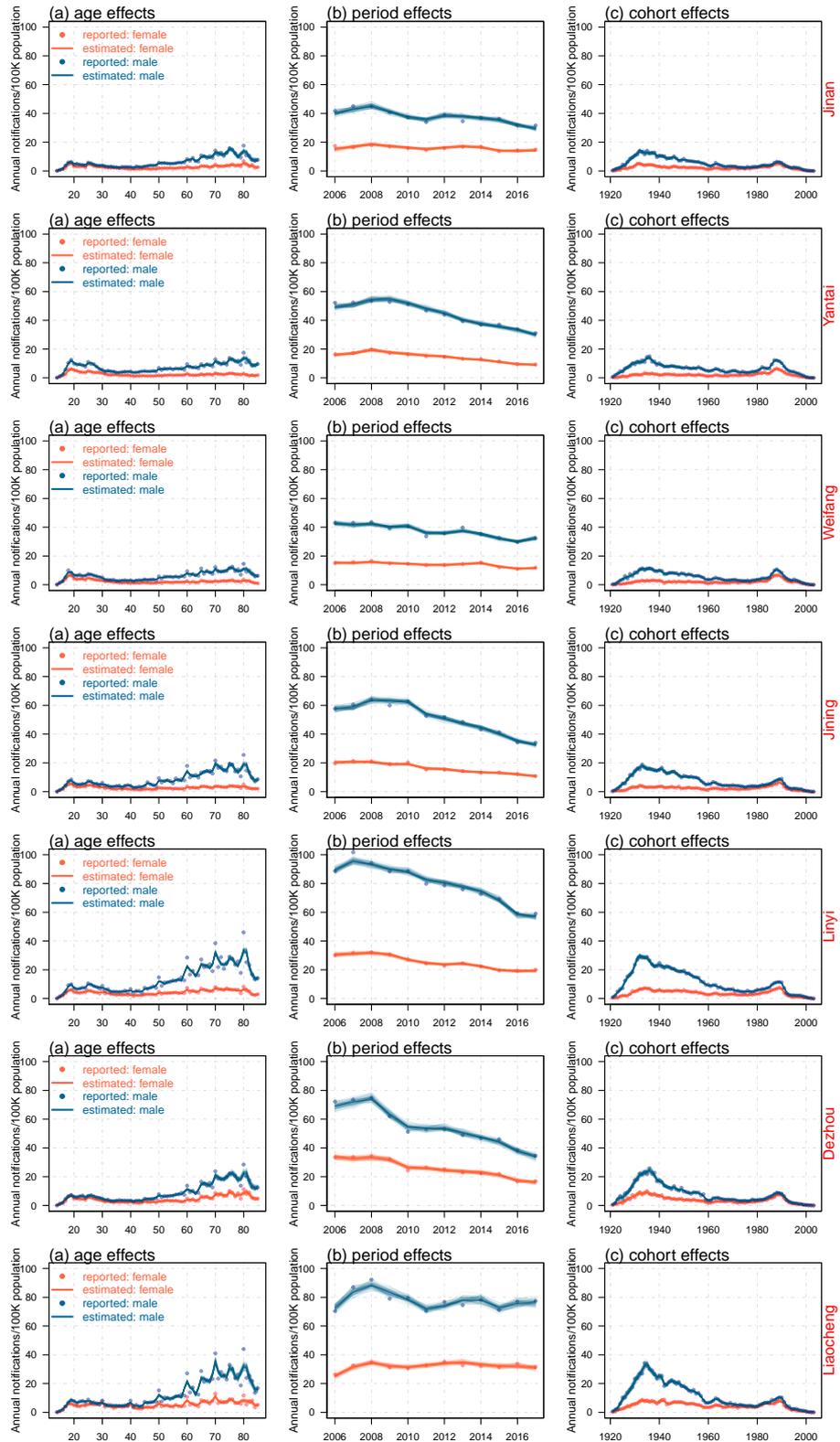


Figure A.12: Estimated effects of age, calendar period and birth cohort on TB notification rate from 2006 to 2017 in seven cities in Shandong, China. From top to bottom, each row show (a) age effects, (b) period effects, and (c) cohort effects from left to right, for each city (*i.e.*, Jinan, Yantai, Weifang, Jining, Linyi, and Dezhou), respectively. Within each panels, dots, lines, and transparent lines colored in blue/red show reported effects, estimated effects, and bootstrap simulations for estimated effects on annual TB notifications (per 100K population), for male/female, respectively.

A.2.4 Description of DR-TB notification in Shandong

In Fig A.13, we present total laboratory-confirmed records and three types of DR-TB notifications from 2004 to 2017 in Shandong. The collected notifications increased from 2004 to 2007 and fell afterwards, until 2010. Sudden rise occurred in 2011 and then the notifications gradually decreased. The general trends of laboratory notifications for DR-TB from 2004 to 2014 were similar to that of total records, including the sharp jump in 2011. However, after 2014, DR-TB notifications increased slowly instead of dropping down.

We present the estimated effects of age, period and cohort for DR-TB tested notification rate in Fig A.15. In general, the estimated effects for MDR notifications are nearly identical to those for RFP-resistant notifications, as MDR is the intersection of INH-resistant and RFP-resistant notifications. While, all of the three effects for INH-resistant notifications show higher resistant risk than MDR and INH. Among these three intrinsic factor, period effects play the most essential role in DR-TB notifications. The period effects of three types of DR-TB are increasing in a main through years, though two temporary falls occurred in 2006–07 and 2014–15 in INH-resistant notifications. With continuous fluctuations through the whole range, age effects and cohort effects are mostly stable and low, indicating little difference between ages and birth cohorts on contributing the resistant risk.

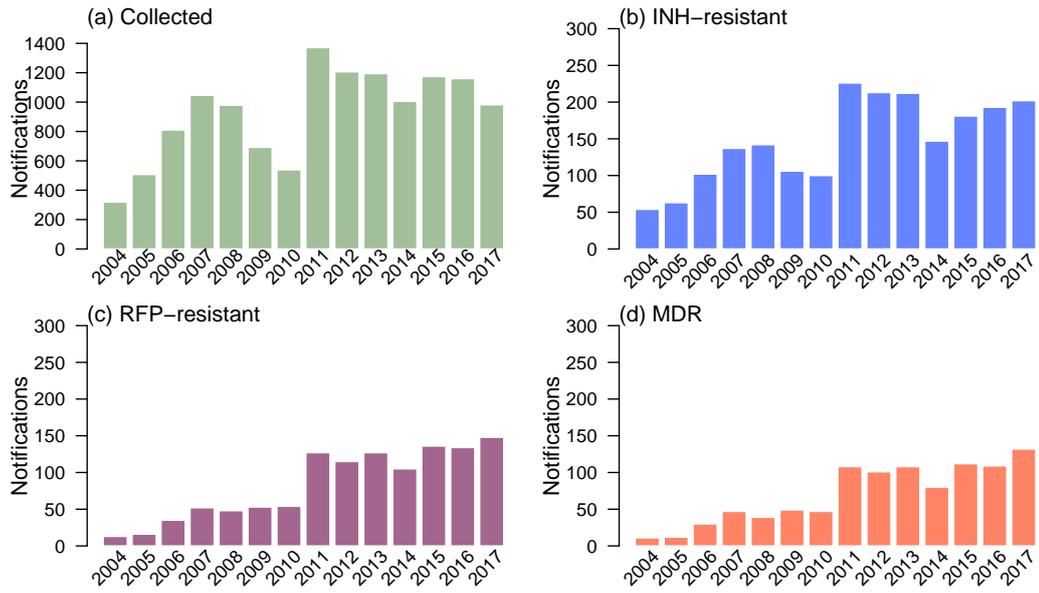


Figure A.13: Numbers of (a) total collected notifications, (b) INH-resistant notifications, (c) RFP-resistant notifications and (d) MDR notifications from 2004 to 2017 in Shandong. The sudden increasing in 2011 was due to enhanced surveillance.

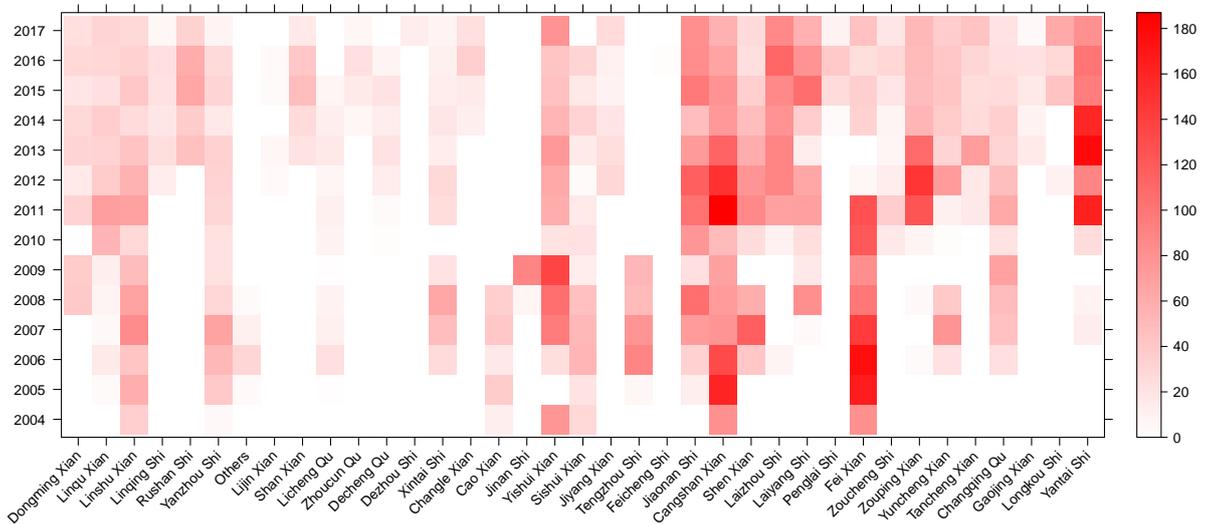


Figure A.14: Number of DR-TB notifications across 38 units in 37 regions in Shandong, from 2004 to 2017.

