Fogarty International Center

Principles of time series analysis

Cécile Viboud and Chelsea Hansen

ANISE 7th Meeting

September 13, 2023

Fogarty International Center

Outline

- Surveillance data
- Seasonality models
- Disease burden models
- Demonstration of disease burden models in R

Surveillance data

Disease Surveillance Data (SARI/SRI in South Africa)

	A	B	C	D	E	H	K	N
1	week	year	case	flu	influenzaA	influenzaB	rsv	tested
758	26	2012	SARI/SRI	9	7	2	20	111
759	27	2012	SARI/SRI	13	11	2	20	118
760	28	2012	SARI/SRI	13	8	6	7	97
761	29	2012	SARI/SRI	11	8	3	11	86
762	30	2012	SARI/SRI	7	4	3	12	104
763	31	2012	SARI/SRI	14	9	5	11	123
764	32	2012	SARI/SRI	12	9	3	4	87
765	33	2012	SARI/SRI	22	11	11	17	123
766	34	2012	SARI/SRI	27	10	17	6	110
767	35	2012	SARI/SRI	24	6	19	5	121
768	36	2012	SARI/SRI	18	6	12	14	103
769	37	2012	SARI/SRI	19	3	16	11	118
770	38	2012	SARI/SRI	20	3	18	11	138
771	39	2012	SARI/SRI	8	1	7	7	91
772	40	2012	SARI/SRI	3	0	3	13	113
773	41	2012	SARI/SRI	2	0	2	4	90

Surveillance data requirements and standardization

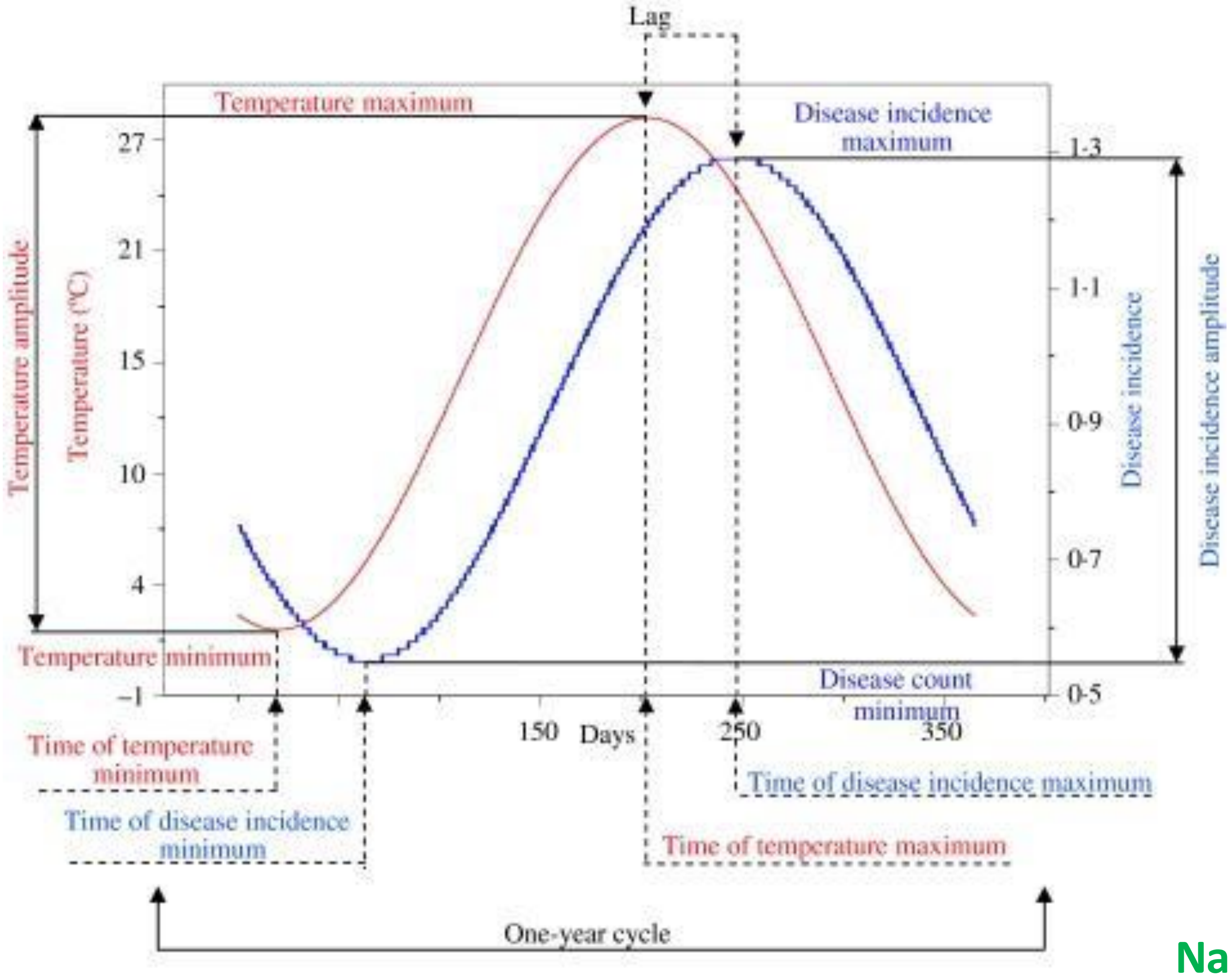
- Serially sampled disease rates at regular time intervals
  - Daily, weekly or monthly virus positive hosts
  - Standardized by the number of specimens tested
  - log, sqrt, time trends, etc
- $\geq$3 years of data for time series analysis
- Data can be spatially disaggregated

Seasonality

Seasonality: what are the research questions?

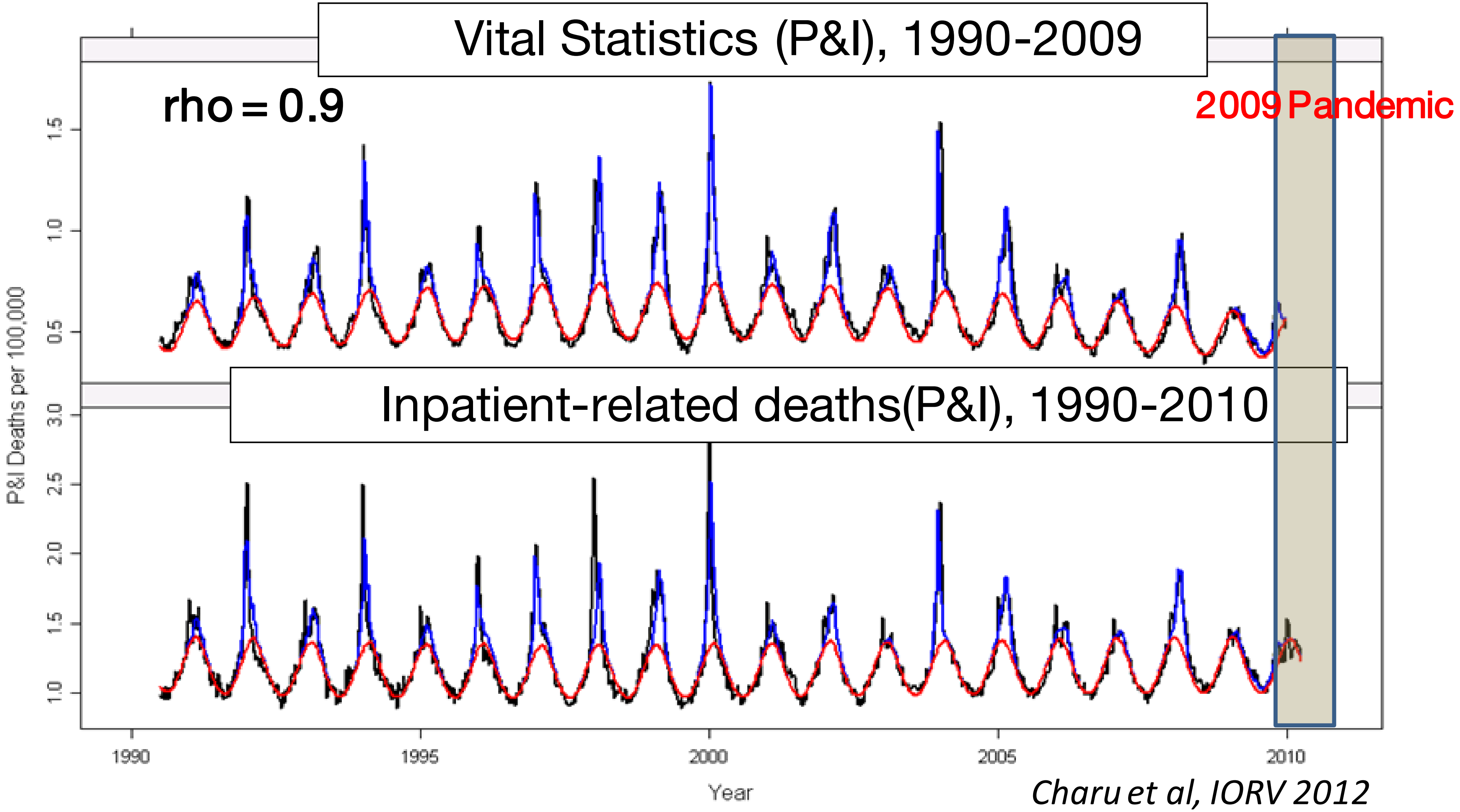
- Seasonality: systematic peaks of infection at certain times of the year
- How seasonal is my disease?
- What are the drivers of seasonality?
  - population aggregation at specific times of the year (school cycles, Christmas, etc)
  - population movements
  - environmental factors
  - birth rate pulses (animal infections)
- Does seasonality differ by geographic location?

