MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Inference for epidemic models

Daniela De Angelis

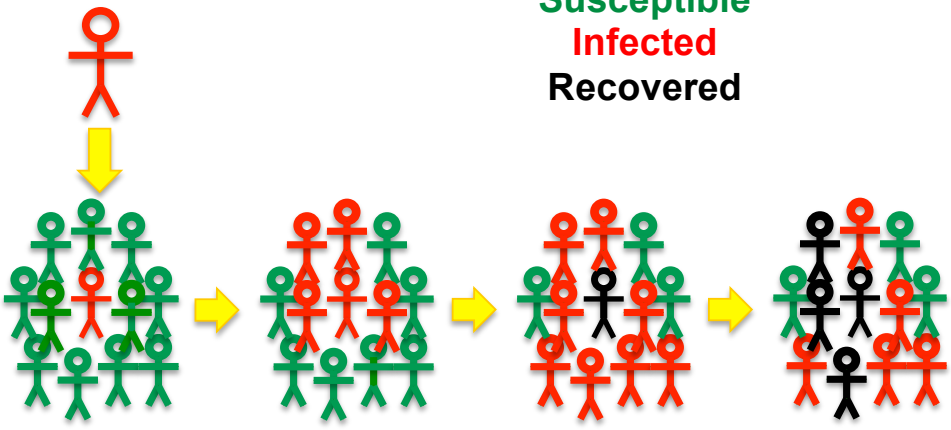
MRC Biostatistics Unit, University of Cambridge

NORDITA Workshop, Stockholm, 29 May 2023

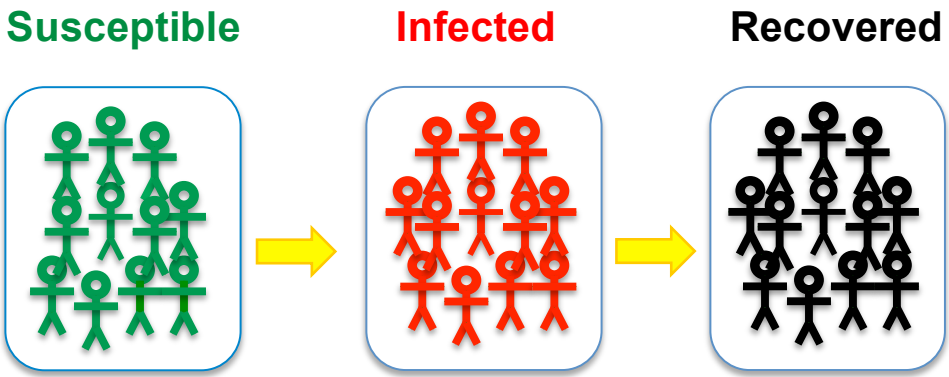
Outline

- Compartmental models of disease dynamics
  - Deterministic/stochastic
  - Key quantities of interest
- Linking models to data
  - Types of data
- Inference approaches - different types of data/models
- Illustration from our COVID19 work

Infectious disease spread

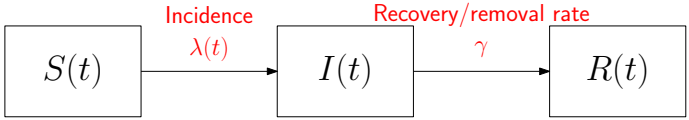


Infectious disease spread



Simple SIR-type models

SIR model:



Delay infectiousness through the addition of a latent infection state: SEIR model

Perhaps infection does not confer lasting immunity: SIS model

$\exists$ many more variations!

Deterministic SIR model [Kermack & McKendrick (1927)]

- Closed population of size $S(t) + I(t) + R(t) = N + 1$
- Initial state $X(0) = (S(0), I(0), R(0)) = (N, 1, 0)$, ODEs:

$$\begin{aligned} \frac{d}{dt} S(t) &= -\lambda(t) S(t) \\ \frac{d}{dt} I(t) &= \lambda(t) S(t) - \gamma I(t) \\ \frac{d}{dt} R(t) &= \gamma I(t) \end{aligned}$$

where *mass action/homogeneous mixing* assumption holds:

$$\begin{aligned} \lambda(t) &= \beta \times I(t) \\ \text{incidence} &= \text{effective contact rate} \times \text{\# infected} \end{aligned}$$

The reproduction number R_0

At time, $t = 0$, for the epidemic to take off we require:

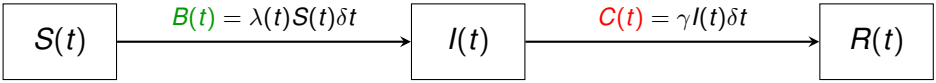
$$\frac{d}{dt} I(t)|_{t=0} > 0 \quad \Rightarrow \quad S(0) = N > \gamma/\beta$$

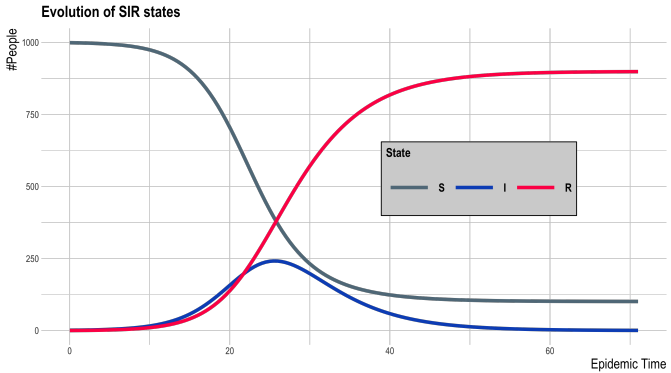
Definition

Let $R_0 := N\beta/\gamma$.
Then the epidemic will not immediately begin to die out if, at $t = 0$, $R_0 > 1$.

Interpretation of R_0:
The number of secondary infections caused by one primary infection in a fully susceptible population.

A Deterministic Epidemic





Simulation Details

- $N = 1000$
- $R_0 = 2.5$
- $\gamma = 0.2\text{days}^{-1}$
- $\delta = 0.5$ day

Deterministic vs. Stochastic

Deterministic

- Model states $X(t) = (S(t), I(t))$ and transitions between them are a deterministic function of time, t, and a parameter $\boldsymbol{\theta} = (\beta, \gamma, \ldots)$.
- $R_0 > 1$ or equivalently $S_0 > \gamma/\beta$ and the epidemic will take-off with certainty.

Stochastic

- No longer a 1-1 relationship between parameter and epidemic; $X_t = \{S_t, I_t\}$ is a stochastic process, dependent on $\boldsymbol{\theta}$, not a function.
- Allows for the possibility of epidemics with $R_0 > 1$ to fail.
- For large I_t, dynamics typically approximate deterministic dynamics.
- $\exists$ many ways to incorporate stochasticity.