A.2.5 Estimation and forecast of DR-TB cases

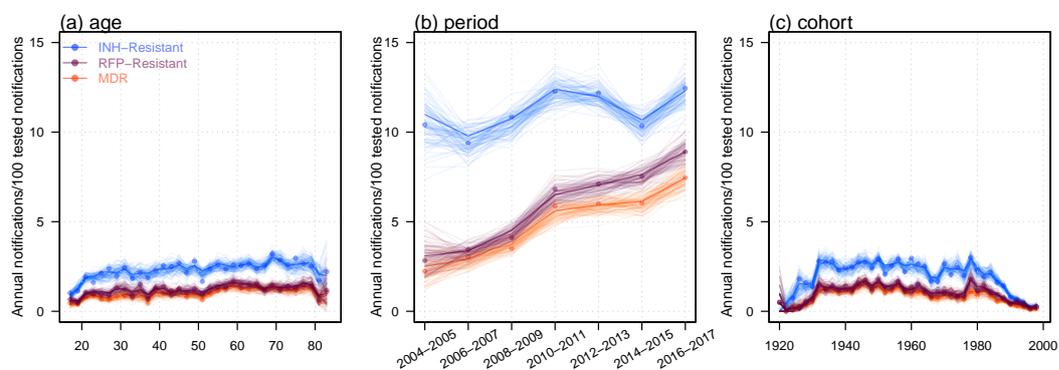


Figure A.15: Estimated effects of (a) age, (b) calendar period and (c) birth cohort on DR-TB tested notification rate from 2004 to 2017 in Shandong, China. Within each panel, dots, lines, and transparent lines colored in blue/brown/orange show reported effects, estimated effects, and bootstrap simulations for estimated TB effects on annual TB notifications (per 100K population), for INH-resistant TB/RFP-resistant TB/MDR-TB, respectively.

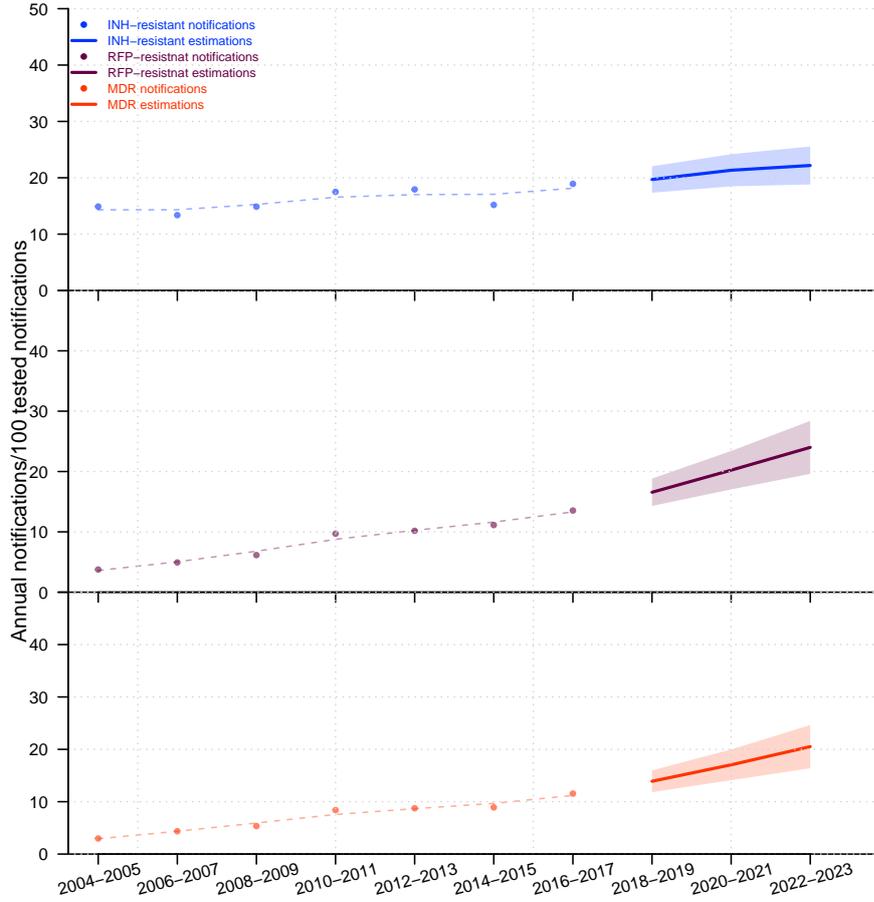


Figure A.16: Reported trends from 2004 to 2017 and forecasts from 2018 to 2023 of DR-TB tested notification rate in Shandong, China for INH-resistant TB, RFP-resistant TB, and MDR-TB, from top to bottom, respectively. Dots, lines, and shaded area colored in blue/brown/orange show reported tested notification rates (per 100 tested notifications), estimated tested notification rates, and 95% confidence interval for estimation, for INH-resistant TB/RFP-resistant TB/MDR-TB, respectively. Dashed lines colored in blue/brown/orange are smoothed splines for indicating the trends of reported DR-TB tested notification rate for INH-resistant TB/RFP-resistant TB/MDR-TB, respectively. Future total collected DR-TB records are predicted by APC models.

A.2.6 Drug correlation

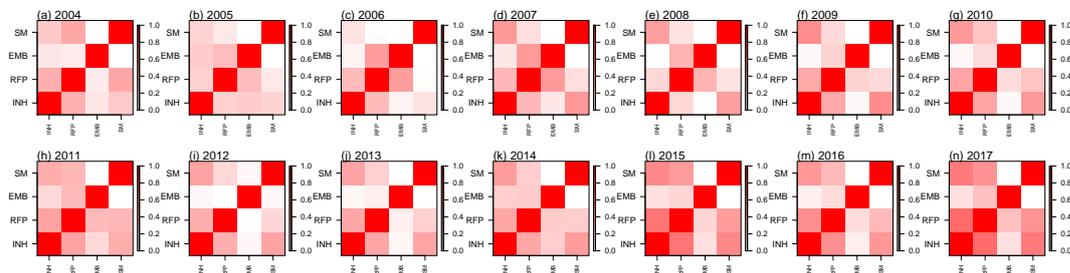


Figure A.17: Correlation between four types of commonly used drugs from 2004 to 2017. Darker color indicates higher correlation.

A.3 MERS-CoV

A.3.1 Assuming Smaller Proportion of Infections Reservoir

In the main text, we assumed the proportion of infections from camel reservoir is 25%, here we changed this ratio to 12%.

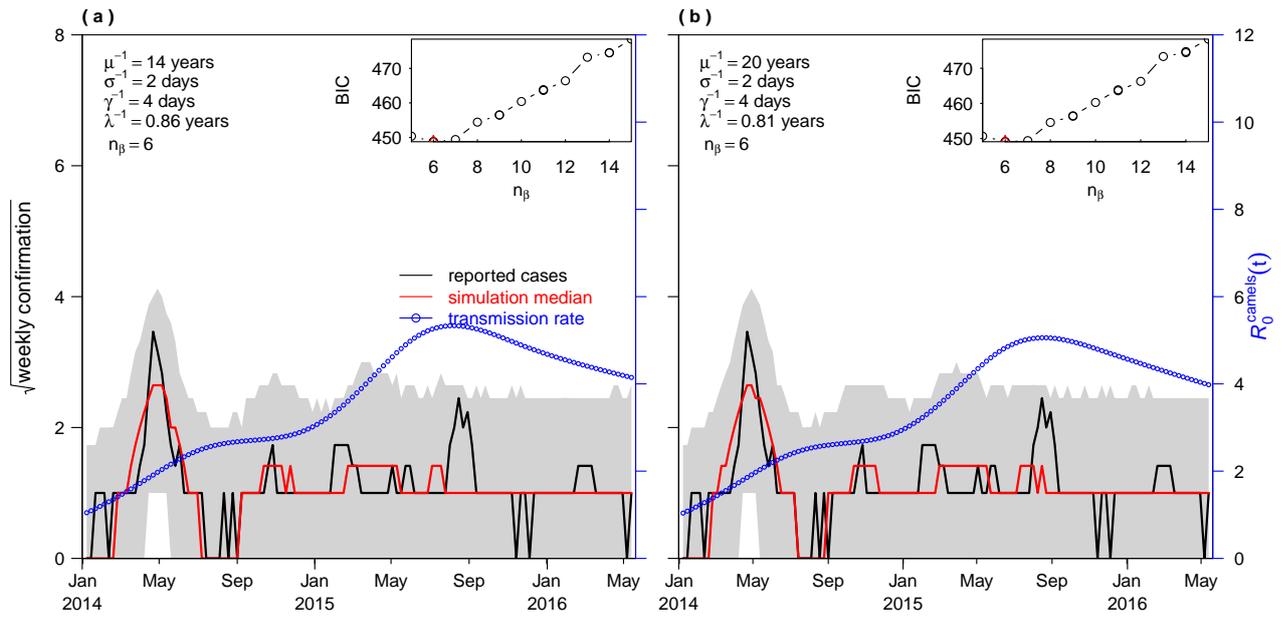


Figure A.18: Model simulations versus observed MERS-CoV spill-over cases. Same model input parameters are used as in Figure 3, except that the proportion of infections from camels is 12%.

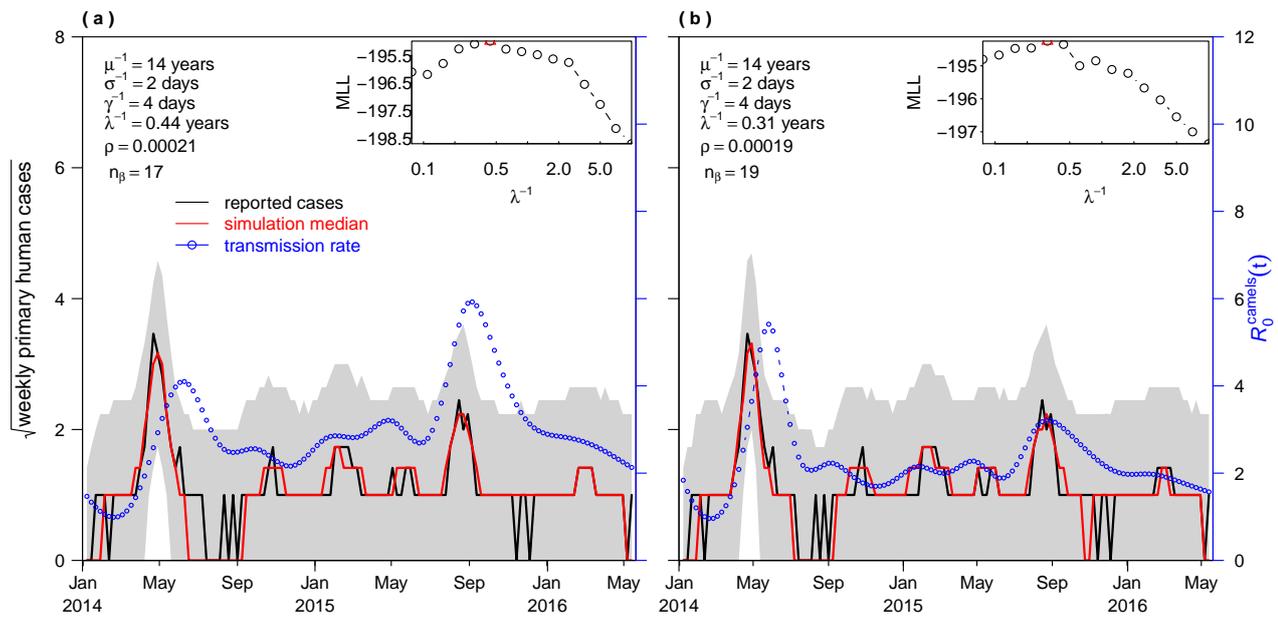


Figure A.19: Model simulations versus observed MERS-CoV spill-over cases. Same model input parameters are used as in Figure 4, except that the proportion of infections from camels is 12%.