Characterizing seasonality



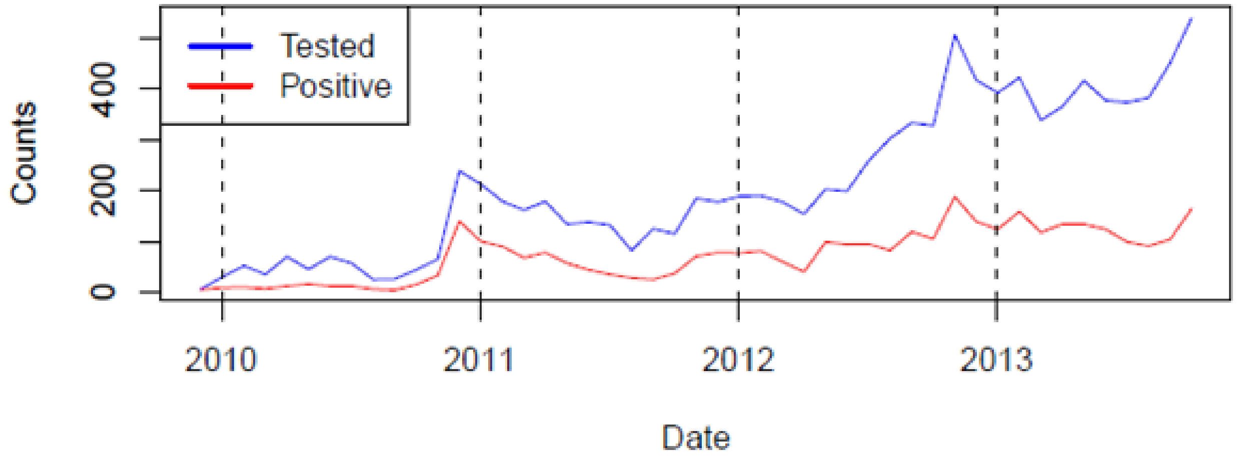
Naumova et al, Epi & Infection, 2007

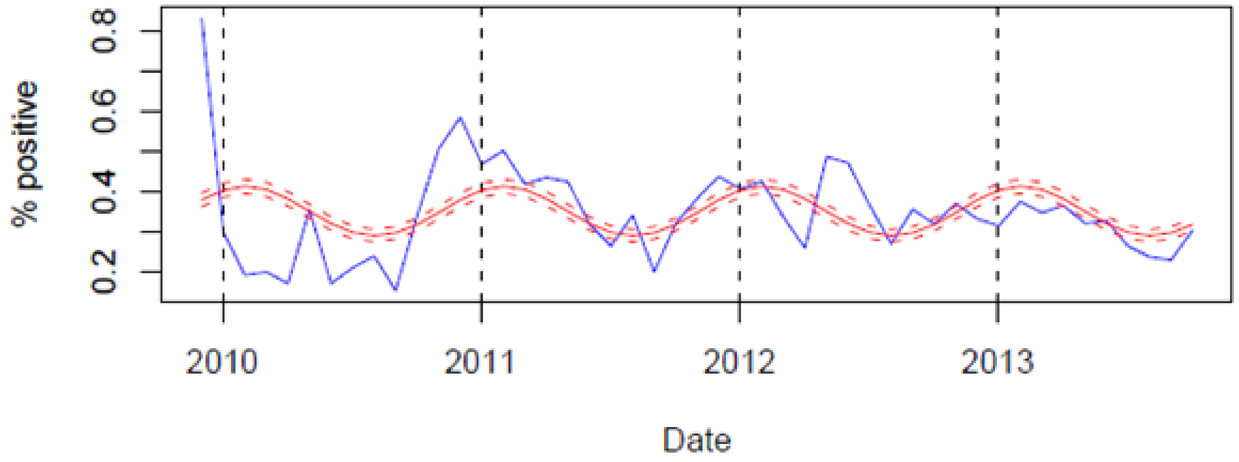
Seasonality of human influenza, USA



Charu et al, IORV 2012

Seasonality in incidence of influenza in US Swine





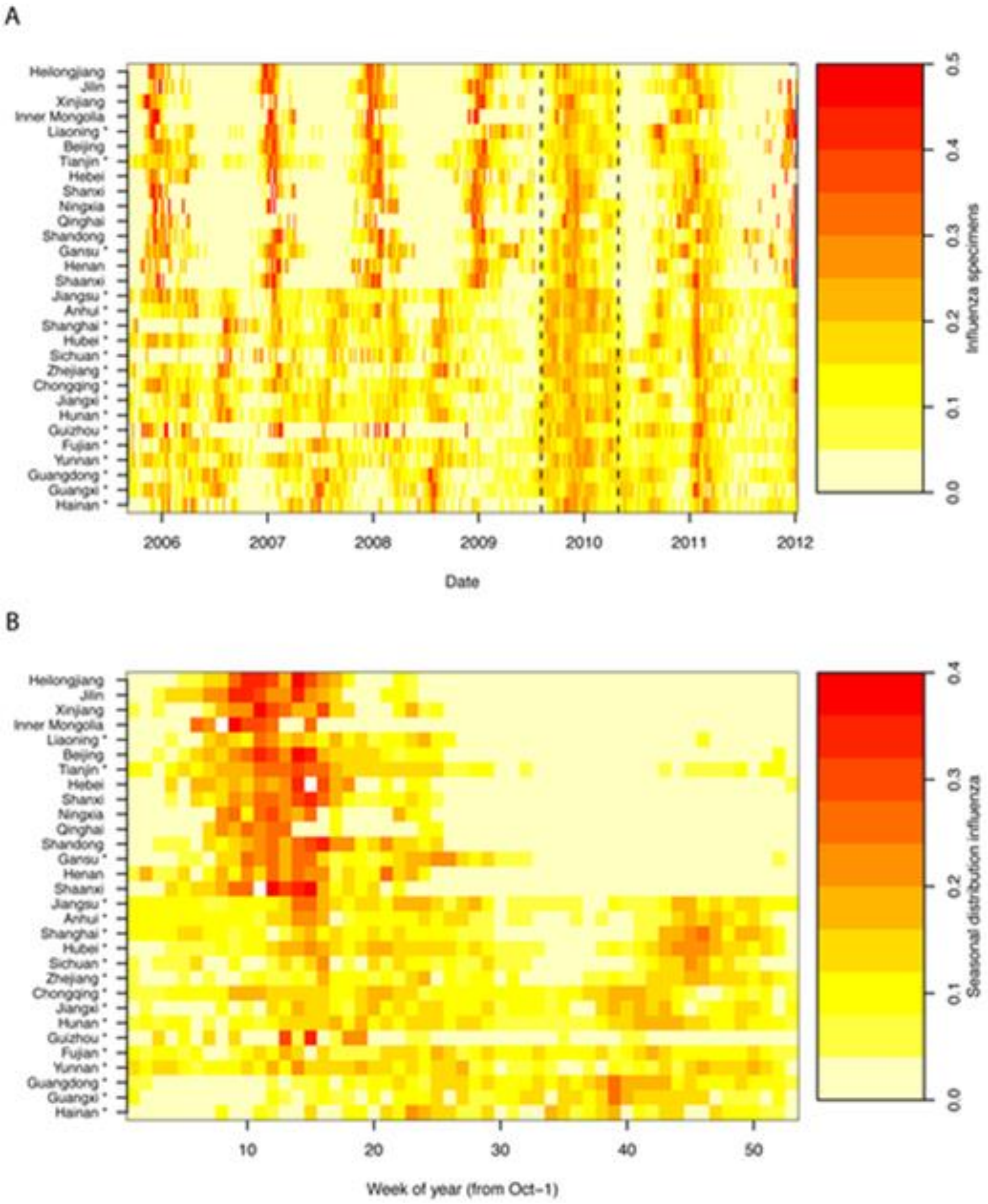
Characterizing seasonal parameters

- Fit linear model to weekly or monthly disease incidence, including:
  - harmonic terms (seasonal factors)
  - time trends

$E(Yti) = \alpha + \beta_1 t_i + \beta_2 t_i^2 +$
$\beta_3 \cos(2\pi\, t_i/52.17) + \beta_4 \sin(2\pi\, t_i/52.17) +$
$\beta_5 \cos(4\pi\, t_i/52.17) + \beta_6 \sin(4\pi\, t_i/52.17) + \varepsilon_i$

Annual amplitude = $\mathrm{sqrt}(\beta_3^2 + \beta_4^2)$
Annual peak timing = $\mathrm{atan}(\beta_3/\beta_4)$

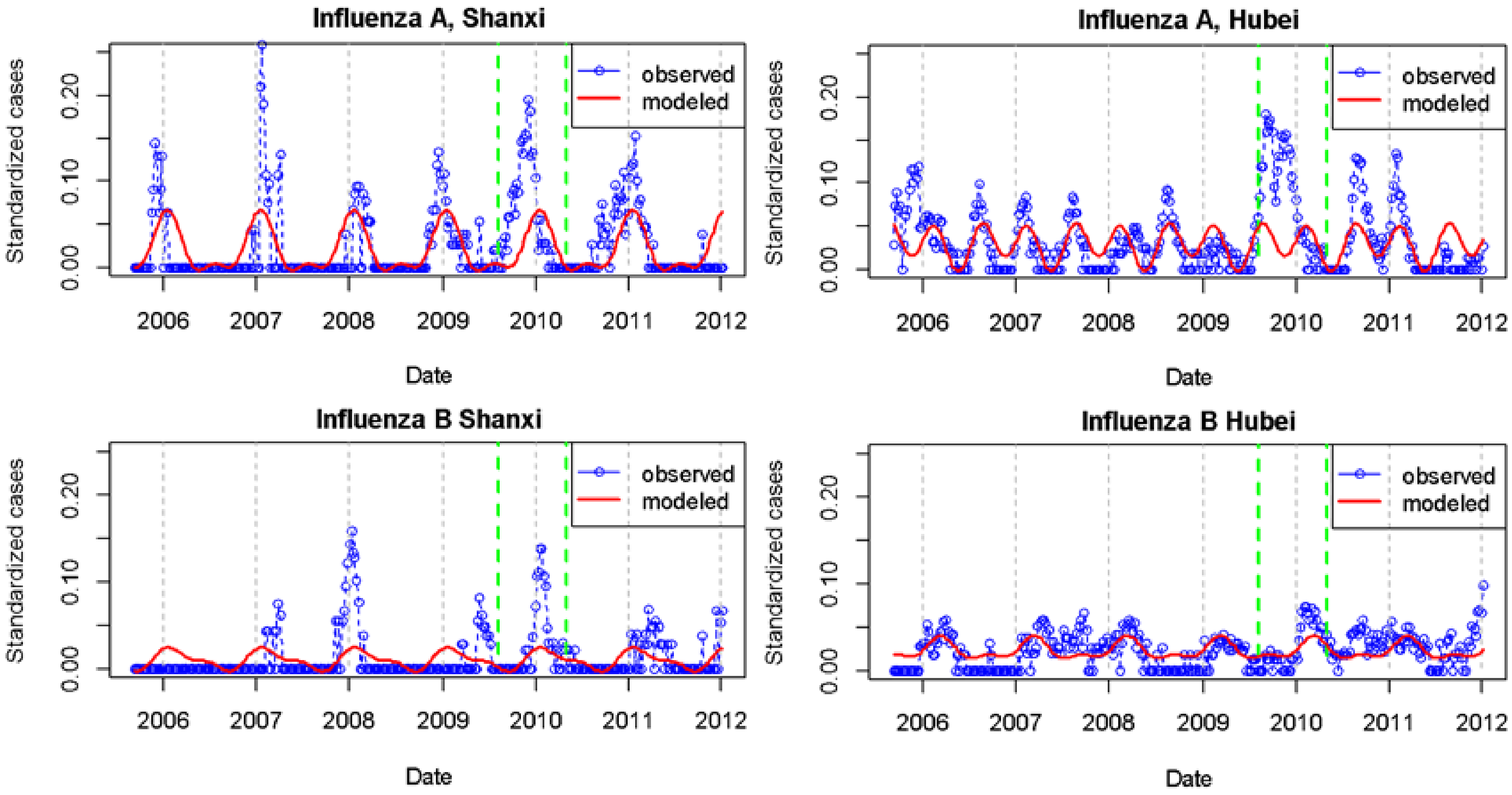
Flu surveillance in Chinese provinces



Heatmaps can be very useful for visualization

Yu et al, Plos Med 2014

Fitting seasonal model to Chinese data

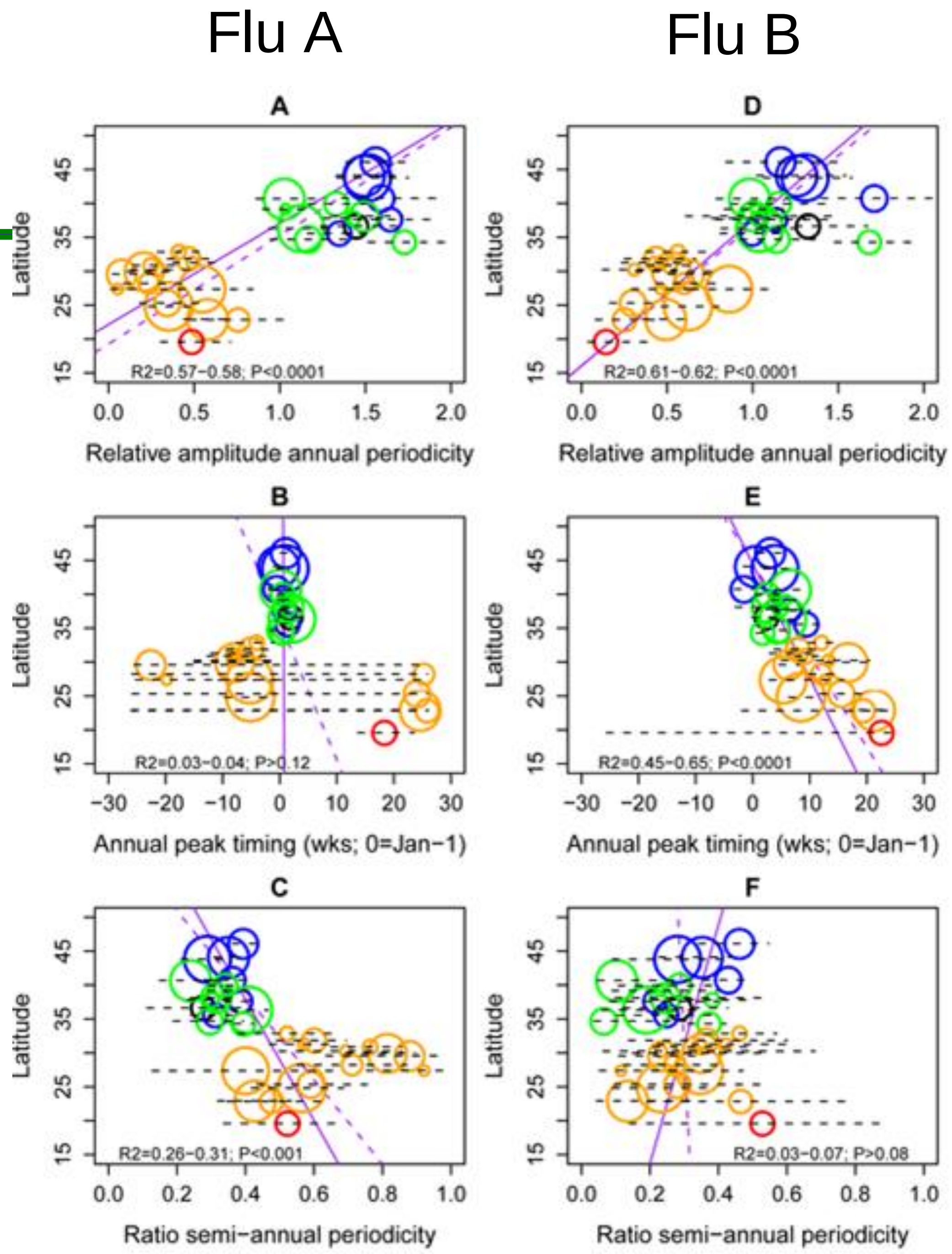


Yu et al, Plos Med 2014

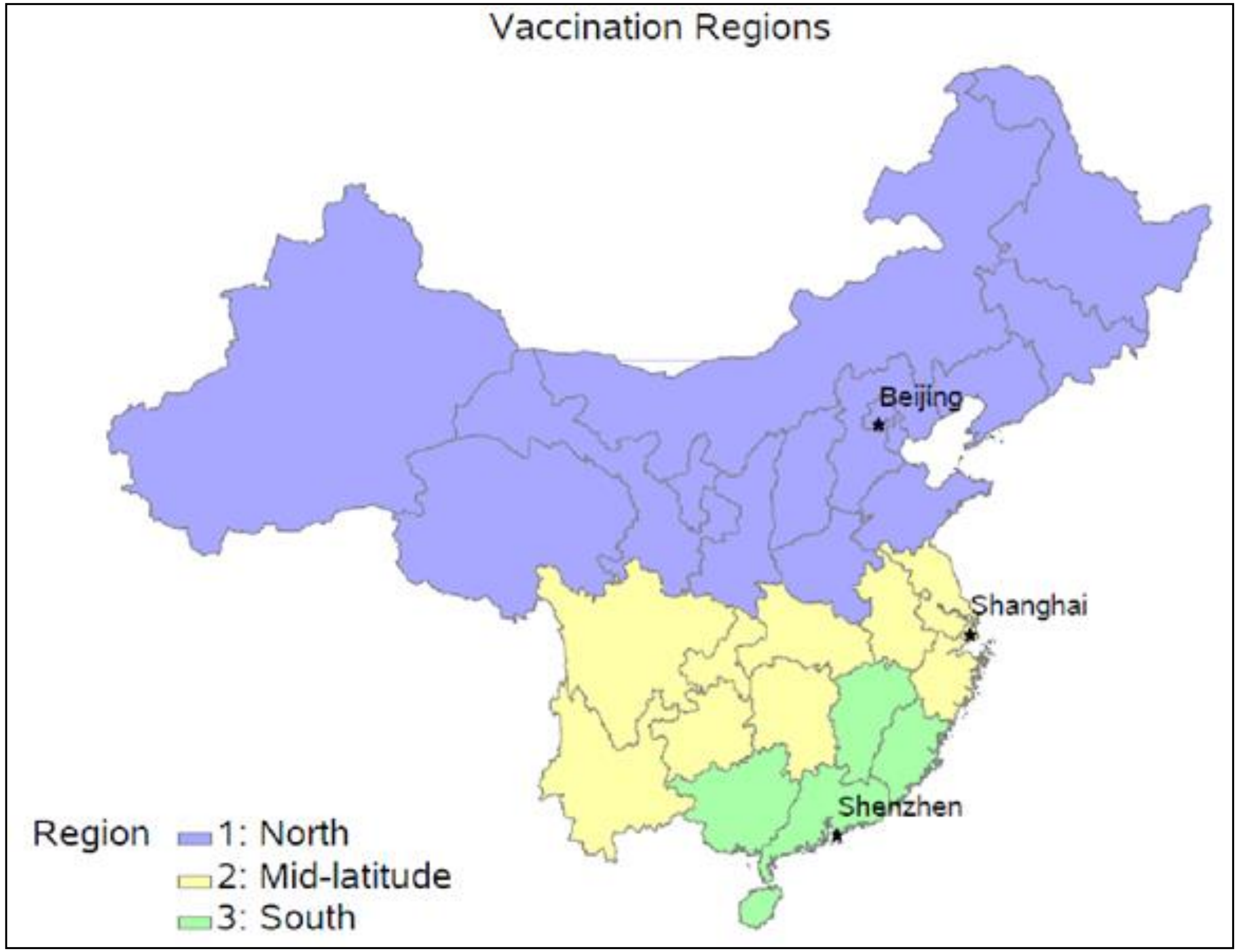
Seasonal parameters

- How seasonal?
- Peak timing?
- 2 peaks within the year?