General Stochastic (SIR) Epidemic model

- Continuous-time Markov Chain - The SIR model on a closed population can be cast as a bivariate stochastic process $X_t = \{S_t, I_t\}$.

General Stochastic (SIR) Epidemic model

- Continuous-time Markov Chain - The SIR model on a closed population can be cast as a bivariate stochastic process $X_t = \{S_t, I_t\}$.
- Time spent in model compartments is exponentially distributed with transition rates

$$\begin{aligned} (s, i) \to (s-1, i+1) &: \beta i \\ (s, i) \to (s, i-1) &: \gamma \end{aligned}$$

General Stochastic (SIR) Epidemic model

- Continuous-time Markov Chain - The SIR model on a closed population can be cast as a bivariate stochastic process $X_t = \{S_t, I_t\}$.
- Time spent in model compartments is exponentially distributed with transition rates

$$\begin{aligned} (s, i) \to (s-1, i+1) &: \beta i \\ (s, i) \to (s, i-1) &: \gamma \end{aligned}$$

- and transition probabilities

$$\begin{aligned} \mathbb{P}\{S_{t+\delta t} = s-1, I_{t+\delta t} = i+1 \mid S_t = s, I_t = i\} &= \beta i s \delta t + o(\delta t) \\ \mathbb{P}\{S_{t+\delta t} = s, I_{t+\delta t} = i-1 \mid S_t = s, I_t = i\} &= \gamma i \delta t + o(\delta t) \end{aligned}$$

General Stochastic (SIR) Epidemic model

- Continuous-time Markov Chain - The SIR model on a closed population can be cast as a bivariate stochastic process $X_t = \{S_t, I_t\}$.
- Time spent in model compartments is exponentially distributed with transition rates

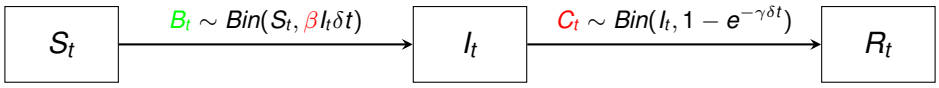
$$\begin{aligned} (s, i) \to (s-1, i+1) &: \beta i \\ (s, i) \to (s, i-1) &: \gamma \end{aligned}$$

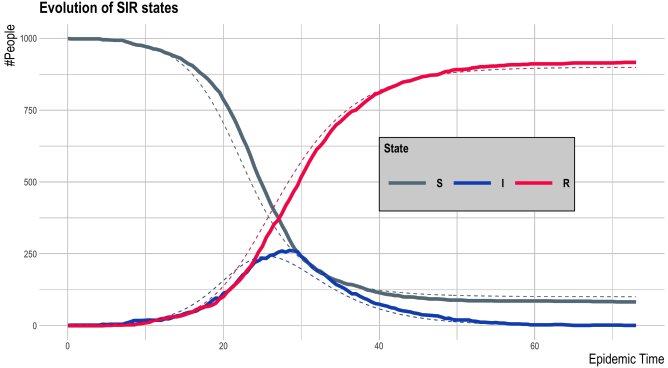
- and transition probabilities

$$\begin{aligned} \mathbb{P}\{S_{t+\delta t} = s-1, I_{t+\delta t} = i+1 \mid S_t = s, I_t = i\} &= \beta i s \delta t + o(\delta t) \\ \mathbb{P}\{S_{t+\delta t} = s, I_{t+\delta t} = i-1 \mid S_t = s, I_t = i\} &= \gamma i \delta t + o(\delta t) \end{aligned}$$

- We refer to this randomness as demographic stochasticity.

Chain-binomial models: In general

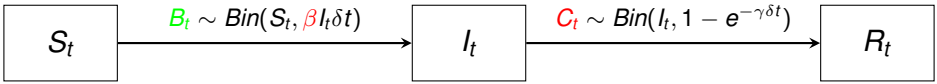


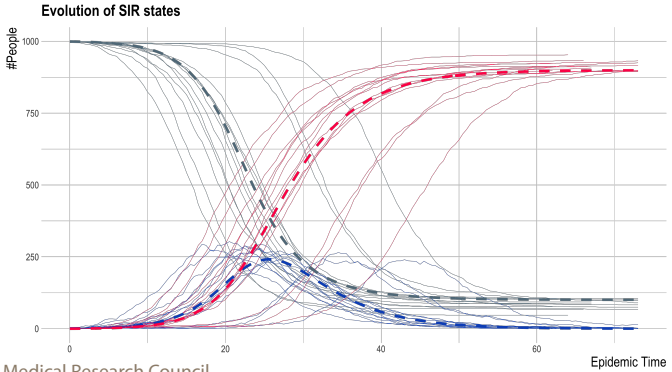


Simulation Details

- $N = 1000$
- $R_0 = 2.5$
- $\gamma = 0.2$ days^{-1}
- $\delta = 0.5$ day

Chain-binomial models: In general



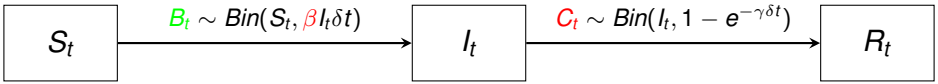


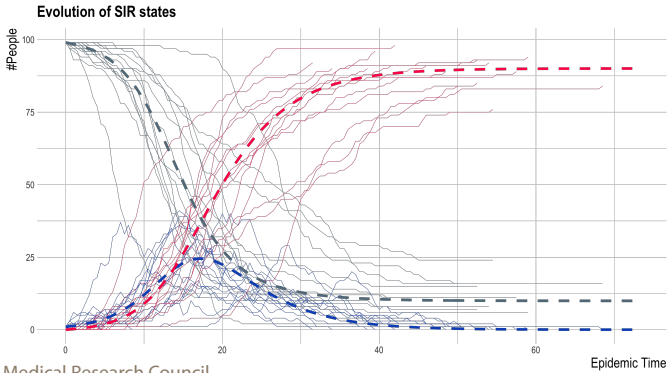
Simulation Details

- $N = 1000$
- $R_0 = 2.5$
- $\gamma = 0.2$ days^{-1}
- $\delta = 0.5$ day

Epidemic timing the main difference

Chain-binomial models: In general





Simulation Details

- $N = 100$
- $R_0 = 2.5$
- $\gamma = 0.2$ days^{-1}
- $\delta = 0.5$ day

Clearer stochastic effects.

Environmental Stochasticity

Assume stochastic fluctuation in the rate of effective contact, β_t, absorbing all extraneous, un-modelled effects on transmission:

e.g. the SIR model

$$\begin{aligned} dS_t/dt &= -\beta_t I_t S_t \\ dI_t/dt &= I_t(\beta_t S_t - \gamma) \\ d\beta_t &= \beta_t \nu dW_t \end{aligned}$$

- W_t is a standard Brownian motion; ν is a volatility parameter
- $\beta_t = \beta \exp(\nu W_t)$ is the *instantaneous rate of secondary infections per susceptible per unit time*.
- $X_t = (S_t, I_t, \beta_t)$ is now the extended state vector, a solution of the above stochastic differential equation.