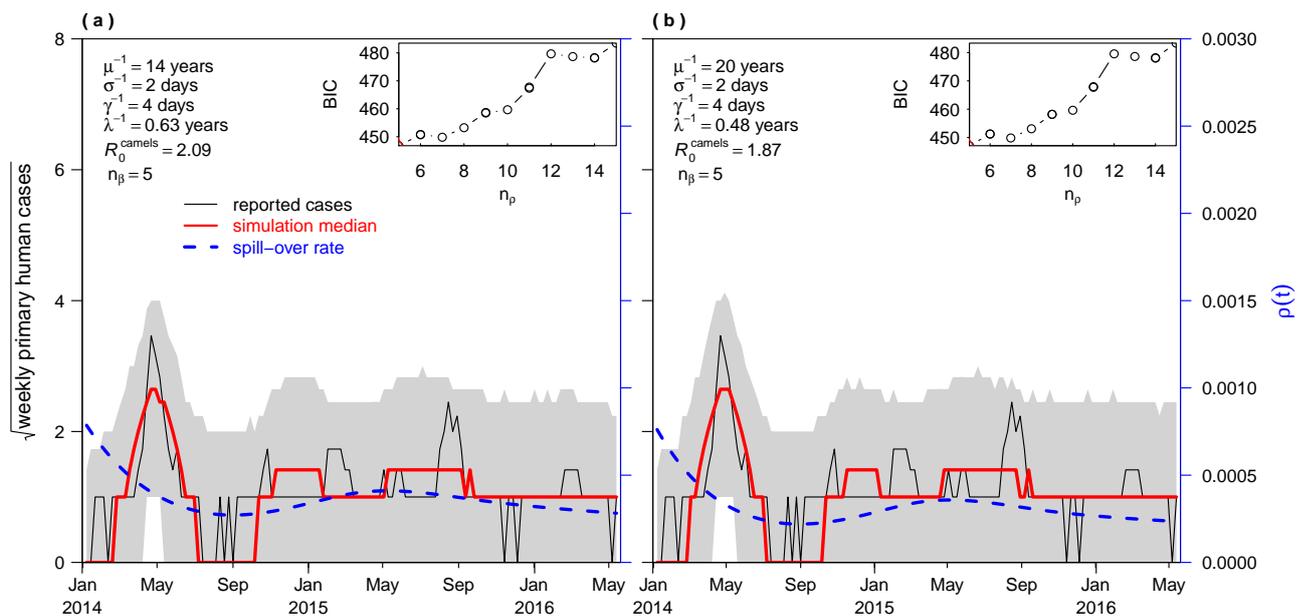


Figure A.20: Model simulations versus observed MERS-CoV spill-over cases. Same model input parameters are used as in Figure 5, except that the proportion of infections from camels is 12%.

Table A.2: Summary of parameter settings and estimates in Figure S2-S4. Same model input parameters are used as in Table 1, except that the proportion of infections from camels is 12%.

| Parameter | Fig.S2a | Fig.S2b | Fig.S3a | Fig.S3b | Fig.S4a | Fig.S4b |
|---------------------------------|----------|----------|----------|----------|----------|----------|
| μ^{-1} (years) | 14 | 20 | 14 | 14 | 14 | 20 |
| λ^{-1} (years) | 0.86 | 0.81 | 0.44 | 0.31 | 0.63 | 0.48 |
| n_{β} | 6 | 6 | 17 | 19 | 5 | 5 |
| $\langle \beta \rangle$ | 327.95 | 315.73 | 268.27 | 206.04 | 191.17 | 170.27 |
| $\mathcal{R}_0^{\text{camels}}$ | 3.594 | 3.46 | 2.94 | 2.258 | 2.095 | 1.866 |
| $\langle \rho \rangle$ | 0.000304 | 3e-04 | 0.00021 | 0.000186 | 0.000367 | 0.000315 |
| τ | 0.00972 | 0.017374 | 0.002918 | 0.005305 | 0.00765 | 0.009408 |
| $S.0$ | 0.969639 | 0.988335 | 0.987012 | 0.99179 | 0.479727 | 0.554365 |
| $E.0$ | 0.002068 | 0.002033 | 0.001639 | 0.000917 | 0.000529 | 0.000437 |
| $I.0$ | 0.002068 | 0.002033 | 0.001639 | 0.000917 | 0.000529 | 0.000437 |
| $R.0$ | 0.026225 | 0.0076 | 0.009711 | 0.006376 | 0.519215 | 0.444761 |
| MLL | -202.7 | -202.86 | -194.98 | -194.25 | -204.12 | -204.28 |
| BIC | 448.78 | 449.11 | 481.54 | 489.73 | 446.8 | 447.12 |

A.3.2 Extension to a Two-Host Model

Besides the model for camels in the main text, we include the following two equations for human

$$\dot{I}_{h1} = \rho\gamma I - \mu_h I_{h1} \quad (\text{A.1a})$$

$$\dot{I}_{h2} = r(I_{h1} + I_{h2}) - \mu_h I_{h2} \quad (\text{A.1b})$$

where I denotes number of infectious camels, I_{h1} denotes primary human cases, and I_{h2} denotes secondary human cases, and r denotes the ‘transmission rate’ among human. We assumed the susceptible pool of human is very large and remains constant (thus synthesised in r). There could be other ways to formulate this model. Alternatively, we could define the situation as in a hospital or in households, where the susceptibles are very limited, $\mu_h^{-1} = 7$ days. Since we know that eventually the proportion of primary cases I_{h1} is only 25% (or 12%) among all human cases. Thus we would expect to achieve at equilibrium (when $\dot{I}_{h2} = 0$) of the second equation, $I_{h1}^e = \frac{1}{3}I_{h2}^e$ (or $I_{h1}^e = \frac{3}{22}I_{h2}^e$). This leads to $r = \frac{3}{4}\mu_h$ (or $r = \frac{88}{100}\mu_h$). We assumed all I_{h1} and I_{h2} are reported.

Then we fit the whole model to all weekly human cases with the moderate primary case ratio 25% as illustration. The results are shown in Figure A.21. The estimated ρ is 0.0005 (95% CI 0.0004, 0.0006) in both cases of $\mu^{-1} = 13$ years and $\mu^{-1} = 20$ years. The estimated λ^{-1} is 1.25 year and 1.2 years, respectively.

This model could be more useful if we have data for separated time series of primary and secondary human cases.

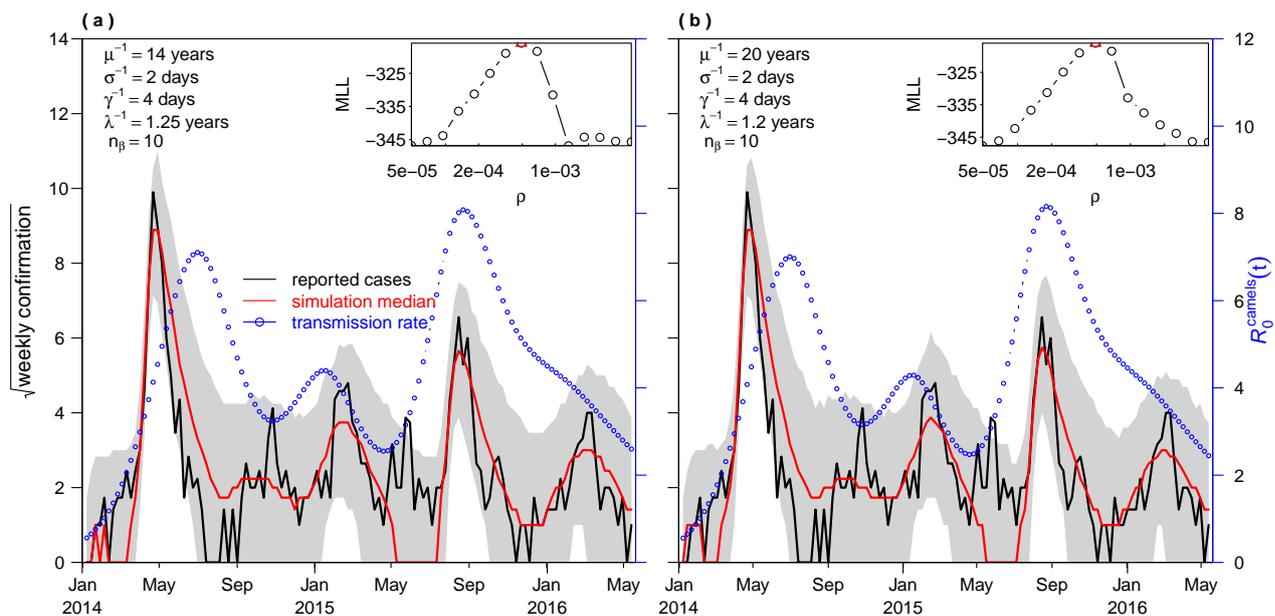


Figure A.21: Two-host model simulations versus observed MERS-CoV human cases. Thin black curve represents reported spill-over cases, bold red curve represents model generated median of 1000 simulations, shaded regions represent the 95% range of simulations, and blue curve (with circles) represents the transmission rate (in units of \mathcal{R}_0). Inset panels show the profile of MLL as a function of the spill-over rate. $\mu_h^{-1} = 7$ days, and $r = \frac{3}{4}\mu_h$.

A.3.3 Climatic Factors

Wavelet analyses are conducted via R package “WaveletComp” (<http://hs-stat.com/WaveletComp/>) for exploring the relationships among weekly MERS-CoV human cases and climatic data for Riyadh Saudi Arabia were obtained from Weather Underground Organization (<https://www.wunderground.com>).

Figures A.22 and A.23 show the results of wavelet coherency analysis for weekly human MERS-CoV cases versus temperature and absolute humidity from June 13, 2012 to October 19, 2016 in Saudi Arabia. Black-dashed line in Figure A.22(c) suggests there is a possible association between temperature and MERS-CoV cases,

where significant synchronization are shown from mid-2013 to mid-2014 and from mid-2015 to early 2016. Synchrony pattern is less evident in MERS-CoV versus absolute humidity. With one-year wavelet period, absolute humidity leads MERS-CoV human cases for about 3 months.