Yu et al, Plos Med 2014



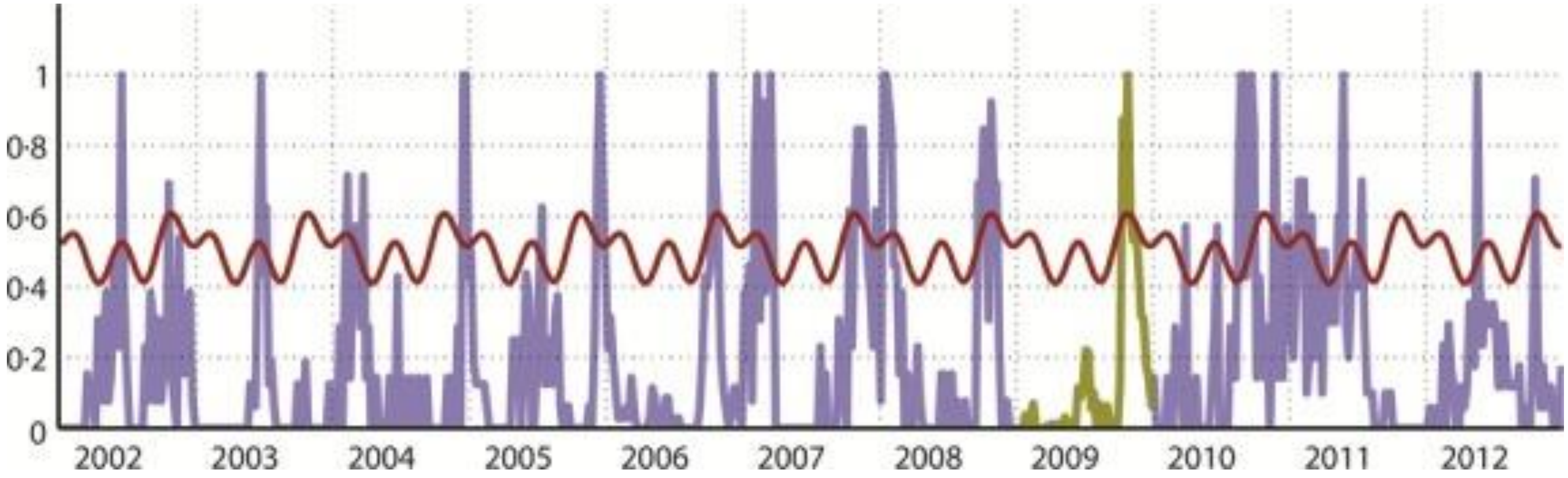
Guiding the timing of annual influenza vaccination campaigns



Yu et al, Plos Med 2014

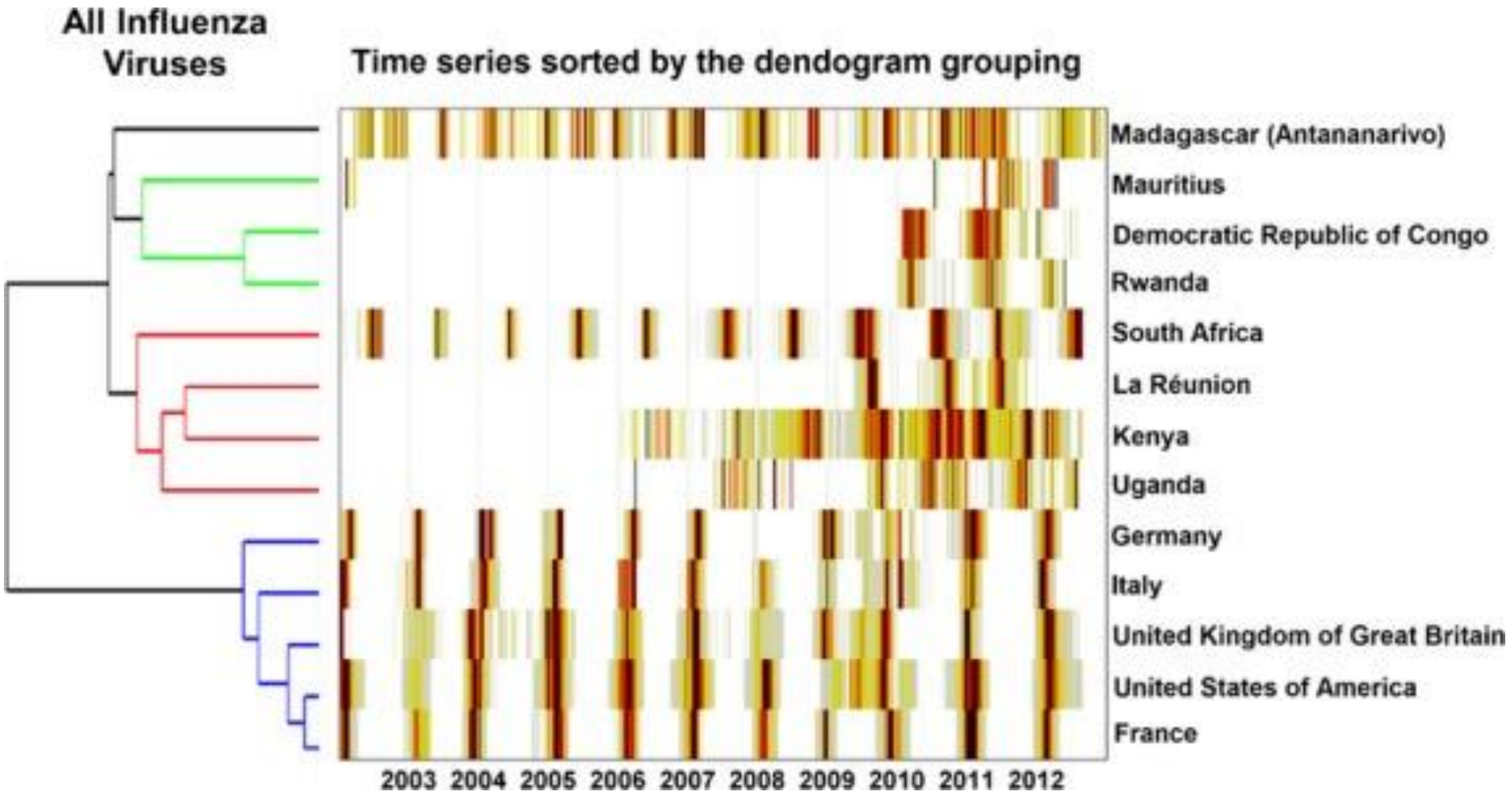
Very weak seasonal signal for influenza in Madagascar

- Peak timing varies between years
- No association with seasonality of tourism
- No association with climate factors



Alonso et al, IORV 2015

Relationship with influenza in other countries? (hierarchical clustering)



Alonso et al, IORV 2015

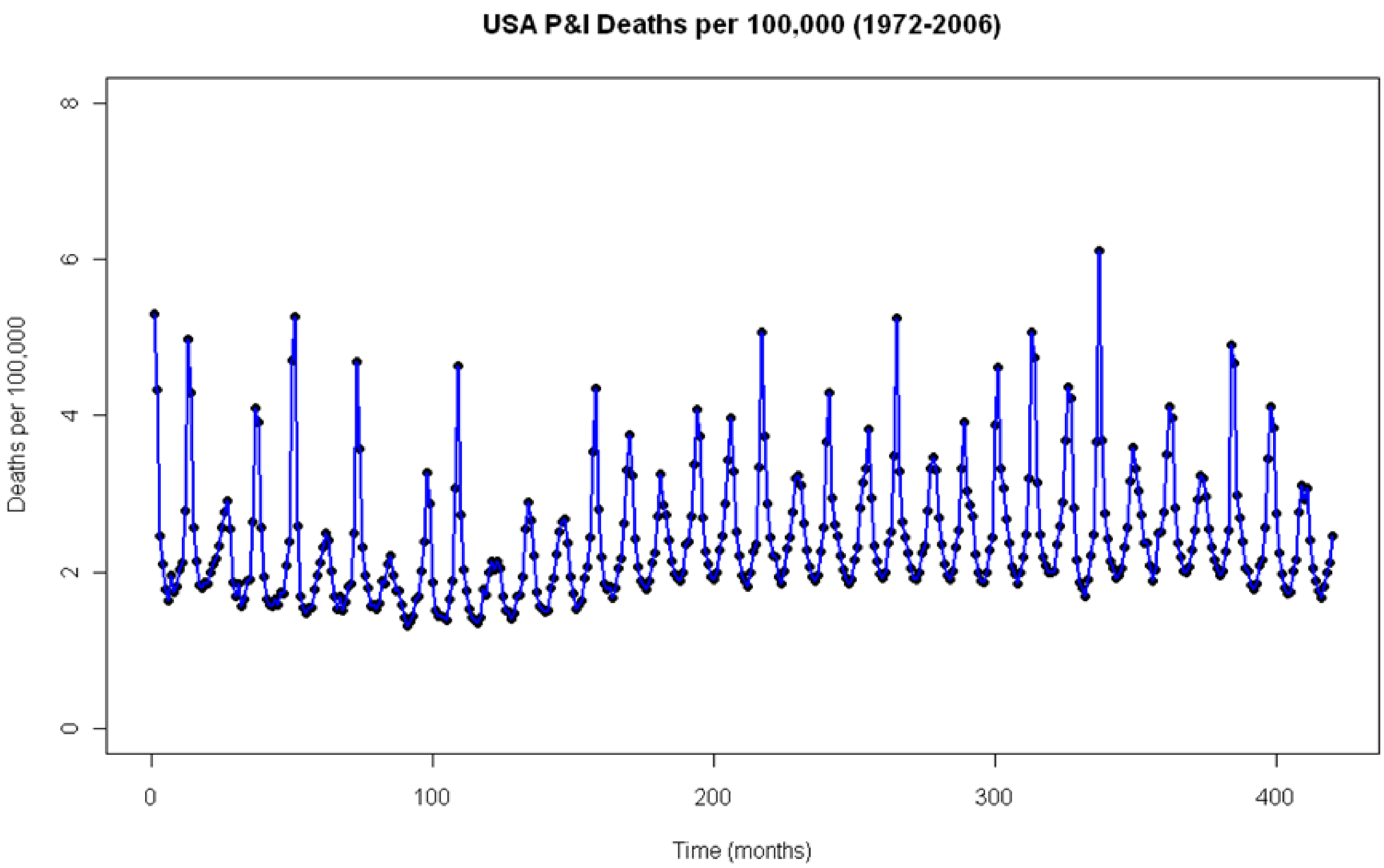
Time series disease burden models

1. Serfling regression models
2. Regression models driven by viral activity

Disease burden models: intro

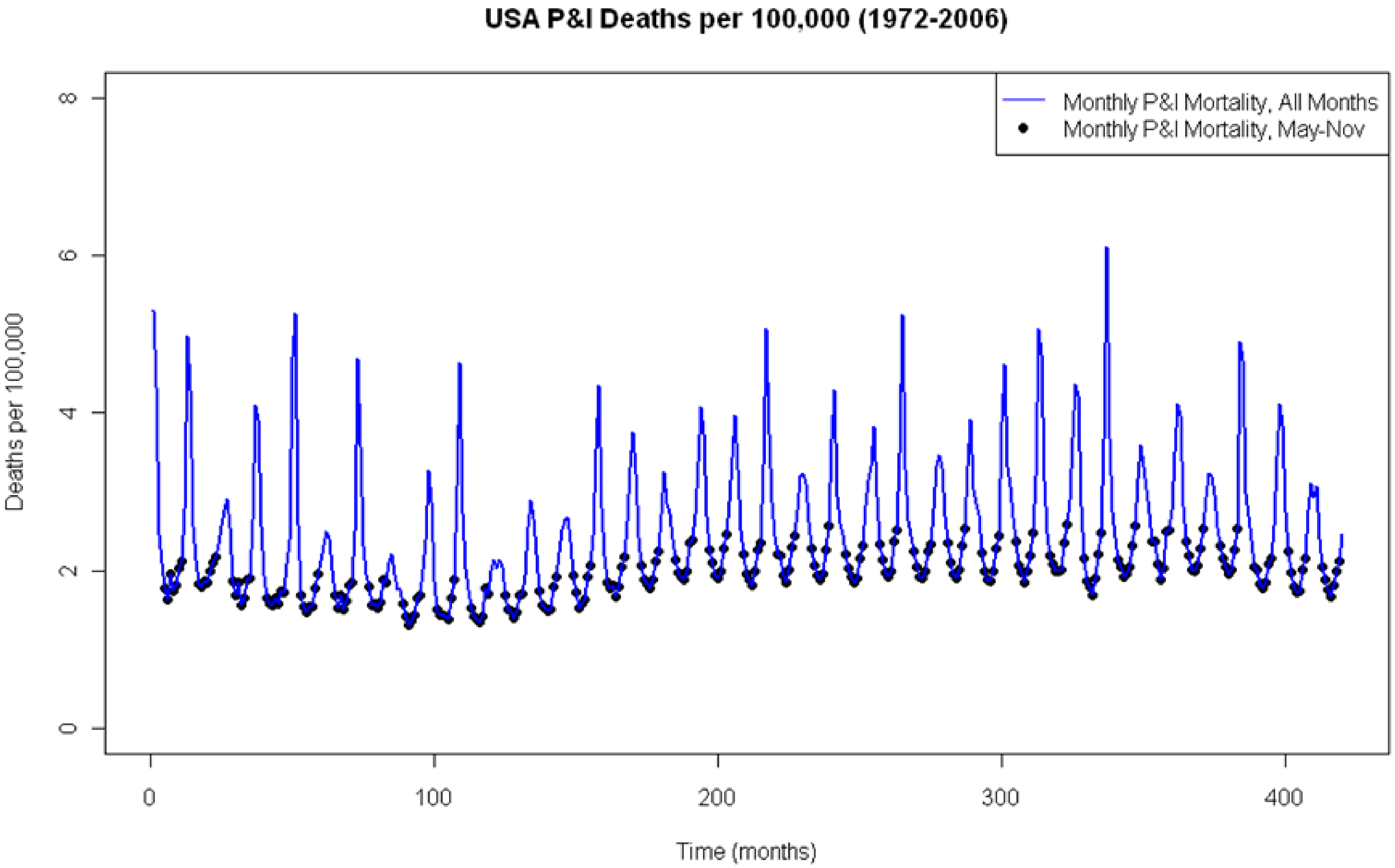
- Motivating question: How can we estimate the contribution of influenza above a background of unrelated deaths, hospitalizations, emergency visits, or other clinical outcome?
- “Excess” pneumonia and influenza deaths, excess respiratory deaths, excess cardio-respiratory deaths