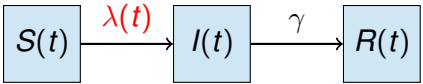
Discrete-time Environmental Stochasticity

In practice, in discrete time with time-steps of size δt

e.g. the SIR model

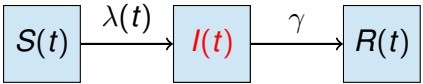
$$
\begin{aligned}
S_{t+\delta t} - S_t &= -\beta_t I_t S_t \delta_t \\
I_{t+\delta t} - I_t &= I_t \left(\beta_t S_t - \gamma\right) \delta t \\
\log\left(\beta_{t+\delta t}\right) &\sim \mathrm{N}\left(\log\left(\beta_t\right), \nu^2 \delta t\right)
\end{aligned}
$$

Key quantities to monitor epidemics



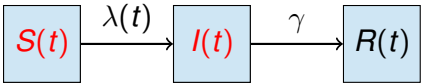
- Incidence or force of infection $\lambda(t) = \beta I(t)$ or $\lambda(t) = \beta I(t)/N$

Key quantities to monitor epidemics



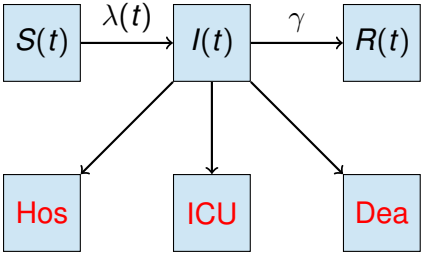
- Incidence or force of infection $\lambda(t) = \beta I(t)$ or $\lambda(t) = \beta I(t)/N$
- Prevalence $\pi = I(t)/N$

Key quantities to monitor epidemics



- Incidence or force of infection $\lambda(t) = \beta I(t)$ or $\lambda(t) = \beta I(t)/N$
- Prevalence $\pi = I(t)/N$
- Reproduction numbers $R_0 = f(\beta, \gamma)$ or $R_0 = f(r, G(\cdot))$ where $G(\cdot)$ is the generation time distribution; and $R_t = f(I(t), G(\cdot))$

Key quantities to monitor epidemics



- Incidence or force of infection $\lambda(t) = \beta I(t)$ or $\lambda(t) = \beta I(t)/N$
- Prevalence $\pi = I(t)/N$
- Reproduction numbers $R_0 = f(\beta, \gamma)$ or $R_0 = f(r, G(\cdot))$ where $G(\cdot)$ is the generation time distribution; and $R_t = f(I(t), G(\cdot))$
- Severity (e.g. infection-fatality risk)

Linking Models To Data

Epidemic data

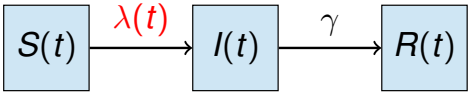
- Direct information on *incidence* of infection hardly ever available (perhaps in small outbreaks)

More typically available

- Final size data
- Temporal data
  - *Prevalence* data
  - *Incidence* of *sequelae* of infection

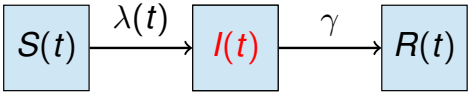
Our goal is prospective (*real-time*) or retrospective estimation

Linking models to data



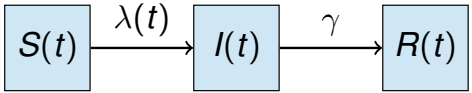
- An observation model required to link data to the SIR system
- An observation model may have inputs:
  - Incidenceof new infections $\lambda(t)S(t)\delta t = \beta I(t)S(t)\delta t$

Linking models to data



- An observation model required to link data to the SIR system
- An observation model may have inputs:
  - Incidenceof new infections $\lambda(t)S(t)\delta t = \beta I(t)S(t)\delta t$
  - Prevalence (e.g. $I(t)/N$)

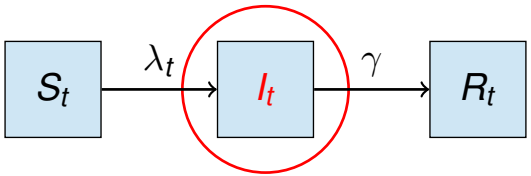
Linking models to data

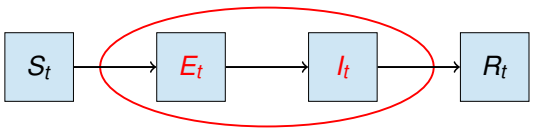


Prev Sym Hos ICU Dea

- An observation model required to link data to the SIR system
- An observation model may have inputs:
  - Incidenceof new infections $\lambda(t)S(t)\delta t = \beta I(t)S(t)\delta t$
  - Prevalence (e.g. $I(t)/N$)
  - Severity (e.g. infection-fatality risk): fraction of incidence that experiences a severe event

Prevalence of *current* infection





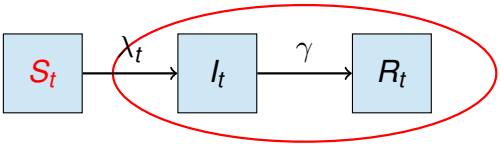
Here, Y_t could represent the number of infections detected out of n_t individuals sampled at random. Data are related to the *prevalence*, π_t, by being considered a realisation of, typically Binomial, distribution:

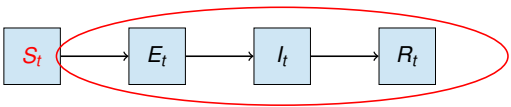
$$Y_t \sim \text{Bin}(n_t, \pi_t) \quad \text{where} \quad \pi_t = I_t/N$$

Alternatively, for *over-dispersed* data, use the Beta-Binomial distribution, with additional dispersion parameter, τ:

$$Y_t \sim \text{BetaBin}\left(n_t, \pi_t/\tau, (1-\pi_t)/\tau\right)$$

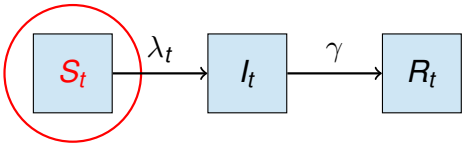
Prevalence of having *ever been infected*

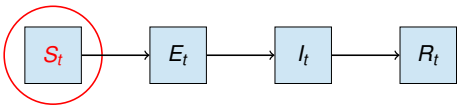




Serological tests/assays measure *antibodies* indicating any previous infection, not necessarily current, informing cumulative incidence