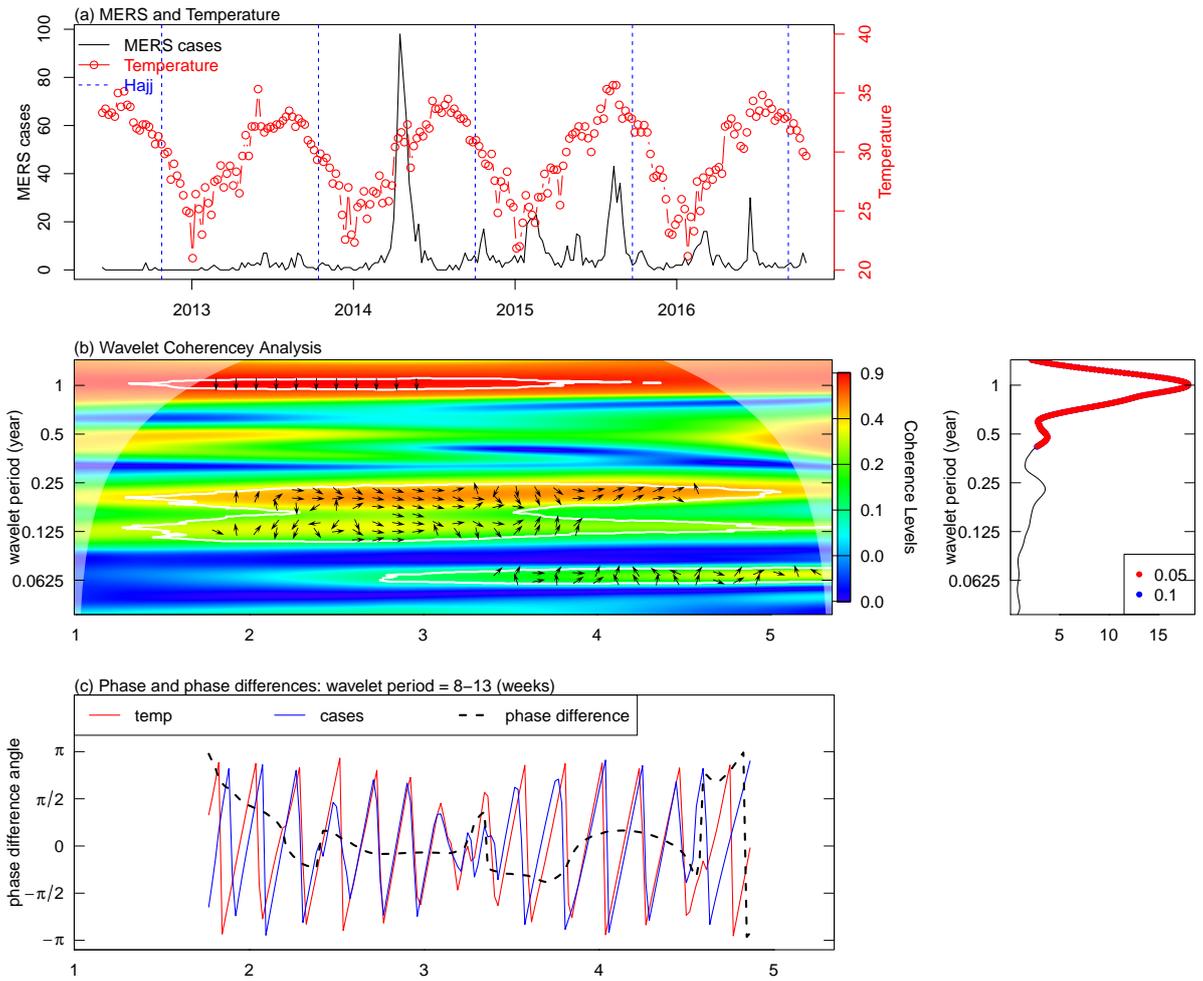


Figure A.22: Wavelet Analysis for MERS-CoV cases and temperature in Saudi Arabia. (a) Time series of weekly reported MERS-CoV cases and weekly average temperature in Saudi Arabia. (b) Left panel is the power spectrum of wavelet coherency analysis for reported MERS-CoV cases and temperature, where red color represents high power and blue for low power; the angle of the arrows indicates the phase and those amidst white line indicates significant phase difference at 5% confidence level; white shades indicates boundary effects. Right panel shows the cross-wavelet average power for different wavelet periods and red dots and blue dots show significance at 5% and 10% levels respectively. (c) Phase and phase difference at wavelet period of 8 to 13 weeks (i.e. 0.15 to 0.25 year); red line and blue line represent temperature and MERS-CoV human cases; black-dashed line indicates the phase difference.

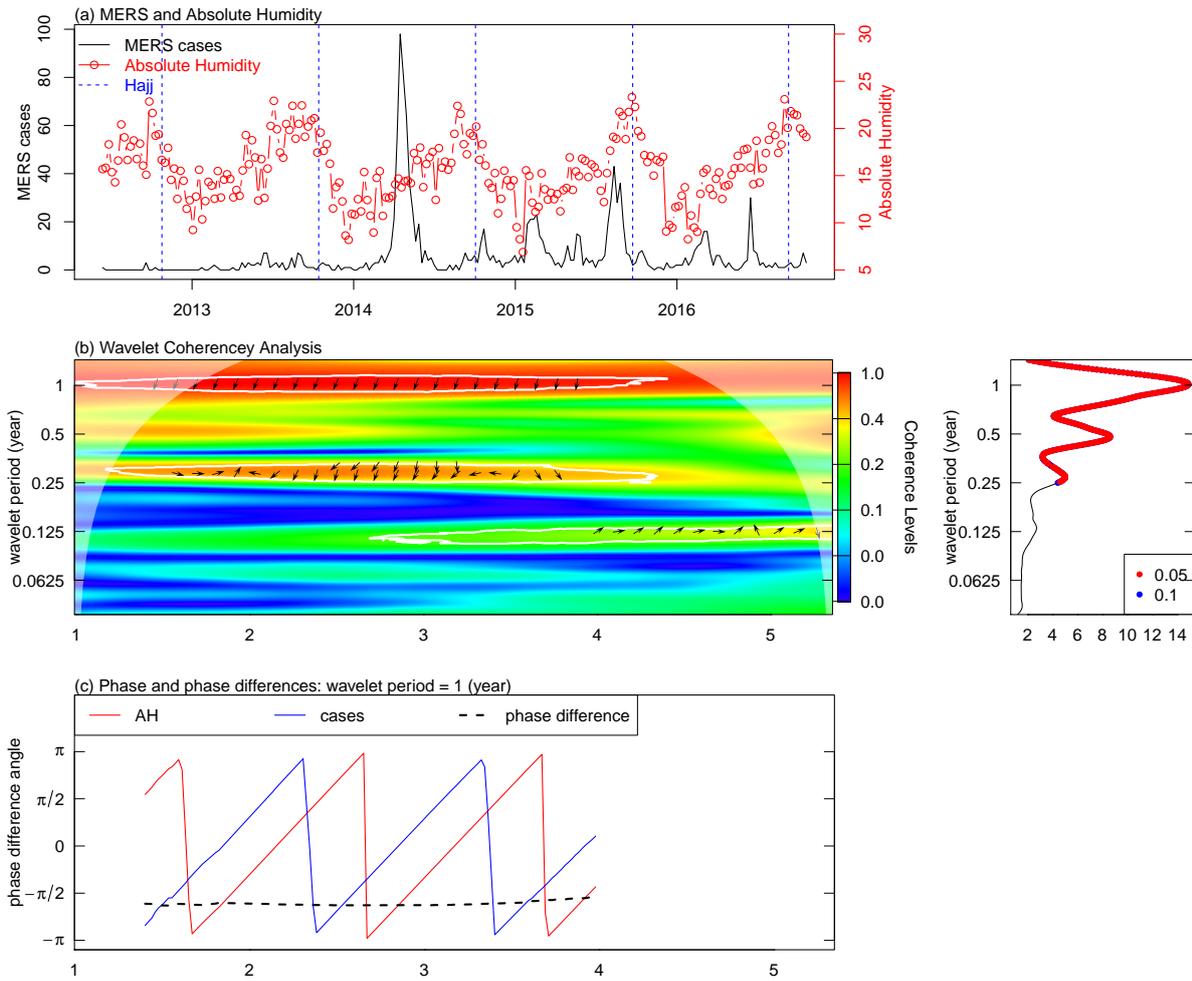


Figure A.23: Wavelet Analysis for MERS-CoV cases and absolute humidity in Saudi Arabia. (a) Time series of weekly reported MERS-CoV cases and weekly average absolute humidity in Saudi Arabia. (b) Left panel is the power spectrum of wavelet coherence analysis for reported MERS-CoV cases and absolute humidity, where red color represents high power and blue for low power; the angle of the arrows indicates the phase and those amidst white line indicates significant phase difference at 5% confidence level; white shades indicates boundary effects. Right panel shows the cross-wavelet average power for different wavelet periods and red dots and blue dots show significance at 5% and 10% levels respectively. (c) Phase and phase difference at wavelet period of 1 year; red line and blue line represent absolute humidity and MERS-CoV human cases; black-dashed line indicates the phase difference.

A.3.4 Equilibrium

For Equation 4.1, we define initial condition $S(0) = S_0 > 0, E(0) = E_0 > 0, I(0) = I_0 > 0$, and $R(0) = R_0 > 0$.

Consider the following equation for the rate of change in total population of camels:

$$N'(t) = \frac{dN}{dt} = \mu N - \mu(S + E + I + R) \geq 0 \quad (\text{A.2})$$

For the region $\Omega = \{(S, E, I, R) \in \mathbb{R}_+^4 : N = S + E + I + R\}$, it can be shown by solving N that all solutions of the system starting in the Ω will remain in Ω for all time t with $t \geq 0$. Thus, the region Ω is positively-invariant and it is sufficient to consider solutions restricted in Ω . In this region, the usual existence, uniqueness, and continuation results hold for the system described in Equation 4.1 (See Hussaini et al. (2017)).

The disease-free equilibrium (DFE) of the model 4.1 obtained at steady state is given by

$$E^0 = (S^0, E^0, I^0, R^0) = (N, 0, 0, 0).$$

Using the next generation operator method (See Van-den Driessche and Watmough (2002)), we obtained associated reproduction number of the model 4.1, denoted by $\mathcal{R}_0 = \rho(FV^{-1})$, where ρ represents the function to find the spectral radius of the next generation matrix, $\mathbf{G} = FV^{-1}$. The matrix F (for the new infection terms) and V (for the remaining transition terms), associated with the model 4.1, are given by

$$F = \begin{bmatrix} 0 & \beta N \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} q_1 & 0 \\ -\sigma & q_2 \end{bmatrix} \quad (\text{A.3})$$

where $q_1 = \sigma + \mu$, $q_2 = \gamma + \mu$ and $q_3 = \lambda + \mu$. Therefore, the basic reproduction number, \mathcal{R}_0 , is given by

$$\mathcal{R}_0 = \frac{\sigma\beta N}{q_1 q_2} \quad (\text{A.4})$$

Theorem A.3.1 below follows the Theorem 2 of Van-den Driessche and Watmough (2002).