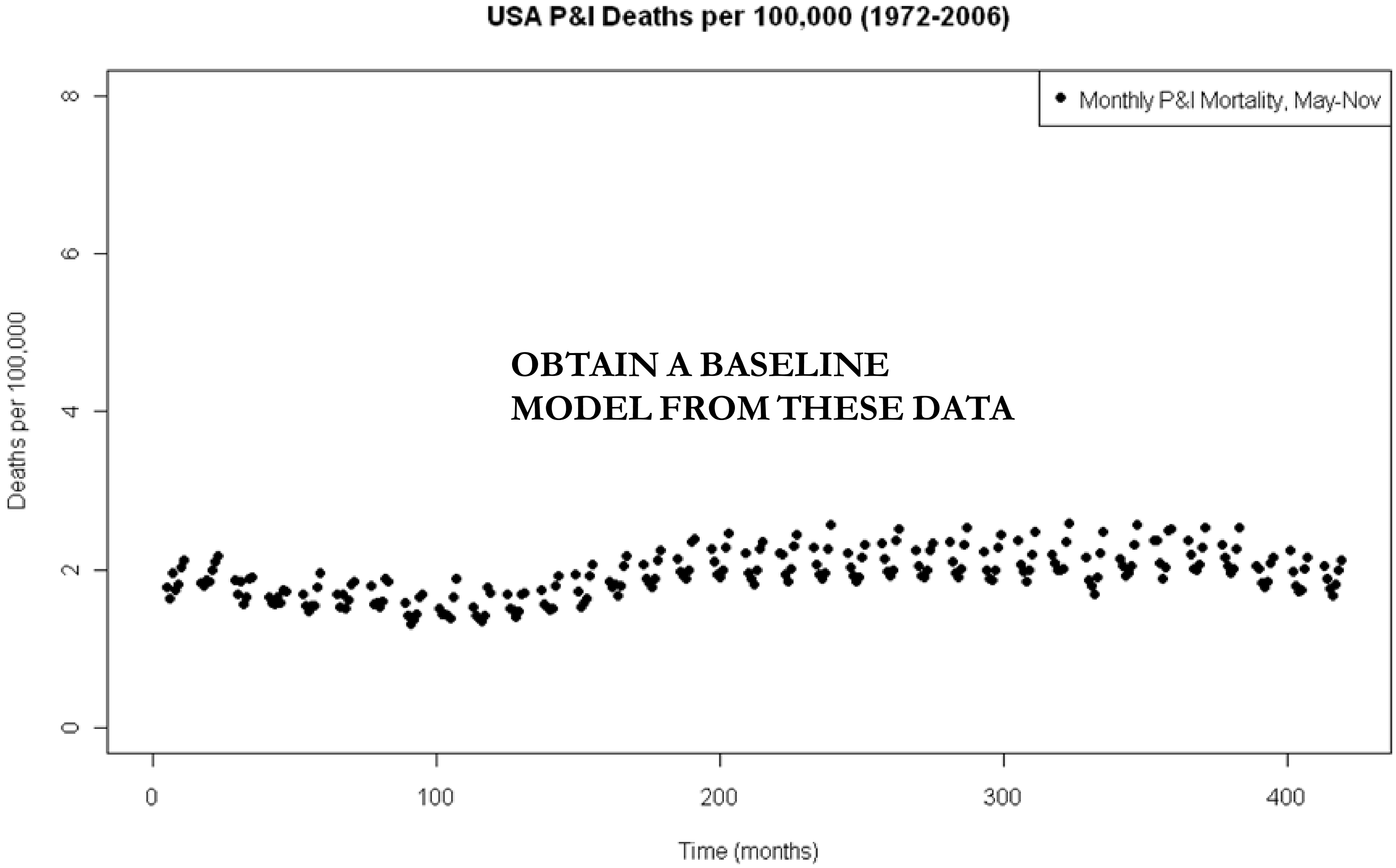
SERFLING MODELS


USA P&I Deaths per 100,000 (1972-2006)
Deaths per 100,000
0
2
4
6
8
0
100
200
300
400
Time (months)

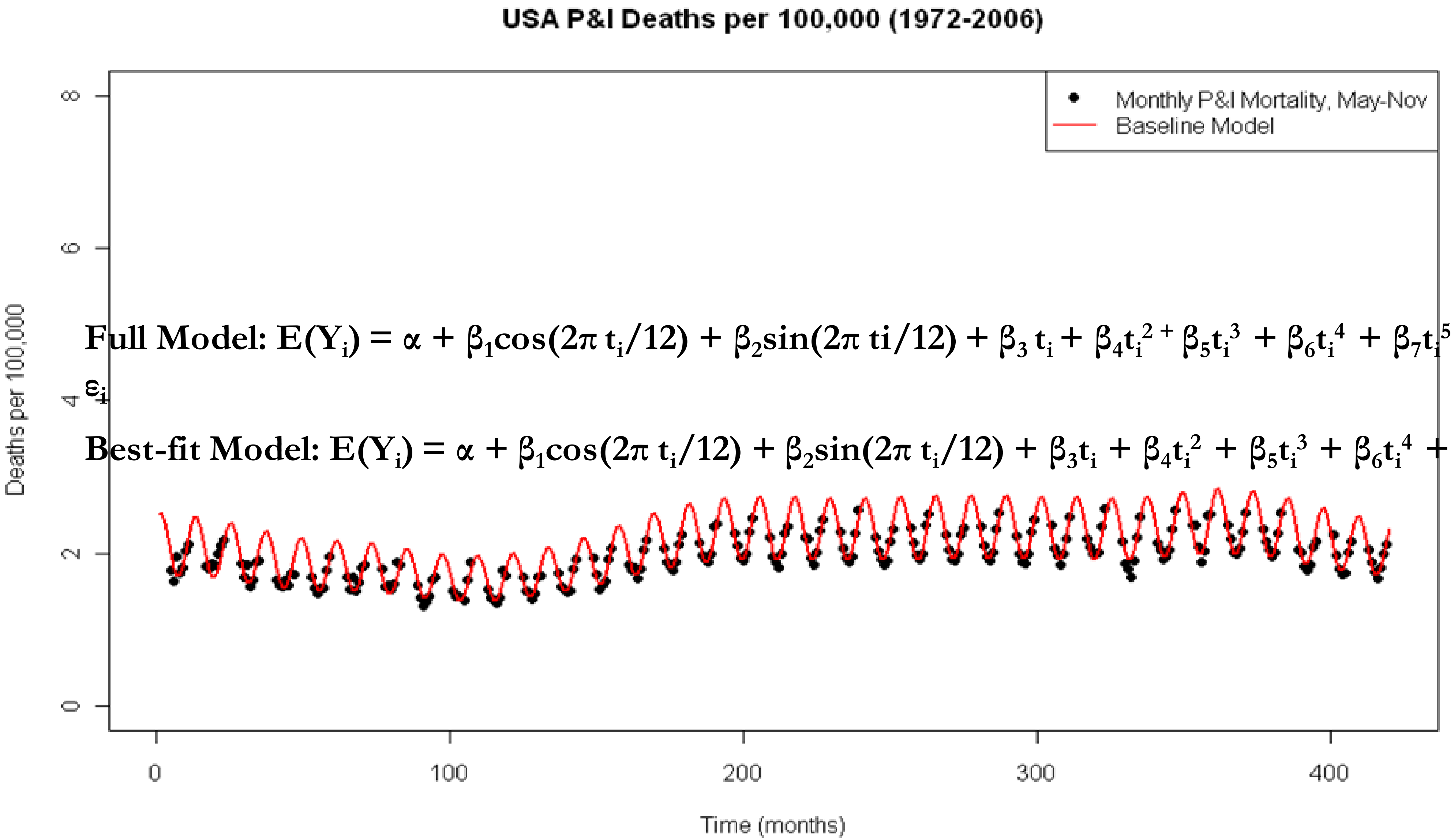
Basic approach: serfling regression

- Step 1: Define influenza, non-influenza period
- Step 2: set a baseline and threshold (95% confidence interval) for mortality during non-influenza period
- Step 3: For each winter, calculate how much mortality above the baseline


USA P&I Deaths per 100,000 (1972-2006)
Monthly P&I Mortality, All Months
Monthly P&I Mortality, May-Nov
Deaths per 100,000
Time (months)
0
2
4
6
8
0
100
200
300
400



OBTAIN A BASELINE MODEL FROM THESE DATA



Full Model: $E(Y_i) = \alpha + \beta_1\cos(2\pi\, t_i/12) + \beta_2\sin(2\pi\, ti/12) + \beta_3\, t_i + \beta_4 t_i^2 + \beta_5 t_i^3 + \beta_6 t_i^4 + \beta_7 t_i^5 + \varepsilon_i$

Best-fit Model: $E(Y_i) = \alpha + \beta_1\cos(2\pi\, t_i/12) + \beta_2\sin(2\pi\, t_i/12) + \beta_3 t_i + \beta_4 t_i^2 + \beta_5 t_i^3 + \beta_6 t_i^4 + \varepsilon_i$

Coefficient	**Estimate**	**Std. Error**	**t-value**	**P-value**
Intercept	2.23E+00	1.76E-02	126.841	< 2.00E-16
$\mathbf{t_i}$	3.73E-03	1.55E-04	24.089	< 2.00E-16
$\mathbf{t_i^2}$	-1.55E-05	2.00E-06	-7.715	3.32E-13
$\mathbf{t_i^3}$	-9.19E-08	5.39E-09	-17.045	< 2.00E-16
$\mathbf{t_i^4}$	3.43E-10	5.10E-11	6.724	1.30E-10
$\mathbf{\cos(2\pi\, t_i/12)}$	-2.71E-01	1.31E-02	-20.753	< 2.00E-16
$\mathbf{\sin(2\pi\, t_i/12)}$	-2.75E-01	1.78E-02	-15.478	< 2.00E-16

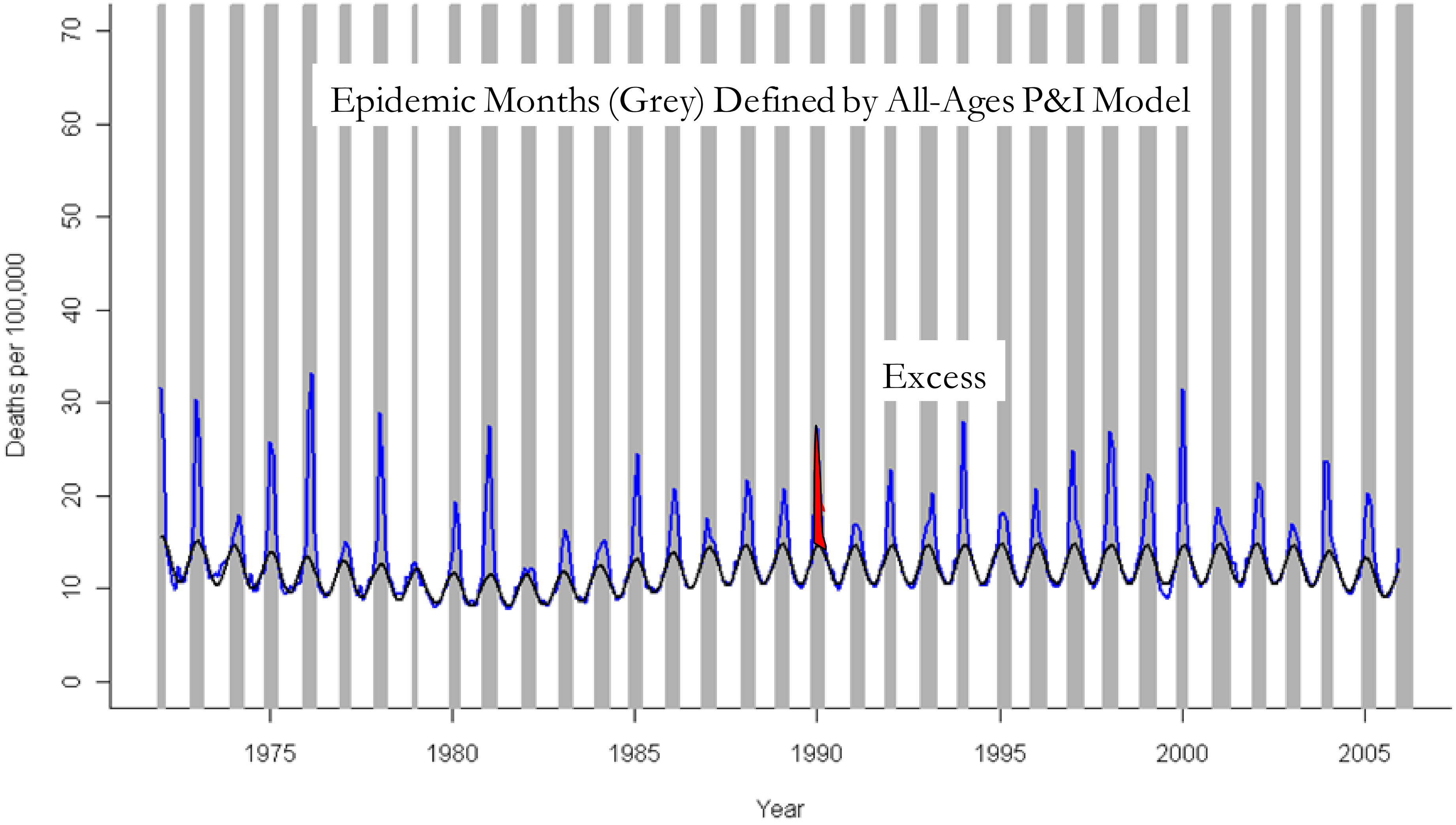
Multiple R-squared: **0.84**

Adjusted R-squared: **0.84**

F-statistic: 209.7 on 6 and 238 DF

– P-value: **< 2.2E-16**

USA P&I Deaths per 100,000 (1972-2006) and Model Baseline 65-89 Year Olds



Seasonal US Excess Mortality Table

Season	Age Group	No. of Epidemic Months	Excess P&I Deaths per 100,000	Excess A-C Deaths per 100,000
1977/1978	65-89	4	30.41	137.05
1978/1979	65-89	1	0.75	7.40
1979/1980	65-89	3	15.68	66.85
1980/1981	65-89	3	28.45	136.71
1981/1982	65-89	3	3.67	27.35
1982/1983	65-89	4	12.68	47.83
1983/1984	65-89	4	9.51	33.90
1984/1985	65-89	4	21.38	122.61
AVERAGES				
-	-	3 (median = 3, range = 1-5)	41.34	78.37