Prevalence of having *ever been infected*





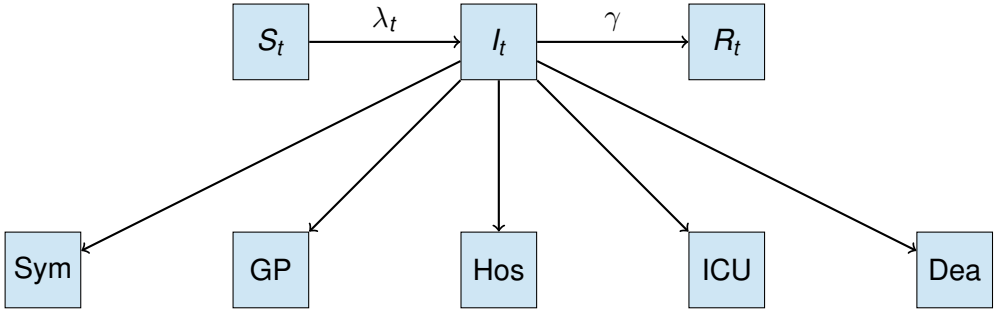
Serological tests/assays measure *antibodies* indicating any previous infection, not necessarily current, informing cumulative incidence or, equivalently, population susceptibility:

$$\pi_t = 1 - S_t/N$$

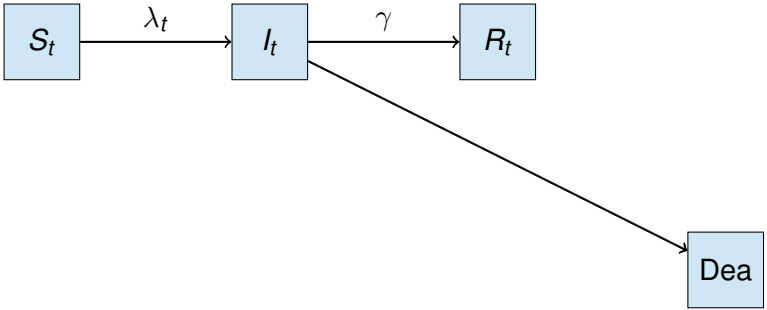
All prevalence test data can include false positives and false negatives, so consider test sensitivity, k_{sens}, and test specificity, k_{spec}:

$$\pi_t = k_{\text{sens}} \left(1 - \frac{S_t}{N}\right) + (1 - k_{\text{spec}}) \frac{S_t}{N}$$

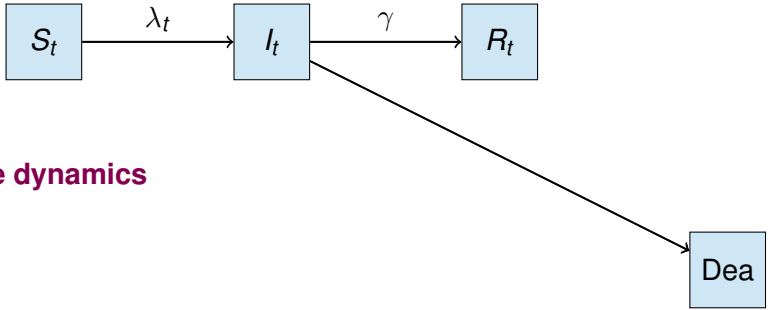
Temporal data on sequelae of infection



Temporal data on sequelae of infection

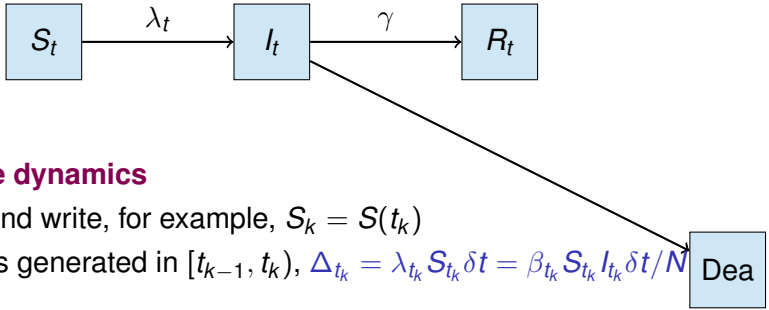


Temporal data on sequelae of infection



Discrete-time dynamics

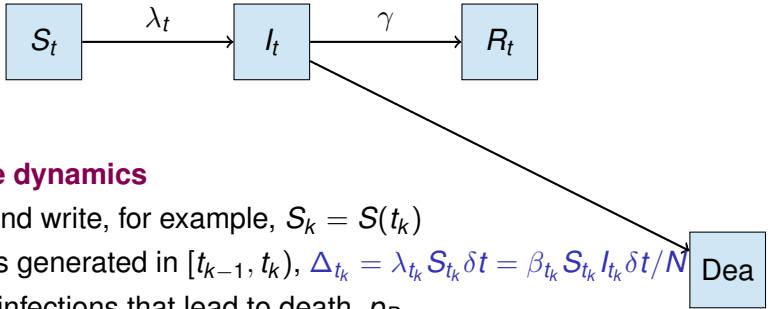
Temporal data on sequelae of infection



Discrete-time dynamics

- Set $t_k = k\delta t$ and write, for example, $S_k = S(t_k)$
- New infections generated in $[t_{k-1}, t_k)$, $\Delta_{t_k} = \lambda_{t_k} S_{t_k} \delta t = \beta_{t_k} S_{t_k} I_{t_k} \delta t / N$

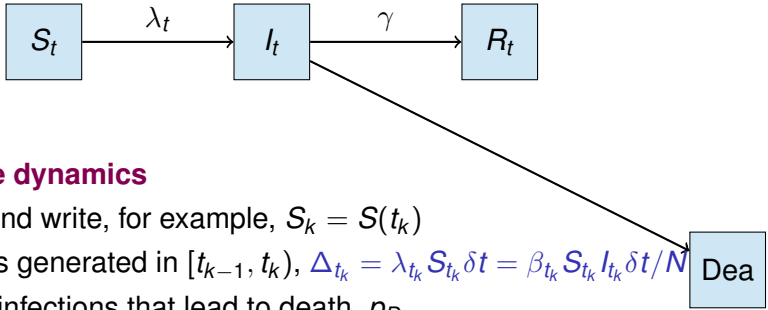
Temporal data on sequelae of infection



Discrete-time dynamics

- Set $t_k = k\delta t$ and write, for example, $S_k = S(t_k)$
- New infections generated in $[t_{k-1}, t_k)$, $\Delta_{t_k} = \lambda_{t_k} S_{t_k} \delta t = \beta_{t_k} S_{t_k} I_{t_k} \delta t / N$
- Proportion of infections that lead to death, p_D.