Theorem A.3.1. *The DFE, Ω_0 , of the model 4.1, is locally-asymptotically stable (LAS) in Ω if $\mathcal{R}_0 < 1$, but unstable if $\mathcal{R}_0 > 1$.*

The following theorem is established.

Theorem A.3.2. *The DFE, Ω_0 , of the model 4.1, is globally-asymptotically stable (GAS) in Ω if $\mathcal{R}_0 \leq 1$.*

Proof. Let $\mathcal{R}_0 \leq 1$. Consider the linear Lyapunov function given by:

$$U = a_0 E + a_1 I, \quad (\text{A.5})$$

where $a_0 = \frac{\mathcal{R}_0}{N}$ and $a_1 = \frac{\beta}{q_2}$, with $\mathcal{R}_0 = \frac{\sigma\beta N}{q_1 q_2}$.

The derivative of the Lyapunov function $U(t)$ is given by

$$\begin{aligned} \dot{U} &= a_0 \dot{E} + a_1 \dot{I} \\ &= a_0 (\beta S I - q_1 E) + a_1 (\sigma E - q_2 I) \\ &= (a_1 \sigma - a_0 q_1) E + (a_0 \beta S - a_1 q_2) I \\ &= \left(\frac{\sigma\beta}{q_2} - \frac{q_1 \mathcal{R}_0}{N} \right) E + \left(\frac{\beta \mathcal{R}_0 S}{N} - \beta \right) I \\ &= \beta \left(\frac{S}{N} \mathcal{R}_0 - 1 \right) I \\ &\leq \beta (\mathcal{R}_0 - 1) I \end{aligned} \quad (\text{A.6})$$

If $\mathcal{R}_0 \leq 1$, then $\dot{U} \leq 0$, which implies that the DFE E^0 of the system 4.1 is GAS. Furthermore, \dot{U} is strictly equal to zero if and only if $S = S_0$ and $E = I = R = 0$. The maximum invariant set in $\{(S, E, I, R) \in \Omega : \dot{U} = 0\}$ is the singleton E^0 . Therefore, by LaSalle's Invariant Principle (See LaSalle (1976)), every solution that starts in Ω approaches E^0 as $t \rightarrow \infty$. ■

The endemic equilibrium (EE) for system 4.1, $E^1 = (S^*, E^*, I^*, R^*)$, is given by

$$\begin{aligned} S^* &= \frac{N}{\mathcal{R}_0} \\ E^* &= \frac{q_1 q_2^2 q_3 (\mathcal{R}_0 - 1)}{\sigma \beta (q_1 q_2 q_3 - \sigma \gamma \lambda)} \\ I^* &= \frac{q_1 q_2 q_3 \mu (\mathcal{R}_0 - 1)}{\beta (q_1 q_2 q_3 - \sigma \gamma \lambda)} \\ R^* &= \frac{q_1 q_2 \gamma \mu (\mathcal{R}_0 - 1)}{\beta (q_1 q_2 q_3 - \sigma \gamma \lambda)} \end{aligned} \tag{A.7}$$

Hence, the following Theorem A.3.3 is claimed.

Theorem A.3.3. *The endemic equilibrium (EE), E^1 , is globally-asymptotically stable (GAS) in Ω whenever $\mathcal{R}_0 > 1$.*

Proof. Define a Lyapunov function as follow

$$V = b_0 \left(S - S^* - S^* \ln \frac{S}{S^*} \right) + b_1 \left(E - E^* - E^* \ln \frac{E}{E^*} \right) + b_2 \left(I - I^* - I^* \ln \frac{I}{I^*} \right) \tag{A.8}$$

where $b_0 = b_1 = 1, b_2 = \frac{\beta S^* I^*}{\sigma E^*}$.

Thus the derivative is given by

$$\begin{aligned}
\dot{V} &= \left(1 - \frac{S^*}{S}\right) \dot{S} + \left(1 - \frac{E^*}{E}\right) \dot{E} + \frac{\beta S^* I^*}{\sigma E^*} \left(1 - \frac{I^*}{I}\right) \dot{I} \\
&= \beta S^* I^* \left(1 - \frac{S^*}{S}\right) \left(1 - \frac{SI}{S^* I^*}\right) - \mu \frac{(S^* - S)^2}{S} \\
&\quad + \beta S^* I^* \left(\frac{SI}{S^* I^*} - \frac{E}{E^*} - \frac{SIE^*}{S^* I^* E} + 1\right) + \beta S^* I^* \left(\frac{E}{E^*} - \frac{I}{I^*} - \frac{EI^*}{E^* I} + 1\right) \\
&\leq \beta S^* I^* \left(3 - \frac{S^*}{S} - \frac{SIE^*}{S^* I^* E} - \frac{EI^*}{E^* I}\right)
\end{aligned} \tag{A.9}$$

Following Yang et al. (2017) and Sun et al. (2017), for the function $v(x) = 1 - x - \ln x$, if $x > 0$, it leads to $v(x) \leq 0$. And if $x = 1$, we have $v(x) = 0$. Thus, $x - 1 \geq \ln(x)$ for any $x > 0$.

Therefore, for the derivative of the Lyapunov function, we have

$$\begin{aligned}
\dot{V} &\leq \beta S^* I^* \left[-\left(\frac{S^*}{S} - 1\right) - \left(\frac{SIE^*}{S^* I^* E} - 1\right) - \left(\frac{EI^*}{E^* I} - 1\right) \right] \\
&\leq \beta S^* I^* \left[-\ln\left(\frac{S^*}{S}\right) - \ln\left(\frac{SIE^*}{S^* I^* E}\right) - \ln\left(\frac{EI^*}{E^* I}\right) \right] \\
&= 0
\end{aligned} \tag{A.10}$$

Furthermore, the equality $\dot{V} = 0$ holds if and only if $S = S^*, E = E^*, I = I^*$ and $R = R^*$. Thus, the endemic equilibrium state E^1 is the only positive invariance set to the system 4.1 contained entirely in $\{(S, E, I, R) \in \Omega : S = S^*, E = E^*, I = I^*, R = R^*\}$. Therefore, it follows from the LaSalle's invariance principle (See LaSalle (1976)) that every solution to the equations in A.9 with initial conditions in Ω converges to stable endemic equilibrium point, E^1 , as $t \rightarrow \infty$. Hence, the positive endemic equilibrium is globally asymptotically stable whenever $\mathcal{R}_0 > 1$.

■

For two-host model demonstrated in equation system A.1, with the initial condition $S(0) = S_0 > 0, E(0) = E_0 > 0, I(0) = I_0 > 0, R(0) = R_0 > 0, I_{H1}(0) > 0$ and $I_{h2}(0) > 0$, and all non-negative parameters and state variable, we have the biologically feasible region as $\Omega_e = \{(S, E, I, R, I_{h1}, I_{h2}) \in \mathbb{R}_+^6 : N_e = S + E + I + R + I_{h1} + I_{h2}\}$. In this region, the usual existence, uniqueness, and continuation results hold for the two-host system A.1.

The disease-free equilibrium (DFE) for the two-host model A.1 obtained at steady state is given by

$$E^0 = (S^0, E^0, I^0, R^0, I_{h1}^0, I_{h2}^0) = (N, 0, 0, 0, 0, 0)$$

Therefore, the basic reproduction number of the human only model is $\mathcal{R}_h = \frac{r}{\mu}$.

Similarly, the theorem A.3.4 below follows the Theorem 2 of Van-den Driessche and Watmough (2002).

Theorem A.3.4. *The DFE of the two-host model A.1, is locally-asymptotically stable (LAS) in Ω_e if $\mathcal{R}_0 \leq 1$, but unstable if $\mathcal{R}_0 > 1$.*

The following theorem is established and the proof is similar to that of model 4.1, demonstrated above.

Theorem A.3.5. *The DFE of the two-host model A.1, is globally-asymptotically stable (GAS) in Ω_e if $\mathcal{R}_0 \leq 1$.*

The endemic equilibrium (EE) for two-host model A.1, $E^1 = (S^*, E^*, I^*, R^*, I_{h1}^*, I_{h2}^*)$ is given by

$$\begin{aligned}
S^* &= \frac{N}{\mathcal{R}_0} \\
E^* &= \frac{q_1 q_2^2 q_3 (\mathcal{R}_0 - 1)}{\sigma \beta (q_1 q_2 q_3 - \sigma \gamma \lambda)} \\
I^* &= \frac{q_1 q_2 q_3 \mu (\mathcal{R}_0 - 1)}{\beta (q_1 q_2 q_3 - \sigma \gamma \lambda)} \\
R^* &= \frac{q_1 q_2 \gamma \mu (\mathcal{R}_0 - 1)}{\beta (q_1 q_2 q_3 - \sigma \gamma \lambda)} \\
I_{h1}^* &= \frac{\rho \gamma \mu q_1 q_2 q_3 (\mathcal{R}_0 - 1)}{\beta \mu_h (q_1 q_2 q_3 - \sigma \gamma \lambda)} \\
I_{h2}^* &= \frac{r \rho \gamma \mu q_1 q_2 q_3 (\mathcal{R}_0 - 1)}{\beta \mu_h (\mu_h - r) (q_1 q_2 q_3 - \sigma \gamma \lambda)}.
\end{aligned} \tag{A.11}$$

Hence, the following Theorem A.3.6 is claimed and the proof is similar to that of model 4.1.