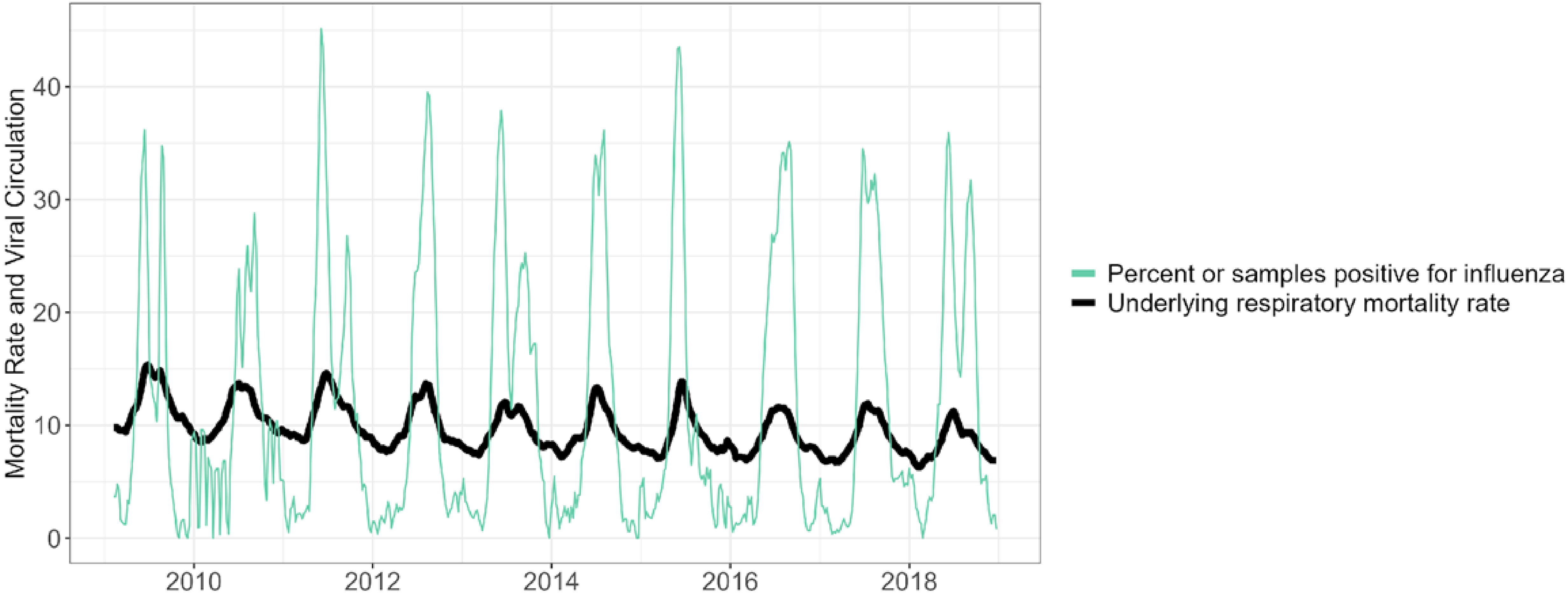
Pros and cons of Serfling approach

- Very flexible: can be used without virological data – especially useful for data on past pandemics
  - However, need at least 3 years of data
- Only works if disease is seasonal
  - Needs clear periods with no viral activity that can be used to create the baseline
  - Cannot be used as is for tropical countries that have year-round influenza circulation
    - There are techniques to a adapt Serfling models for these purposes
- Strong assumption that all excesses above the baseline are due to flu (co-circulation with non-flu pathogens an issue)

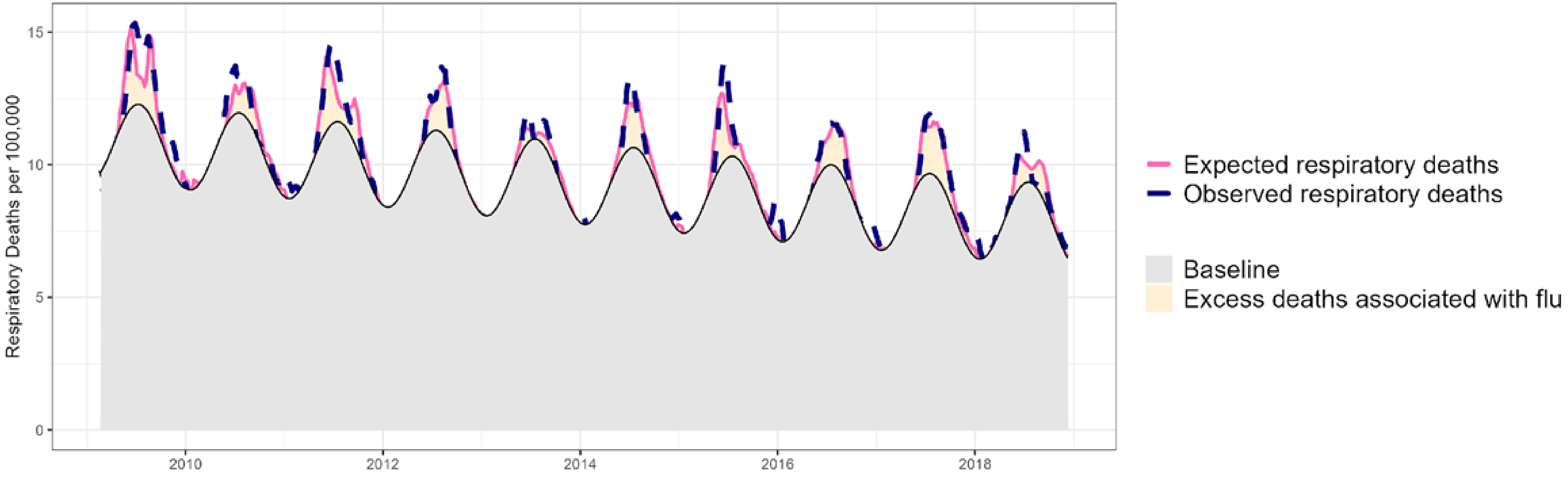
Example using data from South Africa

TIME SERIES DISEASE BURDEN MODELS DRIVEN BY VIRAL COVARIATES

Peaks in influenza viral activity drive peaks in respiratory mortality



Estimating influenza mortality in adults ≥ 60 years in South Africa



$$\text{Respiratory mortality rate }(t) = \alpha + \underbrace{\underbrace{\beta_1 t + \beta_2 t^2}_{\text{Linear \& non-linear trends}} + \beta_3 \sin\underbrace{(2t\pi/52.2) + \beta_4 \cos(2t\pi/52.2)}_{\text{Seasonal harmonic trends}}}_{\text{Baseline}} + \underbrace{\beta_5 flu + \beta_6 RSV}_{\text{Viral circulation}}$$

Estimating influenza and RSV burden

Fitted value:

$$Y1(t) = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 \sin(2t\pi/52.2) + \beta_4 \cos(2t\pi/52.2) + \beta_5 flu + \beta_6 RSV$$

Fitted value with flu term set to 0:

$$Y2(t) = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 \sin(2t\pi/52.2) + \beta_4 \cos(2t\pi/52.2) + \cancel{\beta_5 flu} + \beta_6 RSV$$

Flu burden = Y1(t) – Y2(t)

Fitted value with RSV term set to 0:

$$Y3(t) = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 \sin(2t\pi/52.2) + \beta_4 \cos(2t\pi/52.2) + \beta_5 flu + \cancel{\beta_6 RSV}$$

RSV burden = Y1(t) – Y3(t)

Variations of the same method

$$\text{Respiratory deaths } (t) = \alpha + \underbrace{\textbf{spline(t)}}_{\text{Natural cubic spline replaces terms for trends and seasonality}} + \beta_1 \text{flu} + \beta_2 \text{RSV}$$

Natural cubic spline replaces terms for trends and seasonality

$$\text{Respiratory deaths } (t) = \alpha + \text{Spline}(t) + \underbrace{\boldsymbol{\beta_1}\textbf{FluA/H1N1} + \boldsymbol{\beta_2}\textbf{FluA/H3N2} + \boldsymbol{\beta_3}\textbf{FluB}} + \beta_4 \text{RSV}$$

Individual terms for each flu subtype

$$\text{Respiratory deaths } (t) = \alpha\ \text{Spline}(t) + \underbrace{\boldsymbol{\beta_1}\textbf{Flu\_season1} + \boldsymbol{\beta_2}\textbf{Flu\_season2} + \boldsymbol{\beta_3}\textbf{Flu\_season3}} + \ldots + \beta_i \text{RSV}$$

Individual terms for each flu season

Poisson and Negative binomial regression frequently used when modeling count data

- *requires an offset term for population size and results are exponentiated*

Validity diagnostics for influenza disease burden models

- Regression diagnostics
  - AIC to choose the best model (lower is better)
  - Plotting difference between model and observed data (residuals) to detect outliers
- Checks based on the epidemiology of influenza
  - Flu A vs Flu B dominant seasons
  - RSV vs influenza (age!)
  - Higher rates in 65 yrs and over
  - Multiple years: seasons will little influenza circulation very precious!
- Difficult to estimate disease burden with precision
  - Mild seasons
  - Young children
  - Middle age groups

Examples from literature

HOW HAVE THESE METHODS BEEN USED IN AFRICA?

Stratifying by location of death (in vs. out-of-hospital)

Clinical Infectious Diseases

MAJOR ARTICLE

IDSA Infectious Diseases Society of America

hivma hiv medicine association

OXFORD

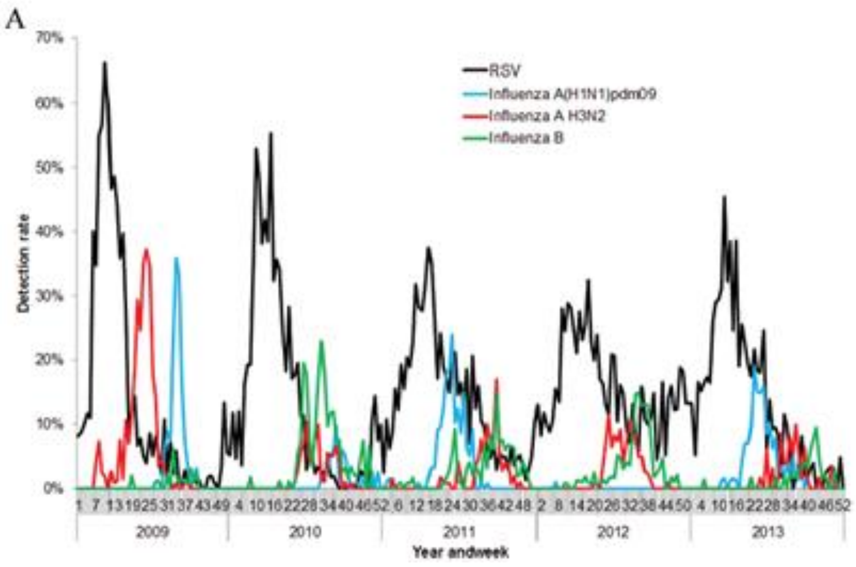
In- and Out-of-hospital Mortality Associated with Seasonal and Pandemic Influenza and Respiratory Syncytial Virus in South Africa, 2009–2013

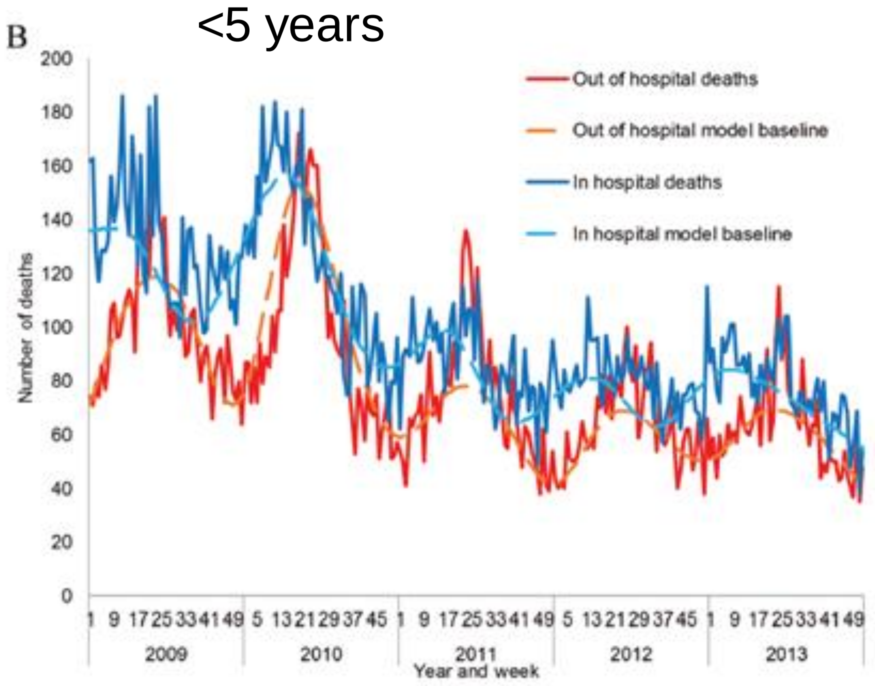
Cheryl Cohen,[1,2] Sibongile Walaza,[1,2] Florette K. Treurnicht,[1,2] Meredith McMorrow,[3,4,5] Shabir A. Madhi,[1,2,6] Johanna M. McAnerney,[1] and Stefano Tempia[3,4]

[1]Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, and [2]School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [3]Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia; [4]Influenza Program, Centers for Disease Control and Prevention, Pretoria, South Africa; [5]US Public Health Service, Rockville, Maryland; and [6]Department of Science and Technology/National Research Foundation: Vaccine Preventable Diseases, University of the Witwatersrand, Johannesburg, South Africa.