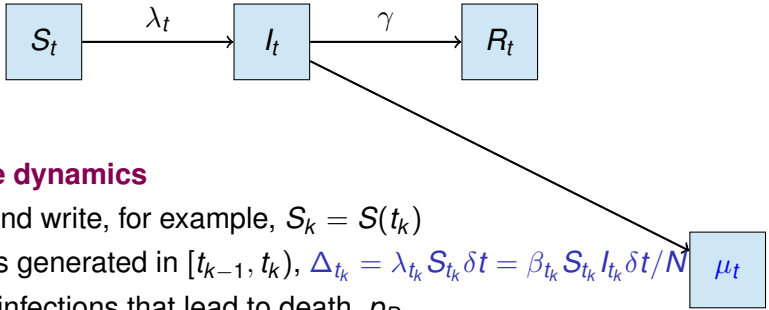
Temporal data on sequelae of infection



Discrete-time dynamics

- Set $t_k = k\delta t$ and write, for example, $S_k = S(t_k)$
- New infections generated in $[t_{k-1}, t_k)$, $\Delta_{t_k} = \lambda_{t_k} S_{t_k} \delta t = \beta_{t_k} S_{t_k} I_{t_k} \delta t / N$
- Proportion of infections that lead to death, p_D.
- Time-to-death governed by discrete distribution $\boldsymbol{f} = (f_0, f_1, \ldots)$.

Temporal data on sequelae of infection

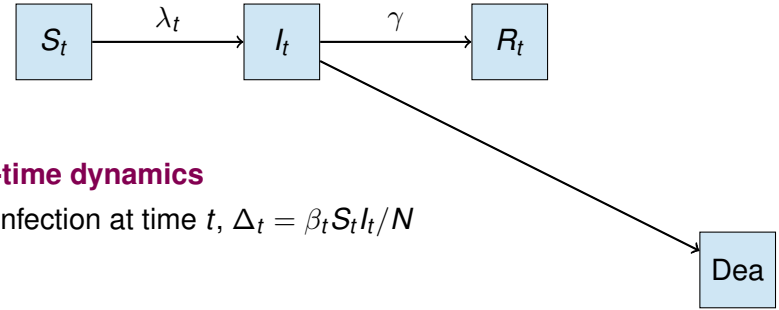


Discrete-time dynamics

- Set $t_k = k\delta t$ and write, for example, $S_k = S(t_k)$
- New infections generated in $[t_{k-1}, t_k)$, $\Delta_{t_k} = \lambda_{t_k} S_{t_k} \delta t = \beta_{t_k} S_{t_k} I_{t_k} \delta t / N$
- Proportion of infections that lead to death, p_D.
- Time-to-death governed by discrete distribution $\boldsymbol{f} = (f_0, f_1, \ldots)$.
- Expected number of deaths per day:

$$\mu_{t_k} = p_D \sum_{l=1}^{k} \Delta_l f_{k-l}$$

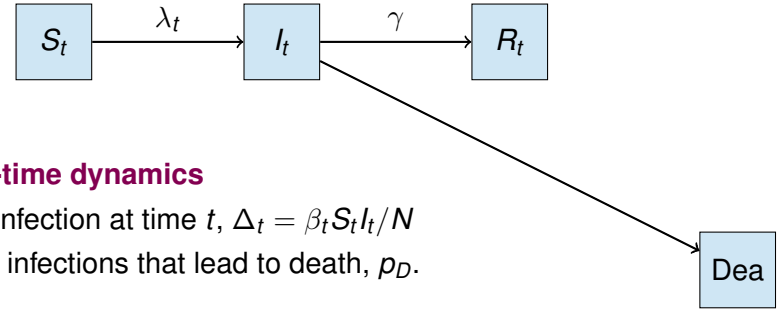
Temporal data on sequelae of infection



Continuous-time dynamics

- Rate of new infection at time t, $\Delta_t = \beta_t S_t I_t / N$

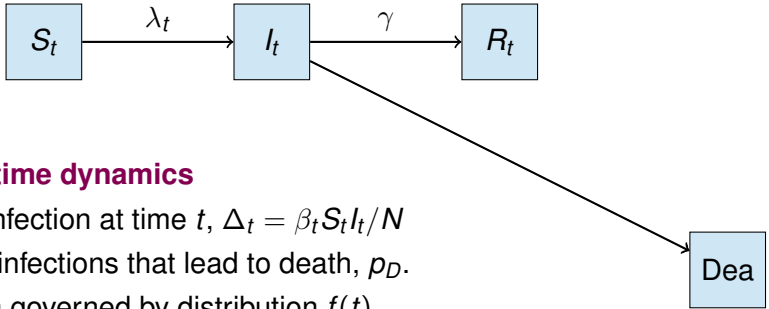
Temporal data on sequelae of infection



Continuous-time dynamics

- Rate of new infection at time t, $\Delta_t = \beta_t S_t I_t / N$
- Proportion of infections that lead to death, p_D.

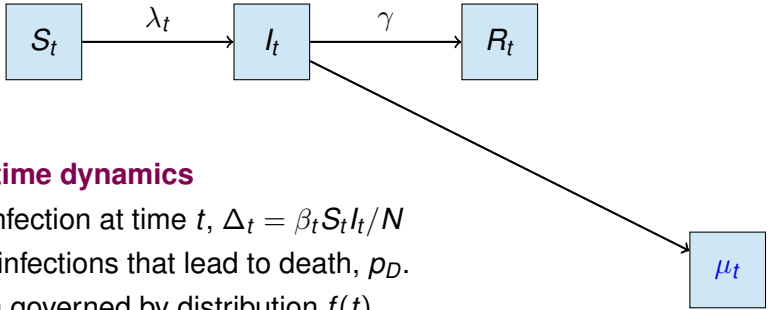
Temporal data on sequelae of infection



Continuous-time dynamics

- Rate of new infection at time t, $\Delta_t = \beta_t S_t I_t / N$
- Proportion of infections that lead to death, p_D.
- Time-to-death governed by distribution $f(t)$.

Temporal data on sequelae of infection

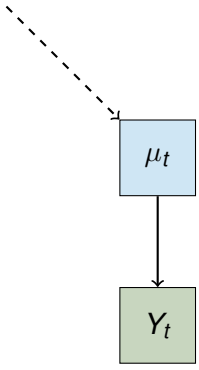


Continuous-time dynamics

- Rate of new infection at time t, $\Delta_t = \beta_t S_t I_t / N$
- Proportion of infections that lead to death, p_D.
- Time-to-death governed by distribution $f(t)$.
- Expected number of deaths per day:

$$\mu_t = p_D \int_0^u \Delta_u f(t-u) du$$

Linking to Data



- Recall the requirement for dispersion parameter η.
- Y_{t_k} the number of cases in $[t_{k-1}, t_k)$
- **Discrete-time**

$$Y_{t_k} \sim \text{NegBin}\left(\mu_{t_k}, \eta\right)$$

- **Continuous-time**

$$Y_t \sim \text{NegBin}\left(\int_t^{t+\delta t} \mu_u du, \eta\right)$$

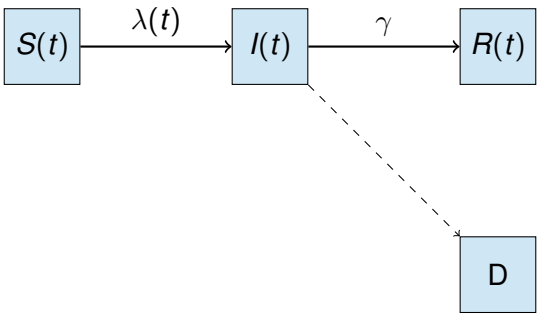
Do we need the convolution?