Theorem A.3.6. *The endemic equilibrium (EE), E^1 , is globally-asymptotically stable (GAS) in Ω_e whenever $\mathcal{R}_0 > 1$.*

A.4 Phylodynamics

A.4.1 Implementation of the model

```

library(pomp2)
library(tidyverse)
library(ggplot2)
library(foreach)
library(doParallel)
registerDoParallel()

```

```

Csnippet("
    int nrate = 6;
    double rate[nrate], trans[nrate];
    double beta;
    double dW = rgammawn(F,dt);
    double combfact;

    beta = R0*(gamma + mu)*(1 + alpha*cos(M_2PI*t/12));

    rate[0] = mu * P;
    rate[1] = beta * I/P * dW/dt;
    rate[2] = mu;
    rate[3] = gamma;
    rate[4] = mu;
    rate[5] = mu;

    trans[0] = rpois(rate[0]*dt);
    reulermultinom(2, S, &rate[1], dt, &trans[1]);
    reulermultinom(2, I, &rate[3], dt, &trans[3]);
    reulermultinom(1, R, &rate[5], dt, &trans[5]);

    S += trans[0] - trans[1] - trans[2];
    I += trans[1] - trans[3] - trans[4];
    R += trans[3] - trans[5];
    H += trans[1];

    // M is the indicator covariate:
    // M = 1 in the first interval following an epidemiological record;
    // M = 0 in subsequent intervals before the next record.
    if (M > 0) K = H;
    else K += trans[1];

```

```

combfact = i*(i-1)/(I+1e-6)/(I+1e-6);
lambda = combfact*beta*S*I/P;
Lambda += lambda*dt;
") -> rproc

```

```

Csnippet("
  double ll = -Lambda;
  ll += (R_FINITE(cases)) ? dnbinom_mu(cases,1/tau,rho*K,1) : 0;
  ll += (branches > 0) ? branches*log(lambda) : 0;
  lik = (give_log) ? ll : exp(ll);
") -> dmeas

```

```

Csnippet("
  cases = rnbinom_mu(1/tau, rho*K);
") -> rmeas

```

```

Csnippet("
  double m = P/(S_0+I_0+R_0);
  S = nearbyint(S_0*m);
  I = nearbyint(I_0*m);
  R = nearbyint(R_0*m);
  H = 0;
  K = 0;
  Lambda = 0;
  lambda = 0;
") -> rinit

```

```

read_csv("rasmock.csv") -> dat

```

```

dat %>%
  filter(month>0) %>%

```

```

select(month,cases,branches,samples) %>%
pfilter(times="month",t0=0,Np=2000,
        params=c(R0=10,gamma=3,mu=0.0012,F=0.012,
                  S_0=0.1,I_0=357/1e6,R_0=0.9-357/1e6,alpha=0.16,
                  P=1e6,rho=0.43,tau=1/15),
        rinit=rinit,
        rprocess=euler(rproc,delta.t=0.01),
        dmeasure=dmeas,
        rmeasure=rmeas,
        accumvars=c("Lambda","H"),
        statenames=c("S","I","R","H","K","lambda","Lambda"),
        paramnames=c("R0","gamma","mu","alpha","P","rho","tau","F",
                      "S_0","I_0","R_0"),
        partrans=parameter_trans(
          log=c("R0","F","tau"),
          logit=c("rho"),
          barycentric=c("S_0","I_0","R_0"),
          toEst=Csnippet("T_alpha = logit((alpha+1)/2);"),
          fromEst=Csnippet("alpha = 2*expit(T_alpha)-1;")
        ),
        covar=covariate_table(
          dat %>%
            mutate(M=if_else(month==min(month) | !is.na(cases),1,0)) %>%
            select(month,i,M),
          times="month",
          order="constant")
) -> pf

pf %>% as.data.frame() %>%
gather(var,val,-month) %>%

```

```

filter(!is.na(val)) %>%
ggplot(aes(x=month,y=val))+
geom_line()+
facet_wrap(~var,scales="free_y")

#####
# local search
#####

sobolDesign(
  lower=c(R0=0,gamma=0,F=0,alpha=0,rho=0,tau=0,mu=0.0012,P=1e6,
          S_0=0.1,I_0=357/1e6,R_0=0.9-357/1e6),
  upper=c(R0=30,gamma=10,F=1,alpha=1,rho=1,tau=1,mu=0.0012,P=1e6,
          S_0=0.1,I_0=357/1e6,R_0=0.9-357/1e6),
  nseq=200
) -> guesses

pf %>%
mif2(
  Nmif=250,
  params=guesses[1,],
  rw.sd=rw.sd(
    R0=0.02,F=0.02,alpha=0.02,rho=0.02,tau=0.02,gamma=0.02
  ),
  paramnames=c("R0","F", "mu", "P", "S_0","I_0","R_0",
               "alpha", "rho", "tau", "gamma"),
  cooling.type="geometric",cooling.fraction.50=0.9) -> mf1

mf1 %>% traces() %>% melt() %>%
ggplot(aes(x=iteration,y=value))+
geom_line()+

```

```

facet_wrap(~variable,scales="free_y")

#####
#   global search
#####

foreach(guess=iter(guesses,"row"),
  .combine=c, .packages=c("pomp2"),
  .errorhandling="remove", .inorder=FALSE) %dopar% {
  mf1 %>% mif2(params=guess)
} -> mifs

foreach(mf=mifs, .combine=rbind, .packages=c("pomp2"),
  .errorhandling="remove", .inorder=FALSE) %dopar% {
  replicate(5,
    mf %>% pfilter() %>% logLik()
  ) %>%
  logmeanexp(se=TRUE) -> ll
  data.frame(as.list(coef(mf)), loglik=ll[1], loglik.se=ll[2])
} -> estimates

estimates %>%
  filter(!is.na(loglik)) %>%
  filter(loglik > max(loglik)-30) %>%
  select(-loglik, -loglik.se) -> starts

foreach(start=iter(starts,"row"),
  .combine=rbind, .packages=c("pomp2"),
  .errorhandling="remove", .inorder=FALSE) %dopar% {

  mf3 %>%
    mif2(params=start) %>%

```

```
mif2() -> mf

replicate(5,
  mf %>% pfilter() %>% logLik()
) %>%
  logmeanexp(se=TRUE) -> ll
data.frame(as.list(coef(mf)), loglik=ll[1], loglik.se=ll[2])
} %>% rbind(estimates) -> estimates
```


Bibliography

- Alagaili, A. N., Briese, T., Mishra, N., Kapoor, V., Sameroff, S. C., de Wit, E., Munster, V. J., Hensley, L. E., Zalmout, I. S., Kapoor, A., et al. (2014), “Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia,” *MBio*, 5, e00884–14.
- Aldous, D. J. (1999), “The Moran process as a Markov chain on leaf-labeled trees,” unpublished manuscript, dated 29 March 1999.
- Anderson, R. and May, R. (1991), *Infectious diseases of humans: dynamics and control*, Oxford university press, New York.
- Andersson, H. and Britton, T. (2000), *Stochastic epidemic models and their statistical analysis*, Springer New York, New York.
- Andrieu, C., Coucet, A., and Holenstein, R. (2010), “Particle Markov chain Monte Carlo methods,” *J Roy Stat Soc B*, 72, 269–342.
- Arregui, S., Iglesias, M., Samper, S., Marinova, D., Martin, C., Sanz, J., and Moreno, Y. (2018), “Data-driven model for the assessment of Mycobacterium tuberculosis transmission in evolving demographic structures,” *Proc Natl Acad Sci*, 115, e3238–3245.
- Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., and Madani, T. A. (2014), “Evidence for camel-to-human transmission of MERS coronavirus,” *New Engl J Med*, 370, 2499–2505.
- Baum, L. (1972), “An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process,” *Inequalities*, 3, 1–8.
- Baum, L. and Eagon, J. (1967), “An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology,” *Bull Am Math Soc*, 73, 360–363.

- Baum, L. and Petrie, T. (1966), “Statistical Inference for Probabilistic Functions of Finite State Markov Chains,” *Ann Math Stat*, 37, 1554–1563.
- Baum, L. and Sell, G. (1968), “Growth transformations for functions on manifolds,” *Pac J Math*, 27, 211–227.
- Baum, L., Petrie, T., Soules, G., and Weiss, N. (1970), “A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains,” *Ann Math Stat*, 41, 164–171.
- Bedford, T. and Neher, R. (2016), “Seasonal Influenza Circulation Patterns and Projections for 2016-2017,” *BoiRxiv*, p. 076638.
- Breban, R., Riou, J., and Fontanet, A. (2013), “Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk,” *Lancet*, 382, 694–699.
- Breto, C., He, D., Ionides, E., and King, A. (2009), “Time series analysis via mechanistic models,” *Ann Appl Stat*, 3, 319–348.
- Cauchemez, S., Nouvellet, P., Cor, A., Jombart, T., Garske, T., Clapham, H., et al. (2016), “Unraveling the drivers of MERS-CoV transmission,” *Proc Natl Acad Sci U S A*, 113, 9081–9086.
- Center of Autituberculosis in Shandong (2018), “Antituberculosis in Shandong,” <http://www.sdf1.org>. Accessed: 2018-09-03.
- Chan, I., Kaushik, N., and Dobler, C. (2017), “Post-migration follow-up of migrants identified to be at increased risk of developing tuberculosis at pre-migration screening: a systematic review and meta-analysis,” *Lancet Infect Dis*, 17, 770–779.
- Chowell, G., Blumberg, S., Simonsen, L., Miller, M. A., and Viboud, C. (2014), “Synthesizing data and models for the spread of MERS-CoV, 2013: key role of index cases and hospital transmission,” *Epidemics*, 9, 40–51.
- Cleveland, R., Cleveland, W., McRae, J., and Terpenning, I. (1990), “STL: A seasonal-trend decomposition procedure based on loess,” *Journal of Official Statistics*, 6, 3–33.
- Del Moral, P. (1996), “New Linear Filtering: Interacting Particle Resolution,” *Markov Process Relat*, 2, 555–580.

- Diekmann, O. and Heesterbeek, J. (2000), *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, John Wiley & Sons, LTD, New York.
- Dodd, P., Looker, C., Plumb, I., Bond, V., Schaap, A., Shanaube, K., and *et al.* (2016), “Age- and Sex-Specific Social Contact Patterns and notifications of *Mycobacterium tuberculosis* Infection,” *Am J Epidemiol*, 183, 156–166.
- Down, D., Meyn, S. P., and Tweedie, R. L. (1995), “Exponential and uniform ergodicity of Markov processes,” *Annals of Probability*, 23, 1671–1691.
- Durrett, R. (2008), *Probability Models for DNA Sequence Evolution*, Springer, 2nd edn.
- Earn, D., Rohani, P., Bolker, B., and Grenfell, B. (2000), “A simple model for complex dynamical transitions in epidemics,” *Science*, 287, 667–670.
- Efron, B. and Tibshirani, R. (1986), “Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy,” *Statist Sci*, 1, 54–75.
- Etheridge, A. (2011), *Some Mathematical Models from Population Genetics*, Springer-Verlag, Berlin.
- Ewens, W. (1979), *Mathematical population genetics*, Springer-Verlag, Berlin.
- Farag, E. A., Reusken, C. B., Haagmans, B. L., Mohran, K. A., Raj, V. S., Pas, S. D., Voermans, J., Smits, S. L., Godeke, G.-J., Al-Hajri, M. M., *et al.* (2015), “High proportion of MERS-CoV shedding dromedaries at slaughterhouse with a potential epidemiological link to human cases, Qatar 2014,” *Infect Ecol Epidemiol*, 5, 28305.
- Feller, W. (1957), *An introduction to probability theory and its applications*, John Wiley & Sons, Inc, New York.
- Finkelman, B., Viboud, C., Koelle, K., Ferran, M., Bharti, N., and Grenfell, B. (2007), “Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients,” *PLOS ONE*, 2, e1296.
- Food and Agriculture Organization of the United Nations (accessed 3 June 2016), “EMPRES Global Animal Disease Information System,” <http://www.empres-i.fao.org>.