Respiratory mortality rates per 100,000 population

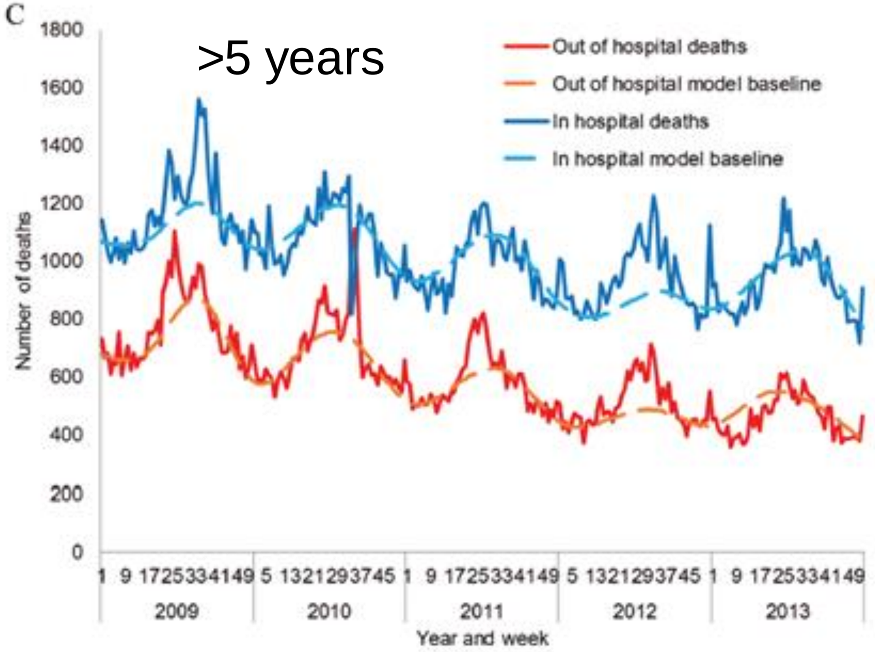
- Flu: 7.8 (all ages), highest in adults >75yrs, 63% out-of-hospital
- RSV: 3.1 (all ages), highest in infants <1 yr, 48% out-of-hospital





Used terms for season and subtype

$$
\begin{aligned}
E(\text{mortality rate}) = {} & \beta_0 + \beta_1 t + [\sum_{y=2009}^{2013} \beta_{2,y} \\
& (\text{Influenza A}(\text{H1N1})\text{pdm09})] + [\sum_{y=2009}^{2013} \beta_{3,y} \\
& (\text{Influenza A}(\text{H3N2})] + [\sum_{y=2009}^{2013} \beta_{4,y} \\
& (\text{Influenza B})] + [\sum_{y=2009}^{2013} \beta_{5,y}(\text{RSV})] + \text{spline}(t)
\end{aligned}
$$



Alternative data sources: Health and Demographic Surveillance Sites

PLOS | ONE

RESEARCH ARTICLE

Estimating influenza and respiratory syncytial virus-associated mortality in Western Kenya using health and demographic surveillance system data, 2007-2013

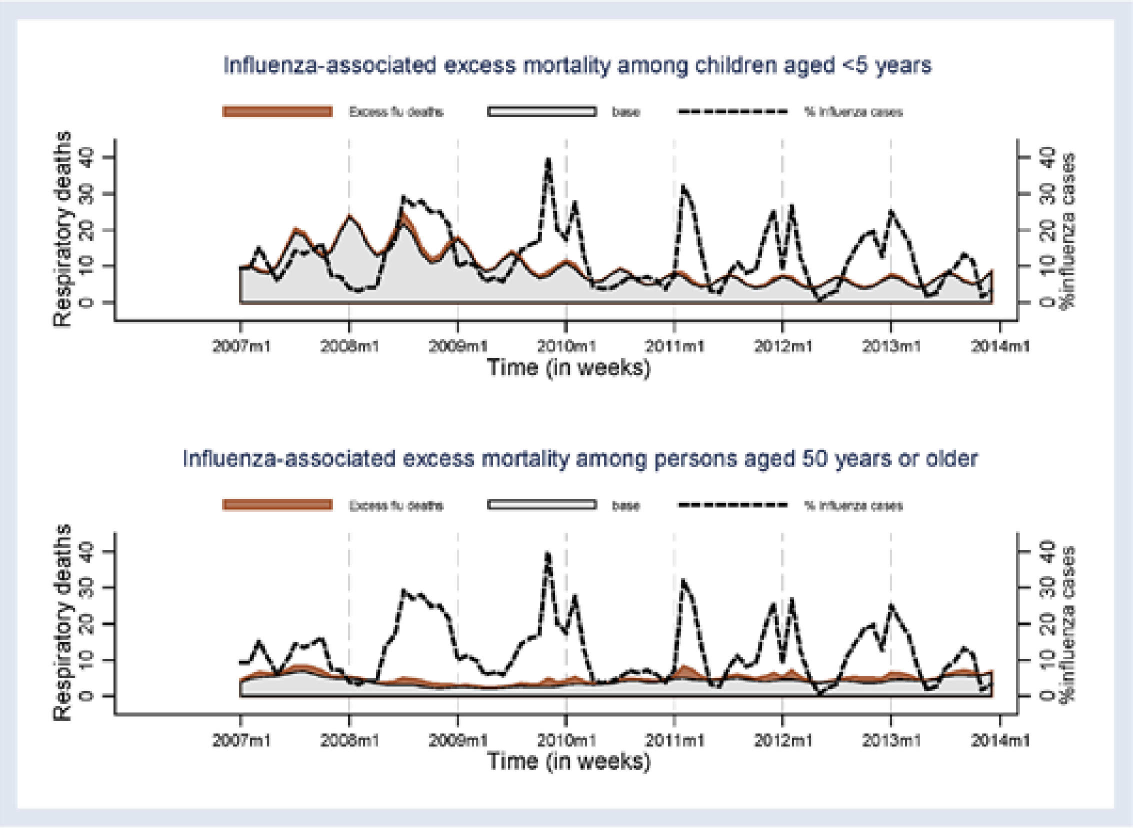
Gideon O. Emukule[1,2]*, Peter Spreeuwenberg[3], Sandra S. Chaves[1,4], Joshua A. Mott[1,4,5], Stefano Tempia[4,6], Godfrey Bigogo[7], Bryan Nyawanda[7], Amek Nyaguara[7], Marc-Alain Widdowson[1], Koos van der Velden[2], John W. Paget[2,3]

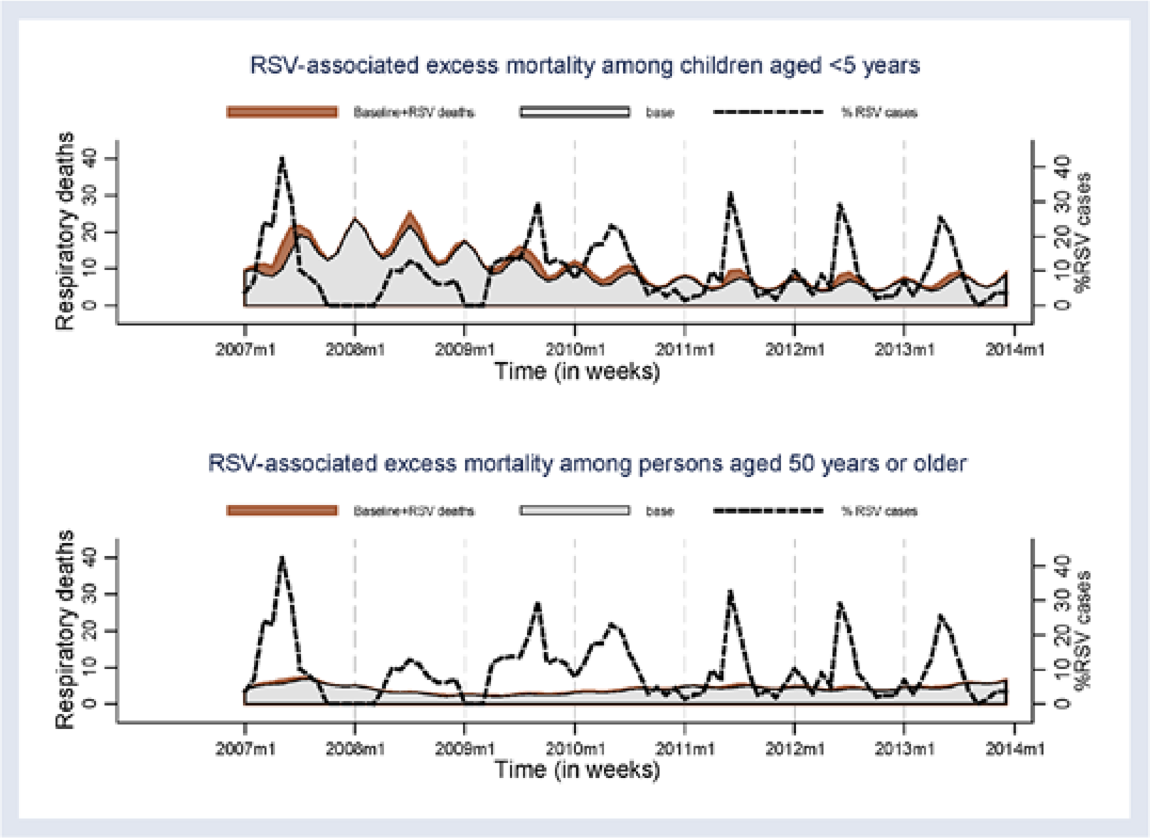
Respiratory mortality rates per 100,000 populations

- Flu: 10.5 (all ages), highest in adults >50 yrs
- RSV: 7.3 (all ages), highest in children <5 yrs

$$E(Y_t) = \beta_0 + \beta_1(t) + \beta_2(t^2) + \ldots + \beta_6(t^6) + \beta_7[\sin(2t\pi/12)] + \beta_8[\cos(2t\pi/12)] + \beta_9[\sin(2t\pi/6)]$$
$$+\beta_{10}[\cos(2t\pi/6)] + \beta_{11}[\sin(2t\pi/3)] + \beta_{12}[\cos(2t\pi/3)] + \beta_{13}[Influenza]$$
$$+\beta_{14}[RSV] + \beta_{15}[Malaria] + \varepsilon_t$$

Included multiple terms for trends and adjusted for malaria





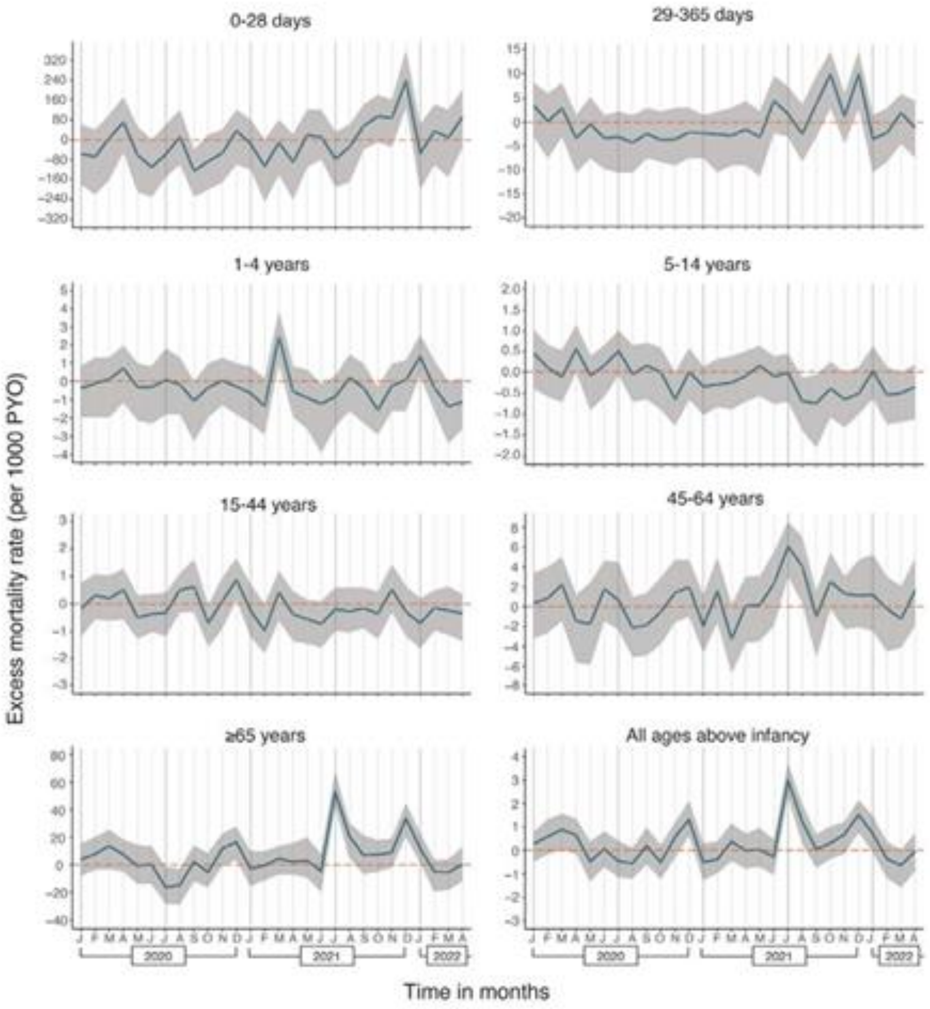
Estimating the burden of COVID-19 mortality – Kenya

Impact of COVID-19 on mortality in coastal Kenya: a longitudinal open cohort study

Authors and affiliations

Otiende M[1*],MS; Nyaguara A[1],PhD; Bottomley C[2],PhD; Walumbe D[1],MS; Mochamah G[1], MS; Amadi D[1], MS; Nyundo C[1], MS; Kagucia EW[1], PhD; Etyang AO[1], PhD; Adetifa IMO[1,2], PhD; Brand SPC[3], PhD; Maitha E[4], MPH; Chondo E[4], BPharm; Nzomo E[5], MBChB; Aman R[6], PhD; Mwangangi M[6],MS; Amoth P[6], PhD; Kasera K[6], MS; Ng'ang'a W[7], MS; Barasa E[1], PhD; Tsofa B[1], PhD; Mwangangi J[1], PhD; Bejon P[1,8], PhD; Agweyu A[1], PhD; Williams TN[1,9], PhD; Scott JAG[1,2], FRCP

- All cause excess mortality 47.5 per 100,000 populations
- Highest burden in adults >65 yrs
- Negative excess mortality in children 1-14 yrs