- Convolutions are relatively expensive to calculate. Can we avoid them?
- Could treat data as removals - arrivals into an absorbing model compartment.

Do we need the convolution?

- Convolutions are relatively expensive to calculate. Can we avoid them?
- Could treat data as removals - arrivals into an absorbing model compartment.

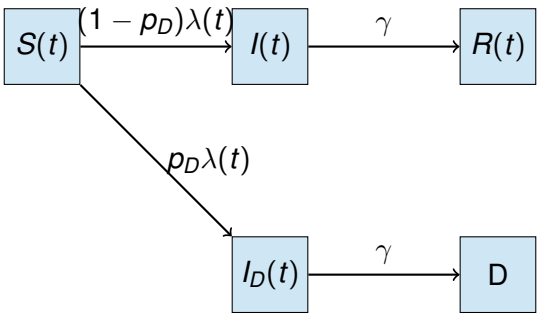
$$\frac{dS}{dt}(t) = -\lambda(t)S(t)$$
$$\frac{dI}{dt}(t) = \lambda(t)S(t) - \gamma I(t)$$
$$\frac{dR}{dt}(t) = \gamma I(t)$$



Do we need the convolution?

- Convolutions are relatively expensive to calculate. Can we avoid them?
- Could treat data as removals - arrivals into an absorbing model compartment.

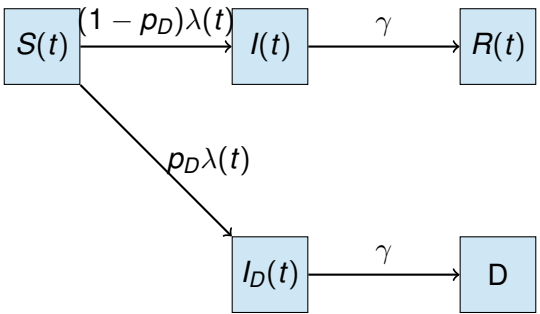
$$\frac{dS}{dt}(t) = -\lambda(t)S(t)$$
$$\frac{dI}{dt}(t) = (1-p_D)\,\lambda(t)S(t) - \gamma I(t)$$
$$\frac{dI_D}{dt}(t) = p_D\lambda(t)S(t) - \gamma I(t)$$
$$\frac{dR}{dt}(t) = \gamma I(t)$$
$$\frac{dD}{dt}(t) = \gamma I_D(t)$$



Do we need the convolution?

- Convolutions are relatively expensive to calculate. Can we avoid them?
- Could treat data as removals - arrivals into an absorbing model compartment.

$$\frac{dS}{dt}(t) = -\lambda(t)S(t)$$
$$\frac{dI}{dt}(t) = (1-p_D)\,\lambda(t)S(t) - \gamma I(t)$$
$$\frac{dI_D}{dt}(t) = p_D\lambda(t)S(t) - \gamma I(t)$$
$$\frac{dR}{dt}(t) = \gamma I(t)$$
$$\frac{dD}{dt}(t) = \gamma I_D(t)$$

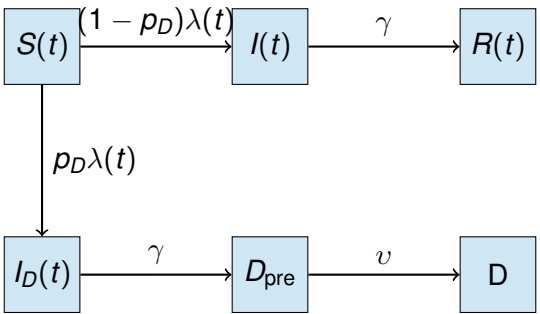


- BUT time to death is the same as the infectious period!

Do we need the convolution?

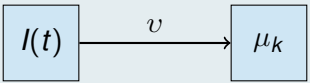
- Convolutions are relatively expensive to calculate. Can we avoid them?
- Could treat data as removals - arrivals into an absorbing model compartment.

$$\frac{dS}{dt}(t) = -\lambda(t)S(t)$$
$$\frac{dI}{dt}(t) = (1-p_D)\,\lambda(t)S(t) - \gamma I(t)$$
$$\frac{dI_D}{dt}(t) = p_D\lambda(t)S(t) - \gamma I(t)$$
$$\frac{dD_{\text{pre}}}{dt} = \gamma I_D(t) - \upsilon D(t)$$
$$\frac{dR}{dt}(t) = \gamma I(t)$$
$$\frac{dD}{dt}(t) = \upsilon I_D(t)$$



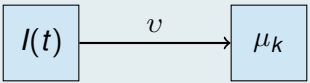
Non-Exponential Delay Times

Exponential



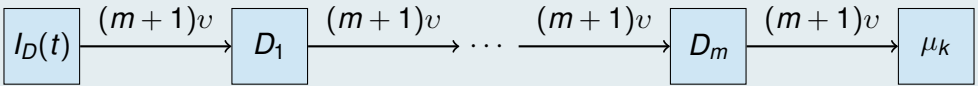
Non-Exponential Delay Times

Exponential



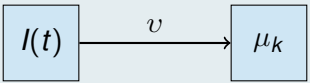
Gamma

- Get more flexible delay times through using composite states.
- Delay time is based on a $\Gamma(m+1, \upsilon/(m+1))$.



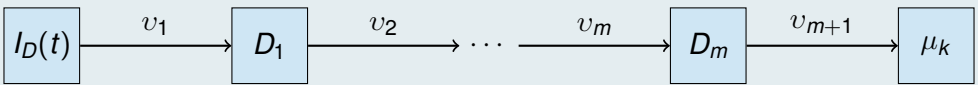
Non-Exponential Delay Times

Exponential



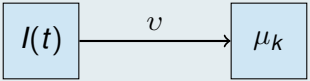
More generally...

- $\frac{1}{\upsilon} = \sum_{i=1}^{m+1} \frac{1}{\upsilon_i}$.
- Non-standard distribution overall.



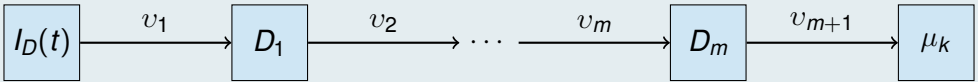
Non-Exponential Delay Times

Exponential



More generally...