- Ge, E., Lai, P., Zhang, X., Yang, X., Li, X., Wang, H., and Wei, X. (2015), “Regional transport and its association with tuberculosis in the Shandong province of China, 2009–2011,” *Journal of Transport Geography*, 46, 232–243.
- Gerdil, C. (2003), “The annual production cycle for influenza vaccine,” *Vaccine*, 21, 1776–1779.
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., and Holmes, E. C. (2004), “Unifying the epidemiological and evolutionary dynamics of pathogens,” *Science*, 303, 327–332.
- Haagmans, B. L., Al Dhahiry, S. H., Reusken, C. B., Raj, V. S., Galiano, M., Myers, R., Godeke, G.-J., Jonges, M., Farag, E., Diab, A., et al. (2014), “Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation,” *Lancet Infect Dis*, 14, 140–145.
- Harris, R., Sumner, T., Knight, G., Evans, T., Cardenas, K., Chen, C., and White, R. (2019), “Age-targeted tuberculosis vaccination in China and implications for vaccine development: a modelling study,” *Lancet Global Health*, 7, e209–e218.
- He, D. and Stone, L. (2003), “Spatio-temporal synchronization of recurrent epidemics,” *Proc Biol Sci*, 270, 1519–1526.
- He, D., Ionides, E., , and King, A. (2010), “Plug-and-play inference for disease dynamics: measles in large and small populations as a case study,” *J R Soc Interface*, 7, 271–283.
- He, D., Lui, R., Wang, L., Tse, C., Yang, L., and Stone, L. (2015), “Global Spatio-temporal Patterns of Influenza in the Post-pandemic Era,” *Sci Rep*, 5, 11013.
- He, X., Zhang, X., Zhao, J., Liu, Y., Yu, C., Yang, G., and Li, H. (2016), “Epidemiological trends of drug-resistant tuberculosis in China from 2007 to 2014: a retrospective study,” *Medicine*, 95, e3336.
- He, X., Tao, N., Liu, Y., Zhang, X., and Li, H. (2017), “Epidemiological trends and outcomes of extensively drug-resistant tuberculosis in Shandong, China,” *BMC Infect Dis*, 17, 555.
- Hein, J., Schierup, M. H., and Wiuf, C. (2005), *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*, Oxford University Press, Oxford.
- Hemida, M., Perera, R., Wang, P., Alhammadi, M., Siu, L., Li, M., Poon, L., Saif, L., Alnaeem, A., and Peiris, M. (2013), “Middle East Respiratory Syndrome (MERS) coronavirus seroprevalence in domestic livestock in Saudi Arabia, 2010 to 2013,” *Euro Surveill*, 18, 20659.

- Hemida, M. G., Chu, D. K., Poon, L. L., Perera, R. A., Alhammadi, M. A., Ng, H.-y., Siu, L. Y., Guan, Y., Alnaeem, A., and Peiris, M. (2014a), “MERS Coronavirus in Dromedary Camel Herd, Saudi Arabia,” *Emerg Infect Dis*, 20, 1231–1234.
- Hemida, M. G., Chu, D. K., Poon, L. L., Perera, R., Alhammadi, M. A., Ng, H., Siu, L. Y., Guan, Y., Alnaeem, A., and Peiris, M. (2014b), “MERS coronavirus in dromedary camel herd, Saudi Arabia,” *Emerg Infect Dis*, 20, 1231–1234.
- Horton, K., Sumner, T., Houben, R., Corbett, E., and White, R. (2018), “A Bayesian approach to understanding sex differences in Tuberculosis disease burden,” *Am J Epidemiol*, p. kwy131.
- Hsieh, Y.-H. (2015), “2015 Middle East Respiratory Syndrome Coronavirus (MERS-CoV) nosocomial outbreak in South Korea: insights from modeling,” *PeerJ*, 3, e1505.
- Hussaini, N., Okuneye, K., and Gumel, A. B. (2017), “Mathematical analysis of a model for zoonotic visceral leishmaniasis,” *Infectious Disease Modelling*, 2, 455–474.
- IBM Big Data & Analytics Hub (2018), “The Four V’s of Big Data,” <https://www.ibmbigdatahub.com/infographic/four-vs-big-data>.
- Iondies, E., Nguyen, D., Atchade, Y., Stoev, S., and King, A. (2015), “Inference for dynamic and latent variable models via iterated, perturbed Bayes maps,” *Proc Natl Acad Sci U S A*, 112, 719–724.
- Ionides, E. (2019), “Iterated Filtering,” https://en.wikipedia.org/wiki/Iterated_filtering.
- Ionides, E., Bréto, C., and King, A. (2006), “Inference for nonlinear dynamical systems,” *Proc Natl Acad Sci U S A*, 103, 18438–18443.
- Ionides, E., Bhadra, A., Atchade, Y., and King, A. (2011), “Iterated filtering,” *Ann Stat*, 39, 1776–1802.
- Ionides, E., Breto, C., Park, J., Smith, R., and King, A. (2017), “Monte Carlo profile confidence intervals for dynamic systems,” *J R Soc Interface*, 14, 20170126.
- Iqbal, S., Winston, C., Bardenheier, B., Armstrong, L., and Navin, T. (2018), “Age-period-cohort analysis of Tuberculosis incidence rates by nativity, United States, 1996–2016,” *Am J Public Health*, 108, S315–S320.

- Johnson, R. (2010), “Statistical inference,” https://www.encyclopediaofmath.org/index.php/Statistical_inference#cite_note-1.
- Keeling, M. and Rohani, P. (2008), *Modeling infectious diseases in humans and animals*, Princeton University Press, Princeton.
- Kendall, E., Fofana, M., and Dowdy, D. (2015), “The burdence of transmitted multi-drug resistance among epidemics of tuberculosis: A transmission model,” *Lancet Respir Med*, 3, 963–972.
- Kermack, W. and McKendrick, A. (1927), “A Contribution to the Mathematical Theory of Epidemics,” *Proc Royal Soc*, 115, 700–721.
- Khalafalla, A. I., Lu, X., Al-Mubarak, A. I., Dalab, A. H. S., Al-Busadah, K. A., and Erdman, D. D. (2015), “MERS-CoV in upper respiratory tract and lungs of dromedary camels, Saudi Arabia, 2013–2014,” *Emerg Infect Dis*, 21, 1153–1158.
- Khan, K., Sears, J., Hu, V. W., Brownstein, J. S., Hay, S., Kossowsky, D., Eckhardt, R., Chim, T., Berry, I., Bogoch, I., and Cetron, M. (2013), “Potential for the international spread of middle East respiratory syndrome in association with mass gatherings in Saudi Arabia,” *PLoS Curr*, 5.
- King, A., Nguyen, D., and Ionides, E. (2016), “Statistical Inference for Partially Observed Markov Processes via the R Package pomp,” *J. Stat. Softw*, 69, 1–43.
- King, A., Ionides, E., Breto, C., Ellner, S., Ferrari, M., Kendall, B., Lavine, M., Nguyen, D., Reuman, D., and Wearing, H. (2019), “pomp: Statistical Inference for Partially Observed Markov Processes (R package, version 2),” <http://kingaa.github.io/pomp>.
- King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008), “Inapparent infections and cholera dynamics,” *Nature*, 454, 877–880.
- Kingman, J. F. C. (1982), “The coalescent,” *Stochastic Processes and their Applications*, 13, 235–248.
- Klein, E., Serohijos, A., Shakhnovich, E., and Pekosz, Z. (2014), “Influenza A H1N1 pandemic strain evolution–divergence and the potential for antigenic drift variants,” *PLOS ONE*, 9, e93632.
- Knoema (2016), “Saudi Arabia foreign trade,” <https://knoema.com/SAFT2014/saudi-arabia-foreign-trade>.
- Kuang, D., Nielsen, B., and Neilsen, J. (2008), “Identification of the age-period-cohort model and the extended chain-ladder model,” *Biometrika*, 95, 979–986.

- Kucharski, A. and Althaus, C. (2015), “The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission,” *Euro Surveill*, 20, 14–18.
- Kupferschmidt, K. (2014), “Emerging Diseases. Soaring MERS Cases in Saudi Arabia Raise Alarms,” *Science*, 344, 457–458.
- LaSalle, J. (1976), *The stability of dynamical systems*, SIAM, Philadelphia.
- Lau, M., Marion, G., Streftaries, G., and Gibson, G. (2011), “A systematic Bayesian integration of epidemiological and genetic data,” *PLOS Comput Biol*, 11, e1004633.
- Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C., Smith, D., Pybus, O., Brockmann, D., and Suchard, M. (2014), “Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2,” *PLoS Pathogens*, 10, e1003932.
- Lessler, J., Rodriguez-Barraquer, I., Cummings, D. A., Garske, T., Van Kerkhove, M., Mills, H., Truelove, S., Hakeem, R., Albarrak, A., Ferguson, N. M., et al. (2014), “Estimating potential incidence of MERS-CoV associated with Hajj pilgrims to Saudi Arabia, 2014,” *PLoS Curr*, 6.
- Longini, I., Koopman, J., Haber, M., and Cotsonis, G. (1987), “Statistical inference for infectious diseases,” *Am J Epidemiol*, 128, 845–859.
- Memish, Z., Zumla, A., Alhakeem, R., Assiri, A., Turkestani, A., Al Harby, K., Alyemni, M., Dhafar, K., Gautret, P., Barbeschi, M., McCloskey, B., Heymann, D., Al Rabeeah, A., and Al-Tawfiq, J. (2014), “Hajj: infectious disease surveillance and control,” *Lancet*, 383, 2073–82.
- Meyer, B., Müller, M. A., Corman, V. M., Reusken, C. B., Ritz, D., Godeke, G. J., Lattwein, E., Kallies, S., Siemens, A., van Beek, J., et al. (2014), “Antibodies against MERS coronavirus in dromedaries, United Arab Emirates, 2003 and 2013,” *Emerg Infect Dis*, 20, 552–559.
- Meyer, B., Juhasz, J., Barua, R., Das, G. A., Hakimuddin, F., Corman, V., Müller, M., Wernery, U., Drosten, C., and Nagy, P. (2016a), “Time Course of MERS-CoV Infection and Immunity in Dromedary Camels,” *Emerg Infect Dis*, 22.
- Meyer, B., Juhasz, J., Barua, R., Das Gupta, A., Hakimuddin, F., Corman, V. M., et al. (2016b), “Time course of MERS-CoV infection and immunity in dromedary camels,” *Emerg Infect Dis*, 22, 2171–2173.