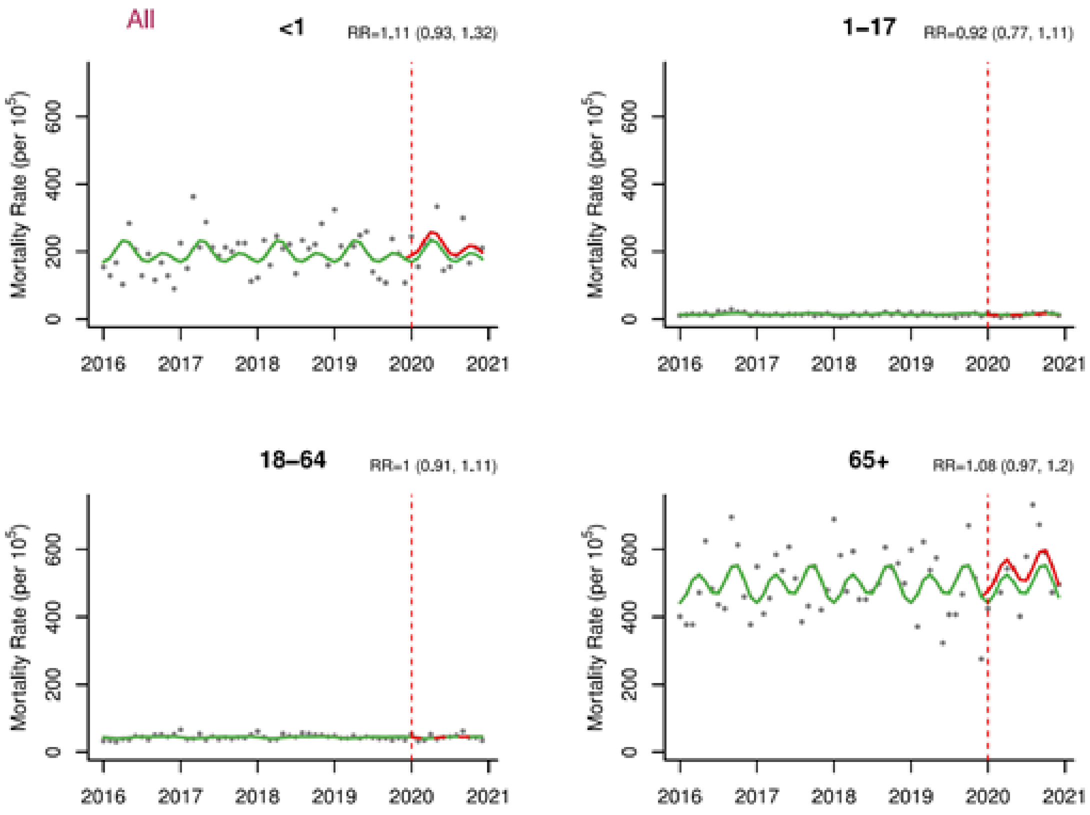


Estimating the burden of COVID-19 mortality – The Gambia

Contents lists available at ScienceDirect
International Journal of Infectious Diseases
journal homepage: www.elsevier.com/locate/ijid

Quantifying excess mortality during the COVID-19 pandemic in 2020 in The Gambia: a time-series analysis of three health and demographic surveillance systems

Nuredin I. Mohammed[1,*], Grant Mackenzie[1,2,3], Esu Ezeani[1], Mamadi Sidibeh[1], Lamin Jammeh[1], Golam Sarwar[1], Aji Kumba Folawiyo Saine[1], Bakary Sonko[1], Pierre Gomez[1], Bai Lamin Dondeh[1], M. Jahangir Hossain[1], Momodou Jasseh[1], Effua Usuf[1], Andrew M. Prentice[1], David Jeffries[1], Umberto Dalessandro[1], Anna Roca[1,*]



Did not find significant excess mortality overall, but results varied by site

Applied these methods to hospitalization data during the COVID-19 pandemic

Clinical Infectious Diseases

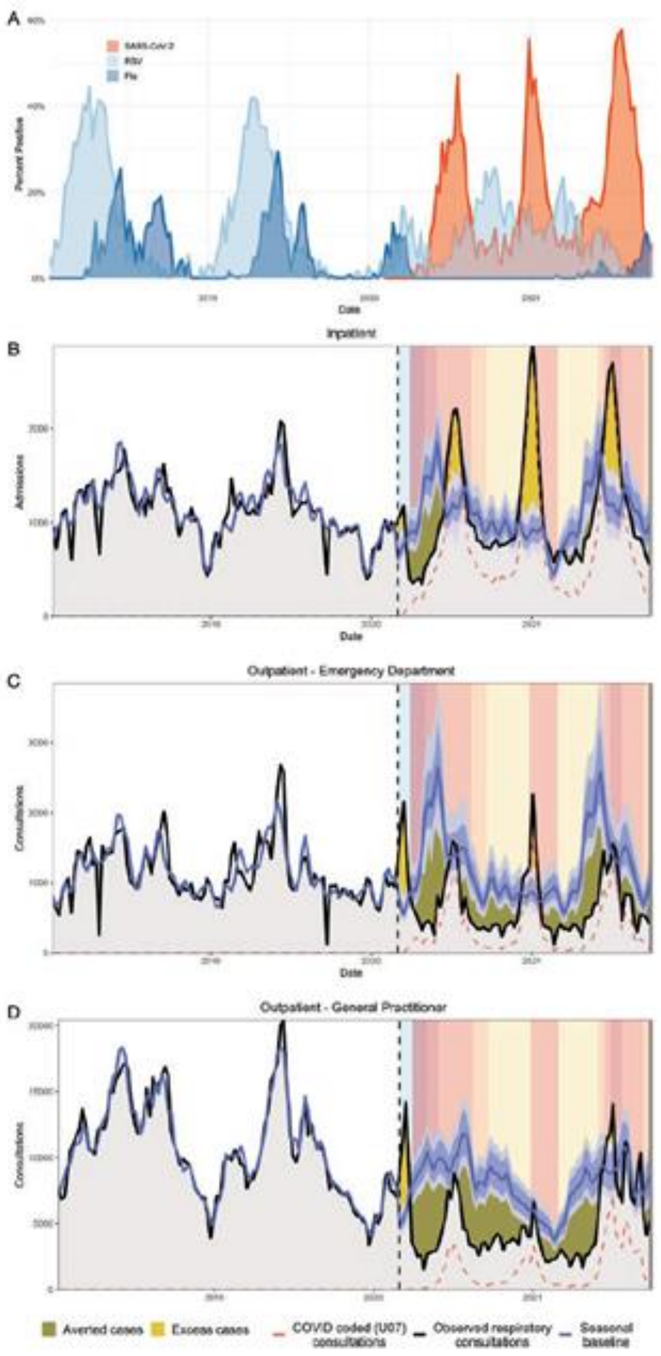
MAJOR ARTICLE

IDSA Infectious Diseases Society of America hivma hiv medicine association OXFORD

Direct and Indirect Effects of the Coronavirus Disease 2019 Pandemic on Private Healthcare Utilization in South Africa, March 2020–September 2021

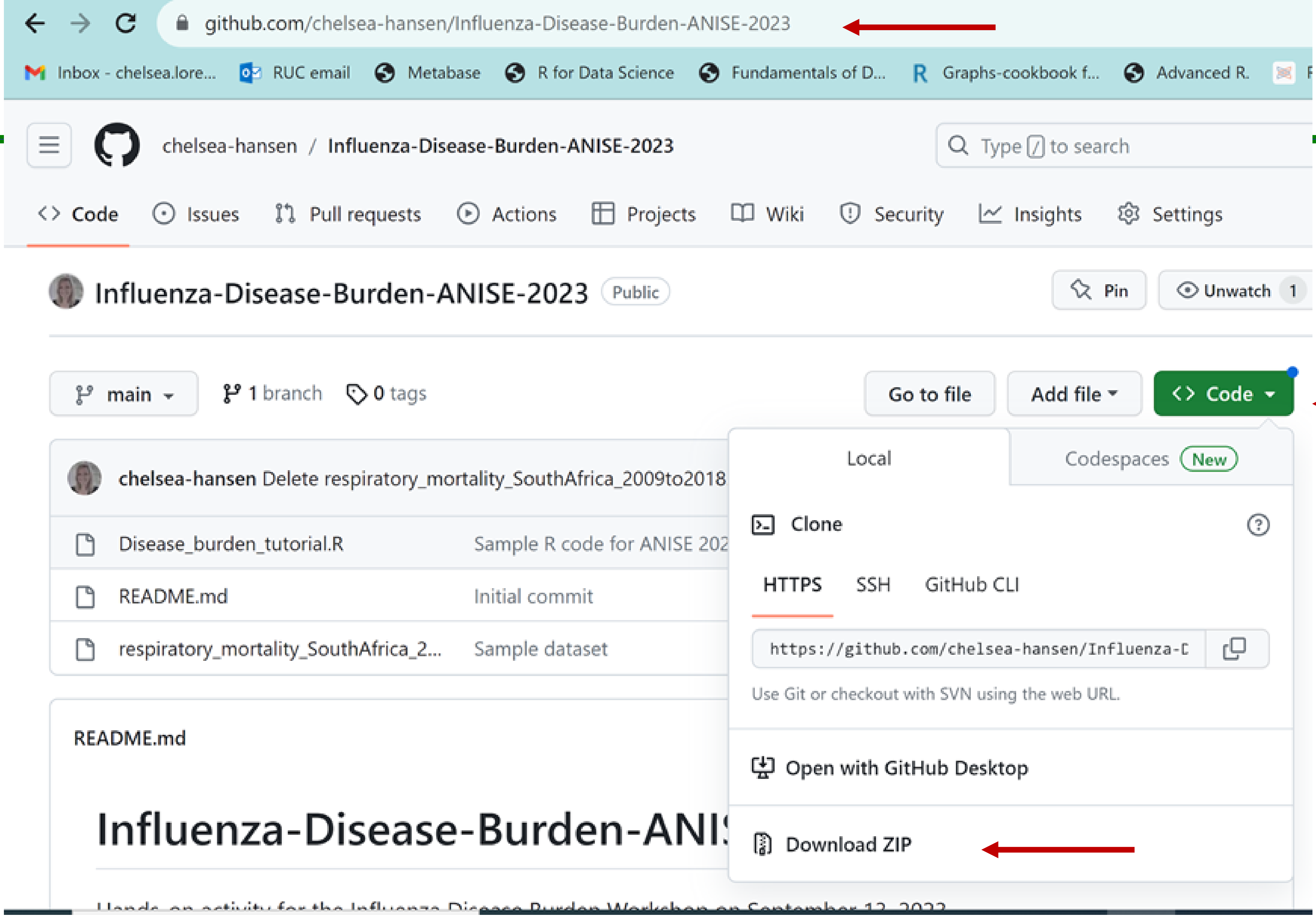
Amanda C. Perofsky,[1,6] Stefano Tempia,[2,3] Jeremy Bingham,[4] Caroline Maslo,[5] Mande Toubkin,[6] Anchen Laubscher,[5] Sibongile Walaza,[2,3] Juliet R. C. Pulliam,[4] Cécile Viboud,[1] and Cheryl Cohen[2,3]

- Hospitalization data rather than mortality data
- Identified 3 waves of excess respiratory admission consistent with SARS-CoV-2 waves
- Observed declines in intestinal, non-COVID respiratory, and non-respiratory admissions



Practical Demonstration using Mortality Data from South Africa

DISEASE BURDEN MODELS IN R

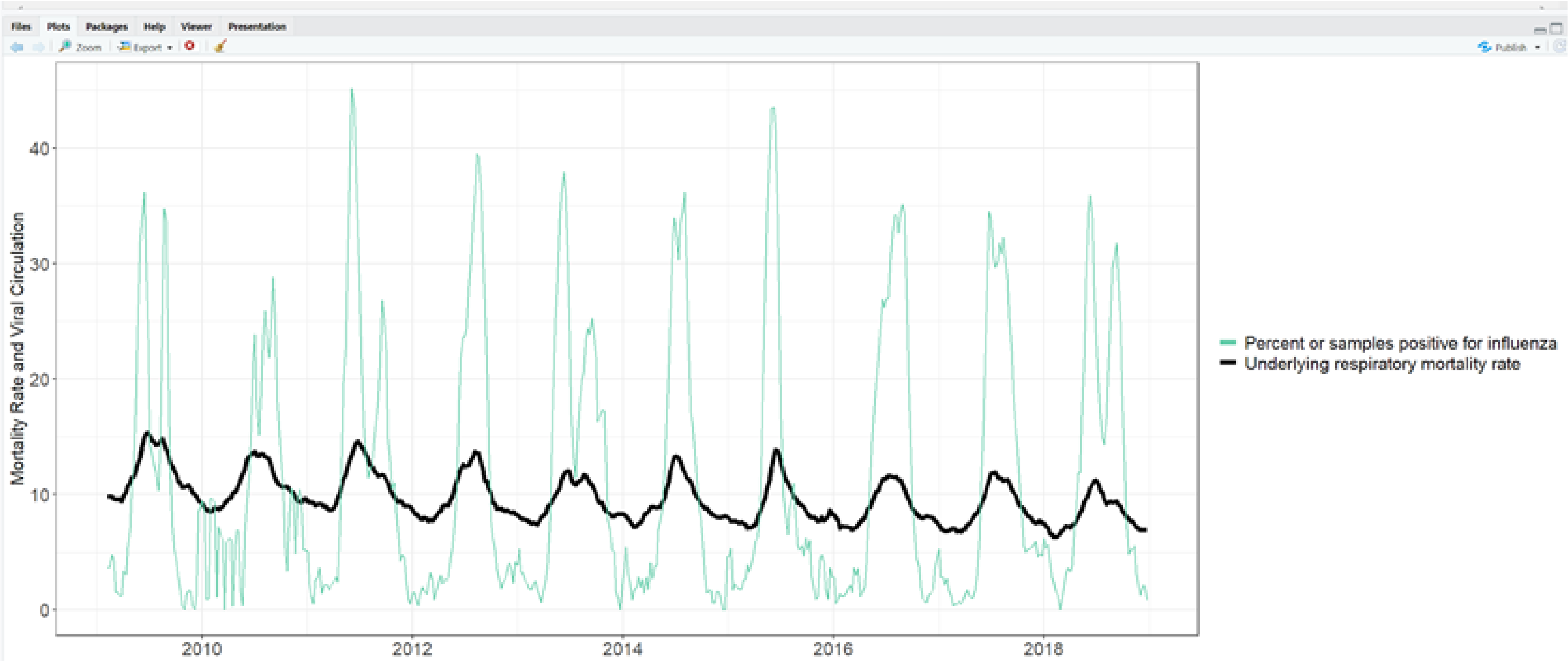


github.com/chelsea-hansen/Influenza-Disease-Burden-ANISE-2023
Inbox - chelsea.lore...
RUC email
Metabase
R for Data Science
Fundamentals of D...
Graphs-cookbook f...
Advanced R.
chelsea-hansen / Influenza-Disease-Burden-ANISE-2023
Type / to search
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
Influenza-Disease-Burden-ANISE-2023
Public
Pin
Unwatch 1
main
1 branch
0 tags
Go to file
Add file
Code
chelsea-hansen Delete respiratory_mortality_SouthAfrica_2009to2018
Disease_burden_tutorial.R
Sample R code for ANISE 202
README.md
Initial commit
respiratory_mortality_SouthAfrica_2...
Sample dataset
README.md
Influenza-Disease-Burden-ANIS
Local
Codespaces
New
Clone
HTTPS
SSH
GitHub CLI
https://github.com/chelsea-hansen/Influenza-D
Use Git or checkout with SVN using the web URL.
Open with GitHub Desktop
Download ZIP

```r
#rm(list=ls()) #run this to clear your environment

#upload packages
library(splines)
library(tidyverse)
library(zoo)

#set working directory - this is where we will save everything
setwd("C:/Users/hansencl/Desktop/ANISE 2023 activity")


#read in the dataset we will use
dat = read.csv("respiratory_mortality_SouthAfrica_2009to2018.csv") %>%
  mutate(resp_rate = rollmean(resp_uc/population*100000,k=5,align="center",fill="extend"))#convert to a mortality rate and smooth


## check the dates variable to make sure R understand it as a date
class(dat$date) #reads as a character
dat$date = as.Date(dat$date, format = '%m/%d/%Y') #convert to date

#check again
class(dat$date) #now it is a date

#have a look at the data
str(dat)
head(dat)
tail(dat)
View(dat)
```