- $\frac{1}{\upsilon} = \sum_{i=1}^{m+1} \frac{1}{\upsilon_i}$.
- Non-standard distribution overall.

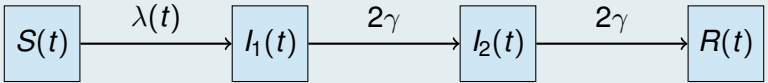


- Intermediate states could represent physically meaningful quantities, e.g. layers of severity, about which we may have some useful information.

Composition transition model states

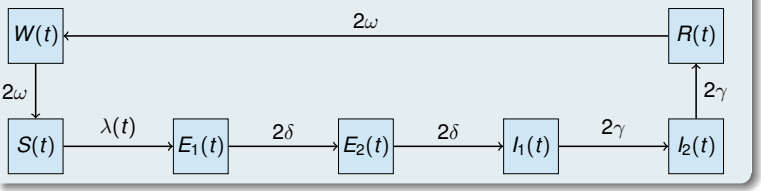
- Similarly, composite transmission states can add greater flexibility.

SIIR



- Useful for longer durations, e.g. the waning of immunity, where it is more likely that those who entered a state longer ago are more likely to move.

SEEIIRW



Multiple Source of Data

- Observations at time t can include a collection of k data sources $\boldsymbol{y}_t = \{\boldsymbol{y}_{1,t}, \ldots, \boldsymbol{y}_{k,t}\}$ *e.g.* both prevalence and incidence-type data
- Introduces multiple, simultaneous links between model and data
- Each link has its own observational model *e.g.*

$$\mathrm{Y}_t^{\mathrm{prev}} \sim \mathrm{Bin}\left(n_t, \pi_t\right), \ldots, \mathrm{Y}_t^{\mathrm{hosp}} \sim \mathrm{NegBin}\left(\mu_t, \eta\right)$$

  where π_t and μ_t may depend on model states $\boldsymbol{x}_t$ and/or system parameters β, γ, and η is component of the observational model.
- Conditionally on the model and parameters, these data are typically treated as independent
- More realistically, dependencies in the data [Corbella *et al* 2022] should be accounted for.

In generality: (Markov) State Space Model formulation

Denote $\boldsymbol{X}_t = \{S_t, I_t, R_t\}$, the system evolution described by the initial density $f_{\boldsymbol{\theta}}(\boldsymbol{x}_0; \boldsymbol{\theta})$ and the one-step transitions

$$\boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{\theta} \sim f_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\theta})$$
$$\boldsymbol{Y_t}|\boldsymbol{X}_t, \psi \sim f_{\psi}(\boldsymbol{y}_t|\boldsymbol{x}_t; \psi)$$

- where $\boldsymbol{\theta}$ and ψ are system and observational process parameters respectively
- $\boldsymbol{y}_t$ data observed at time t
- different degree of noise (system/observational) in both of the above results in models with different level of complexity

In generality: (Markov) State Space Model formulation

- Only observational noise

$$\begin{aligned} \boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{\theta} &= g_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}) \\ \boldsymbol{Y_t}|\boldsymbol{X}_t, \boldsymbol{\psi} &\sim f_{\boldsymbol{\psi}}(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\psi}) \end{aligned}$$

  where $g_{\boldsymbol{\theta}}(\cdot)$ is a deterministic function of $\boldsymbol{\theta}$

- Both system and observational noise

$$\begin{aligned} \boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{\theta} &\sim f_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\theta}) \\ \boldsymbol{Y_t}|\boldsymbol{X}_t, \boldsymbol{\psi} &\sim f_{\boldsymbol{\psi}}(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\psi}) \end{aligned}$$

Inference

Parameter estimation - only observational noise

$$\boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{\theta} = g_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1})$$
$$\boldsymbol{Y_t}|\boldsymbol{X}_t, \boldsymbol{\psi} \sim f_{\boldsymbol{\psi}}(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\psi})$$

- *Relatively straightforward* to estimate parameters $\boldsymbol{\theta}, \boldsymbol{\psi}$ from observed data $y_{1:T}$
- Simplest: minimize the sum of squares of differences between observed data and model prediction

Parameter estimation - only observational noise

Likelihood-based inference

- Maximum likelihood estimation

$$\mathcal{L}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta}, \psi) = f_{\boldsymbol{\theta},\psi}(\boldsymbol{y}_{1:T} \mid \boldsymbol{x}_{1:T}; \boldsymbol{\theta}, \psi)$$

- Bayesian inference

$$p(\boldsymbol{\theta}, \psi \mid \boldsymbol{y}_{1:T}) = \frac{\mathcal{L}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta}, \psi) p(\boldsymbol{\theta}, \psi)}{p(\boldsymbol{y}_{1:T})} \propto \mathcal{L}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta}, \psi) p(\boldsymbol{\theta}, \psi)$$

with $p(\boldsymbol{\theta}, \psi)$ the *prior* distribution for $(\boldsymbol{\theta}, \psi)$

Bayesian Inference

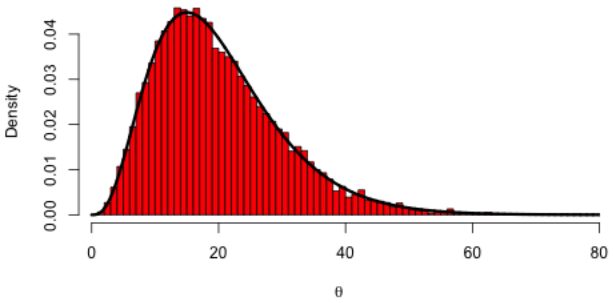
- Apart from trivial cases, the posterior distribution is not a standard distribution

Bayesian Inference

- Apart from trivial cases, the posterior distribution is not a standard distribution

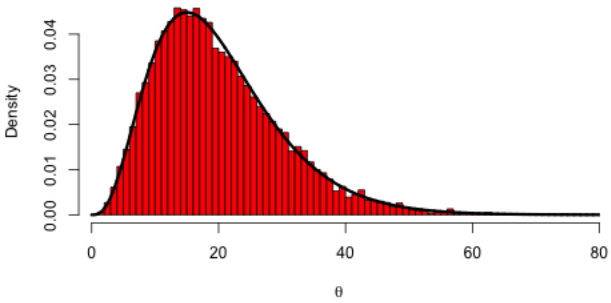
Bayesian Inference

- Apart from trivial cases, the posterior distribution is not a standard distribution
- If we can obtain sample from it, we can generate a large sample $(\theta_1, \psi_1), \ldots, (\theta_N, \psi_N)$.