- Meyn, S. P. and Tweedie, R. L. (2009), *Markov chains and stochastic stability*, Springer-Verlag, London, second edn., Available at <http://probability.ca/MT/>.
- Müller, M. A., Meyer, B., Corman, V. M., Al-Masri, M., Turkestani, A., Ritz, D., Sieberg, A., Aldabbagh, S., Bosch, B.-J., Lattwein, E., et al. (2015), “Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a nationwide, cross-sectional, serological study,” *Lancet Infect Dis*, 15, 559–564.
- National Bureau of Statistics (2011), “The Sixth National Population Census,” <http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm>, Accessed: 2018-09-03.
- Nielsen and B (2015), “apc: An R package for age-period-cohort analysis,” *R Journal*, 7, 52–64.
- Paules, C. and Subbarao, K. (2017), “Influenza,” *Lancet*, 390, 697–708.
- Paules, C., Sullivan, S., Subbarao, K., and Fauci, A. (2018), “Chasing Seasonal Influenza – The Need for a Universal Influenza Vaccine,” *N Engl J Med*, 378, 7–9.
- People’s Government of Shandong Province (2017), ““Thirteenth Five-Year” Plan of TB control in Shandong Province,” <http://news.sdchina.com/show/4208778.html>, Accessed: 2018-09-03.
- Piedra, P. A., Gaglani, M. J., Kozinetz, C. A., Herschler, G., Riggs, M., Griffith, M., Fewlass, C., Watts, M., Hessel, C., Cordova, J., and Glezen, W. P. (2005), “Herd immunity in adults against influenza-related illnesses with use of the trivalent-live attenuated influenza vaccine (CAIV-T) in children,” *Vaccine*, 23, 1540–1548.
- Poletto, C., Pelat, C., Levy-Bruhl, D., Yazdanpanah, Y., Boelle, P.-Y., and Colizza, V. (2014), “Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach,” *Euro Surveill*, 19, 20824.
- Poletto, C., Colizza, V., and Boëlle, P.-Y. (2016), “Quantifying spatiotemporal heterogeneity of MERS-CoV transmission in the Middle East region: a combined modelling approach,” *Epidemics*, 15, 1–9.
- Pybus, O., Rambaut, A., and Harvey, P. (2000), “An Integrated Framework for the Inference of Viral Population History From Reconstructed Genealogies,” *Genetics*, 144, 1429–1437.

- Rasmussen, D. A., Ratmann, O., and Koelle, K. (2011), “Inference for nonlinear epidemiological models using genealogies and time series,” *PLoS Computational Biology*, 7, e1002136.
- Rasmussen, D. A., Boni, M. F., and Koelle, K. (2014), “Reconciling phylodynamics with epidemiology: the case of dengue virus in southern Vietnam,” *Molecular Biology and Evolution*, 31, 258–271.
- Reusken, C. B., Haagmans, B. L., Müller, M. A., Gutierrez, C., Godeke, G.-J., Meyer, B., Muth, D., Raj, V. S., Smits-De Vries, L., Corman, V. M., et al. (2013), “Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study,” *Lancet Infect Dis*, 13, 859–866.
- Reusken, C. B., Messadi, L., Feyisa, A., Ularanu, H., Godeke, G. J., Danmarwa, A., Dawo, F., Jemli, M., Melaku, S., Shamaki, D., et al. (2014), “Geographic distribution of MERS coronavirus among dromedary camels, Africa,” *Emerg Infect Dis*, 20, 1370–1374.
- Reusken, C. B., Raj, V. S., Koopmans, M. P., and Haagmans, B. L. (2016), “Cross host transmission in the emergence of MERS coronavirus,” *Curr Opin Virol*, 16, 55–62.
- Rösch, A. and Schmidbauer, H. (2018), “WaveletComp 1.1: A guided tour through the R package,” http://www.hs-stat.com/projects/WaveletComp/WaveletComp_guide_tour.pdf.
- Sabir, J. S., Lam, T. T.-Y., Ahmed, M. M., Li, L., Shen, Y., Abo-Aba, S. E., Qureshi, M. I., Abu-Zeid, M., Zhang, Y., Khiyami, M. A., et al. (2016), “Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia,” *Science*, 351, 81–84.
- Schwarz, G. et al. (1978), “Estimating the dimension of a model,” *Ann Stat*, 6, 461–464.
- Shandong Provincial Bureau of Statistics (2018), “Shandong Statistical Yearbook,” http://www.stats-sd.gov.cn/art/2018/4/13/art_6131_404684.html, Accessed: 2018-09-03.
- Sharma, A., Hill, A., Kurbatova, E., van, der, W. M., and *et al.* (2017), “Estimating the future burden of multidrug-resistant and extensively drug-resistant tuberculosis in India, the Philippines, Russia, and South Africa: a mathematical modelling study,” *Lancet Infect Dis*, 17, 707–715.

- Sims, C. (1980), “Macroeconomics and Reality,” *Econometrica*, 48, 1–48.
- Skowronski, D., Chambe, r. C., De Serres, G., Dickinson, J., and *et al.* (2018), “Early Season Co-circulation of Influenza A(H3N2) and B(Yamagata): Interim Estimates of 2017/2018 Vaccine Effectiveness, Canada, January 2018,” *Euro Surveil*, 23, pii=18–00035.
- Smith, R., Ionides, E., and King, A. (2017), “Infectious disease dynamics inferred from genetic data via sequential Monte Carlo,” *Mol Biol Evol*, 34, 2065–2084.
- Smith, T. and Wakefield, J. (2016), “A review and comparison of age-period-cohort models for cancer notifications,” *Statist Sci*, 31, 591–610.
- Sun, G., Xie, J., Huang, S., Jin, Z., Li, M., and Liu, L. (2017), “Transmission dynamics of cholera: Mathematical modeling and control strategies,” *Commun Nonlinear Sci Numer Simulate*, 45, 235–244.
- Sun, Y., Zhu, L., Lu, Z., and Jia, Z. (2016), “Notification Rate of Tuberculosis among Migrants in China 2005–2014: A Systematic Review and Meta-analysis,” *Chin Med J (Engl)*, 129, 1856–1860.
- Tamerius, J., Nelson, M., Zhou, S., Viboud, C., Miller, M., and Alonso, W. (2011), “Global influenza seasonality: Reconciling patterns across temperate and tropical regions,” *Environ Health Perspect*, 119, 439–445.
- Tang, X., Zhao, S., Chiu, A., Wang, X., Yang, L., and He, D. (2018), “Analyzing increasing Trends of Guillain-Barré Syndrome (GBS) and dengue cases in Hong Kong using meteorological data,” *PLOS ONE*, 12, e0187830.
- Tao, N., He, X., Zhang, X., Liu, Y., Yu, C., and Li, H. (2017), “Trends and characteristics of drug-resistant tuberculosis in rural Shandong, China,” *Int J Infect Dis*, 65, 8–14.
- Technical Guidance Group of the Fifth National TB Epidemiological Survey (2012), “The fifth national tuberculosis epidemiological survey in 2010,” *Chin J Antituberc*, 34.
- The Animal Ageing and Longevity Database (accessed on June 3, 2016), “http://genomics.senescence.info/species/entry.php?species=Camelus_dromedarius,” .
- United Nations (accessed on June 3, 2016), “Food and Agriculture Organization of the United Nations Statistics Division,” <http://faostat3.fao.org/download/Q/QA/E>.

- Van-den Driessche, P. and Watmough, J. (2002), “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Math Biosci*, 180, 29–48.
- Veleda, D., Montagne, R., and Araujo, M. (2012), “Cross-Wavelet Bias Corrected by Normalizing Scales,” *J Atmos Ocean Tech*, 29, 1401–1408.
- Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., and Frost, S. D. W. (2009), “Phylodynamics of Infectious Disease Epidemics,” *Genetics*, 183, 1421–1430.
- Wakeley, J. (2001), “The coalescent in an island model of population subdivision with variation among demes,” *Theoretical population biology*, 59, 133–144.
- Wakeley, J. and Aliacar, N. (2001), “Gene genealogies in a metapopulation,” *Genetics*, 159, 893–905.
- Wang, W., Zhang, W., Alayi, A., Yan, C., Zhang, W., and Cao, M. (2016), “The Characteristics of TB Epidemic and TB/HIV Epidemic: 2007–2013 Retrospective Study in Urumqi, Xinjiang Province, China,” *PLoS ONE*, 11, e0164947.
- Wernery, U., Corman, V. M., Wong, E. Y., Tsang, A. K., Muth, D., Lau, S. K., Khazanehdari, K., Zirkel, F., Ali, M., Nagy, P., et al. (2015), “Acute Middle East respiratory syndrome coronavirus infection in livestock dromedaries, Dubai, 2014,” *Emerg Infect Dis*, 21, 1019–1022.
- WHO (2014), “Vigilant on new Middle East respiratory syndrome developments,” <http://www.emro.who.int/media/news/mers-developments.html>.
- WHO (2015), “WHO End TB Strategy,” https://www.who.int/tb/post2015_strategy/en, Accessed: 2019-01-30.
- WHO (2016), “Middle East respiratory syndrome coronavirus (MERS-CoV),” <http://www.who.int/emergencies/mers-cov/en/>.
- WHO (2018a), “Global tuberculosis report 2018,” https://www.who.int/tb/publications/global_report/en/, Accessed: 2019-01-30.
- WHO (2018b), “Tuberculosis in China, WHO,” <http://www.wpro.who.int/china/mediacentre/factsheets/tuberculosis/en/>, Accessed: 2018-09-03.
- Wu, B., Yu, Y., Xie, W., Liu, Y., Zhang, Y., Hu, D., and Li, Y. (2017), “Epidemiology of tuberculosis in Chongqing, China: a secular trend from 1992 to 2015,” *Sci Rep*, 7, 7832.

- Wu, P., Cowling, B., Schooling, C., Wong, I., Johnston, J., Leung, C., Tam, C., and Leung, G. (2008), “Age-period-cohort analysis of tuberculosis notifications in Hong Kong from 1961 to 2005,” *Thorax*, 63, 312–316.
- Xu, B., Jiang, Q., Xiu, Y., and Diwan, V. (2005), “Diagnostic delays in access to tuberculosis care in counties with or without the National Tuberculosis Control Programme in rural China,” *Int J Tuberc Lung Dis*, 9, 784–790.
- Yang, C., Wang, X., Gao, D., and Wang, J. (2017), “Impact of awareness programs on cholera dynamics: two modeling approaches,” *Bull Math Bio*, 7, 2109–2131.
- Yang, C., Lu, L., Warren, J., Wu, J., Jiang, Q., Zuo, T., Gan, M., and *et al.* (2018a), “Internal migration and the transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, and genomic analysis,” *Lancet Infect Dis*, 18, 788–795.
- Yang, J., Lau, Y., Wu, P., Feng, L., Wang, X., and *et al.* (2018b), “Variation in Influenza B Virus Epidemiology by Lineage, China,” *Emerg Infect Dis*, 24, 1536–1540.
- Yang, Y. and Land, K. (2013), *Age-period-cohort analysis: new models, methods, and empirical applications*, Chapman and Hall/CRC, New York.
- Zhang, S., Yan, H., Zhang, J., Zhang, T., Li, X., and Zhang, Y. (2010), “The experience of college students with pulmonary tuberculosis in Shaanxi, China: a qualitative study,” *BMC Infect Dis*, 10, 174.
- Zumla, A., Hui, D. S., and Perlman, S. (2015), “Middle East respiratory syndrome,” *Lancet*, 386, 995–1007.