	date	agegrp	pi_uc	resp_uc	resp_cardio_uc	natural	population	perc_pos_flu	perc_pos_fluAh1n1	perc_pos_fluAh3n2	perc_pos_fluB	perc_pos_rsv	resp_rate
1	2009-02-08	60+ yrs	169	394	1351	3282	3923521	3.678693e-02	0.000000e+00	3.678693e-02	0.000000e+00	4.336099e-01	9.797322
2	2009-02-15	60+ yrs	159	392	1439	3417	3923521	3.678693e-02	0.000000e+00	3.678693e-02	0.000000e+00	4.336099e-01	9.797322
3	2009-02-22	60+ yrs	147	356	1284	3051	3923521	4.777594e-02	0.000000e+00	4.777594e-02	0.000000e+00	4.577857e-01	9.797322
4	2009-03-01	60+ yrs	158	374	1384	3279	3923521	4.336140e-02	0.000000e+00	4.336140e-02	0.000000e+00	5.549057e-01	9.578131
5	2009-03-08	60+ yrs	177	406	1339	3205	3923521	1.596414e-02	0.000000e+00	1.596414e-02	0.000000e+00	5.632079e-01	9.527157
6	2009-03-15	60+ yrs	143	351	1259	3122	3923521	1.410754e-02	0.000000e+00	1.410754e-02	0.000000e+00	5.792148e-01	9.542449
7	2009-03-22	60+ yrs	163	382	1340	3247	3923521	1.216272e-02	0.000000e+00	1.216272e-02	0.000000e+00	5.371523e-01	9.608716
8	2009-03-29	60+ yrs	154	359	1306	3172	3923521	1.216272e-02	0.000000e+00	1.216272e-02	0.000000e+00	4.812601e-01	9.394623
9	2009-04-05	60+ yrs	168	387	1299	3176	3923521	3.333333e-02	0.000000e+00	3.333333e-02	0.000000e+00	4.289562e-01	9.868687
10	2009-04-12	60+ yrs	151	364	1321	3168	3923521	3.030303e-02	0.000000e+00	3.030303e-02	0.000000e+00	3.986532e-01	10.235704
11	2009-04-19	60+ yrs	188	444	1405	3266	3923521	5.445762e-02	0.000000e+00	5.445762e-02	0.000000e+00	3.741765e-01	10.699573
12	2009-04-26	60+ yrs	191	454	1535	3475	3923521	7.045089e-02	0.000000e+00	7.045089e-02	0.000000e+00	2.900015e-01	11.086980
13	2009-05-03	60+ yrs	176	450	1553	3556	3923521	1.093398e-01	0.000000e+00	1.037842e-01	5.555556e-03	1.607085e-01	11.459095
14	2009-05-10	60+ yrs	206	463	1532	3540	3923521	1.230640e-01	0.000000e+00	1.175084e-01	5.555556e-03	1.047138e-01	11.530460
15	2009-05-17	60+ yrs	183	437	1588	3636	3923521	1.834343e-01	0.000000e+00	1.778788e-01	5.555556e-03	9.434343e-02	12.019816
16	2009-05-24	60+ yrs	199	458	1622	3636	3923521	2.459947e-01	0.000000e+00	2.459947e-01	0.000000e+00	9.627580e-02	12.580537
17	2009-05-31	60+ yrs	228	550	1899	4033	3923521	3.192271e-01	0.000000e+00	3.192271e-01	0.000000e+00	7.066253e-02	13.222817
18	2009-06-07	60+ yrs	242	560	1847	4207	3923521	3.375604e-01	0.000000e+00	3.375604e-01	0.000000e+00	5.590062e-02	14.089386
19	2009-06-14	60+ yrs	249	589	1837	4093	3923521	3.617942e-01	0.000000e+00	3.617942e-01	0.000000e+00	5.906844e-02	14.991636
20	2009-06-21	60+ yrs	274	607	1870	4026	3923521	3.013160e-01	0.000000e+00	3.013160e-01	0.000000e+00	6.032421e-02	15.277094
21	2009-06-28	60+ yrs	284	635	2063	4474	3923521	2.346493e-01	0.000000e+00	2.346493e-01	0.000000e+00	7.758612e-02	15.338264
22	2009-07-05	60+ yrs	266	606	1979	4252	3923521	1.426741e-01	0.000000e+00	1.330433e-01	9.630819e-03	6.524499e-02	15.088488
23	2009-07-12	60+ yrs	281	572	1878	4029	3923521	1.362179e-01	0.000000e+00	1.265871e-01	9.630819e-03	5.619613e-02	14.695984
24	2009-07-19	60+ yrs	215	540	1862	4118	3923521	1.248931e-01	2.350427e-02	9.175803e-02	9.630819e-03	3.984998e-02	14.512475

```r
#make a figure to show respiratory mortality and influenza circulation - plot from slides
plot1 = ggplot(data=dat)+
  theme_bw()+
  geom_line(aes(x=date, y=resp_rate,color="Underlying respiratory mortality rate"),size=2)+
  geom_line(aes(x=date, y=perc_pos_flu*100, color="Percent or samples positive for influenza"))+
  labs(x=NULL, y="Mortality Rate and Viral Circulation")+
  scale_color_manual(name=NULL, values=c("aquamarine3","black"))+
  theme(legend.text = element_text(size=15),
        axis.text = element_text(size=15),
        axis.title = element_text(size=15))
plot1 #have a look at the plot
#save the plot
ggsave(plot=plot1, "mortality and viral time series.png",height=5,width=13, units="in")
```



```r
#Make a separate variable for flu and RSV for each season
dat = dat %>%
  mutate(season = year(date),
         flu = perc_pos_flu,
         rsv = perc_pos_rsv) %>%
  pivot_wider(names_from = season, values_from=c("flu", "rsv"), values_fill=0)

#make a flu season matrix
flu_matrix = as.matrix(dat[1:nrow(dat),c(grep("flu_",names(dat)))])
View(flu_matrix)
```

Filter

	flu_2009	flu_2010	flu_2011	flu_2012	flu_2013	flu_2014	flu_2015	flu_2016	flu_2017	flu_2018
1	3.678693e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
2	3.678693e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
3	4.777594e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
4	4.336140e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
5	1.596414e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
6	1.410754e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
7	1.216272e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
8	1.216272e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
9	3.333333e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
10	3.030303e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
11	5.445762e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
12	7.045089e-02	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
13	1.093398e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
14	1.230640e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
15	1.834343e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
16	2.459947e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
17	3.192271e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
18	3.375604e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
19	3.617942e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
20	3.013160e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
21	2.346493e-01	0.000000e+00	0.000000000	0.000000000	0.000000e+00	0.000000000	0	0	0	0
499	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	3.031221e-01
500	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	3.174962e-01
501	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	2.848957e-01
502	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	2.504986e-01
503	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	1.856740e-01
504	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	1.292377e-01
505	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	7.695055e-02
506	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	4.808673e-02
507	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	5.245181e-02
508	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	5.158496e-02
509	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	5.548405e-02
510	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	3.167452e-02
511	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	2.008032e-02
512	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	1.265823e-02
513	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	2.059474e-02
514	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	2.059474e-02
515	0	0	0	0	0.000000e+00	0.000000e+00	0.000000000	0.000000000	0.000000000	7.936508e-03

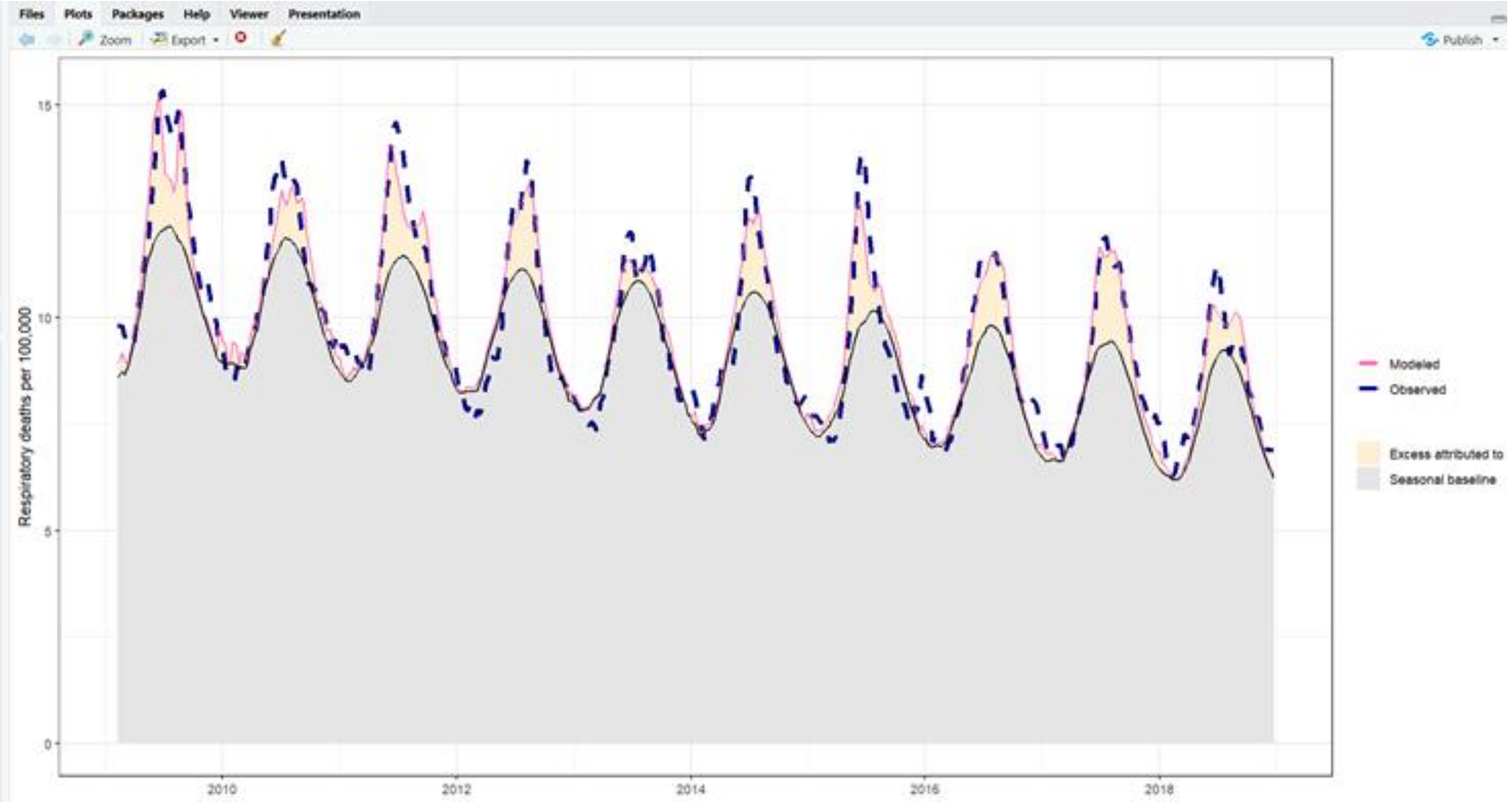
Showing 484 to 515 of 515 entries, 10 total columns

```
# Seasonal harmonic regression with covariates for flu and rsv --------
weeknum=seq(1,dim(dat)[1])
cos1= cos(2*pi/52.17*weeknum)
sin1= sin(2*pi/52.17*weeknum)

#single covariate for flu and single covariate for rsv
model_1 = glm(dat$resp_rate~ weeknum + weeknum^2 + cos1 + sin1  + dat$perc_pos_flu + dat$perc_pos_rsv, family=gaussian(link="identity"))
summary(model_1)



#try the same model with a covariate for each flu season
model_2 = glm(dat$resp_rate~ weeknum + weeknum^2 + cos1 + sin1  + flu_matrix + dat$perc_pos_rsv, family=gaussian(link="identity"))
summary(model_2)



#try a coviariate for each flu subtype
model_3 = glm(dat$resp_rate~ weeknum + weeknum^2 + cos1 + sin1  + dat$perc_pos_fluAh1n1 + dat$perc_pos_fluAh3n2 + dat$perc_pos_fluB + dat$perc_pos_rsv, family=gaussian(link="identity"))
summary(model_3)


#compare AIC
AIC(model_1)#single flu term
AIC(model_2)#term for each flu season
AIC(model_3)#term for each flu subtype
```

```
> #compare AIC
> AIC(model_1)#single flu term
[1] 977.2667
> AIC(model_2)#term for each flu season
[1] 880.5742
> AIC(model_3)#term for each flu subtype
[1] 932.4801
>
```

Which model is the best?

	pop	excess.flu.rate.estimated	excess.rsv.rate.estimated	baseline	observed	predicted	season
2009	3923521	45.26520	-15.095414	502.6307	547.2340	532.8005	2009
2010	4024870	28.35326	-11.643074	542.1487	559.6234	558.8589	2010
2011	4140047	44.26559	-10.107334	525.2288	561.7588	559.3871	2011
2012	4264528	30.26849	-12.466530	516.4708	519.3623	534.2728	2012
2013	4391538	14.88264	-9.315801	491.0736	490.7022	496.6404	2013
2014	4527167	28.05373	-9.378393	474.1540	488.1267	492.8294	2014
2015	4665551	35.95456	-10.461340	457.2344	473.9021	482.7276	2015
2016	4806961	31.42475	-12.882069	440.3146	464.5485	458.8573	2016
2017	4950886	41.11200	-11.712566	429.9153	463.2480	459.3148	2017
2018	5098435	27.04145	-12.871752	399.9543	421.3066	414.1240	2018



Please join for the afternoon session for hands-on practice with this exercise!