Bayesian Inference

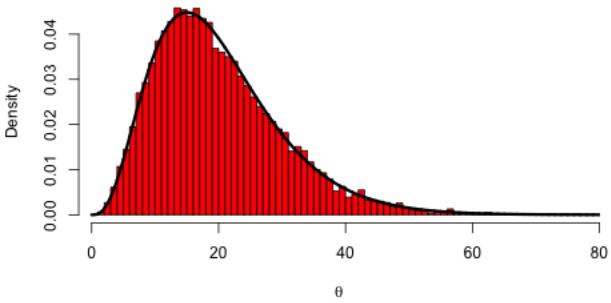
- Apart from trivial cases, the posterior distribution is not a standard distribution
- If we can obtain sample from it, we can generate a large sample $(\boldsymbol{\theta_1}, \boldsymbol{\psi_1}), \ldots, (\boldsymbol{\theta_N}, \boldsymbol{\psi_N})$.
- Can then use features of the sample to estimate posterior features of interest *e.g.*:



Bayesian Inference

- Apart from trivial cases, the posterior distribution is not a standard distribution
- If we can obtain sample from it, we can generate a large sample $(\boldsymbol{\theta_1}, \boldsymbol{\psi_1}), \ldots, (\boldsymbol{\theta_N}, \boldsymbol{\psi_N})$.
- Can then use features of the sample to estimate posterior features of interest *e.g.*:

$$\mathbb{E}[\theta|y] \approx \frac{1}{N}\sum_{i=1}^{N}\theta_i$$

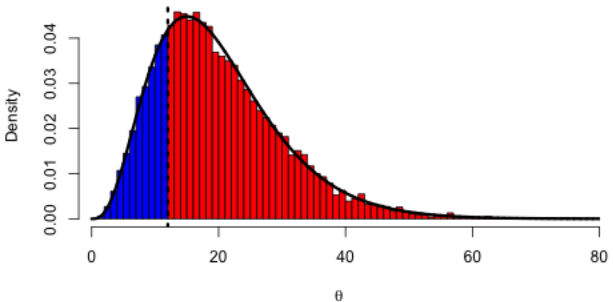


Bayesian Inference

- Apart from trivial cases, the posterior distribution is not a standard distribution
- If we can obtain sample from it, we can generate a large sample $(\boldsymbol{\theta_1}, \boldsymbol{\psi_1}), \ldots, (\boldsymbol{\theta_N}, \boldsymbol{\psi_N})$.
- Can then use features of the sample to estimate posterior features of interest *e.g.*:

$$\mathbb{E}[\theta|y] \approx \frac{1}{N}\sum_{i=1}^{N}\theta_i$$

$$\mathbb{P}(\theta < c) \approx \frac{1}{N}\sum_{i=1}^{N}1(\theta_i < c)$$



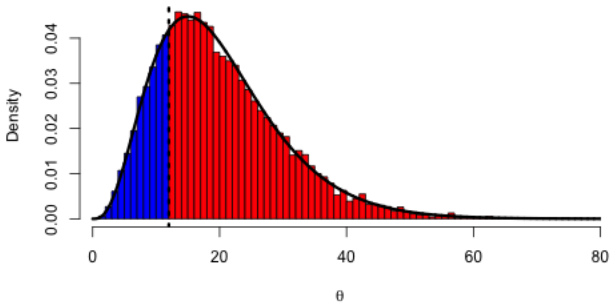
Bayesian Inference

- Apart from trivial cases, the posterior distribution is not a standard distribution
- If we can obtain sample from it, we can generate a large sample $(\boldsymbol{\theta_1}, \boldsymbol{\psi_1}), \ldots, (\boldsymbol{\theta_N}, \boldsymbol{\psi_N})$.
- Can then use features of the sample to estimate posterior features of interest *e.g.*:

$$\mathbb{E}[\theta|y] \approx \frac{1}{N}\sum_{i=1}^{N}\theta_i$$

$$\mathbb{P}(\theta < c) \approx \frac{1}{N}\sum_{i=1}^{N}1(\theta_i < c)$$

This is known as Monte Carlo estimation.



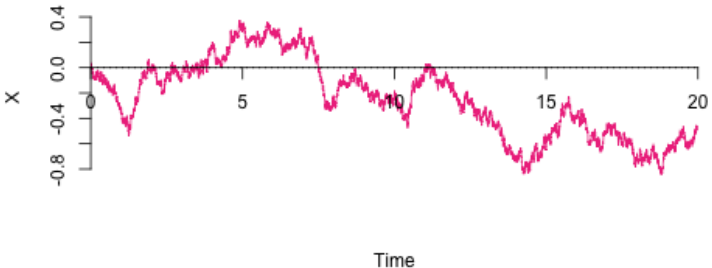
Monte Carlo Sampling from Posteriors

- We cannot obtain *independent* samples directly from the posterior
- Can instead generate a sequence of *dependent* random variables whose distribution converges to the posterior distribution of interest
- Idea underlying Markov chain Monte Carlo (*MCMC*) (e.g. Gelman et al.2004; Gamerman and Lopes 2006)

Markov chains

- A Markov chain is any random sequence of numbers where the future of the sequence depends only on the current state and not on its history.
- Most simple example is the random walk

$$\theta_{t+1} \sim \mathrm{N}\left(\theta_t, \sigma^2\right)$$



Metropolis-Hastings algorithm Metropolis et al.1953; Hastings 1970

1. **Initialise:** Set $i = 0$, pick starting state θ_0, ψ_0
2. **Set:** $i = i + 1$.
3. **Sample:** $(\tilde{\theta}, \tilde{\psi}) \sim q\left(\cdot|\theta_{i-1}, \psi_{i-1}\right)$.
4. **Calculate:** acceptance probability $\alpha = \min\left\{1, \frac{p\left(\tilde{\theta},\tilde{\psi}\right) q\left(\theta_{i-1},\psi_{i-1}|\tilde{\theta},\tilde{\psi}\right)}{p\left(\theta_{i-1},\psi_{i-1}\right) q\left(\tilde{\theta},\tilde{\psi}|\theta_{i-1},\psi_{i-1}\right)}\right\}$
5. **Sample:** $u \sim U[0, 1]$
6. **Accept/Reject:** Set $\theta_i = \tilde{\theta}, \psi_i = \tilde{\psi}$ if $u < \alpha$, else set $\theta_i = \theta_{i-1}$ and $\psi_i = \psi_{i-1}$. Return to **2**

Choice of $q\left(\cdot|\theta_{i-1}, \psi_{i-1}\right)$ crucial for convergence - a vast literature exists on this choice

Parameter estimation - both system and observational noise

$$\boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{\theta} \sim f_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\theta})$$
$$\boldsymbol{Y_t}|\boldsymbol{X}_t, \boldsymbol{\psi} \sim f_{\boldsymbol{\psi}}(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\psi})$$

- Particularly challenging for both Bayesian and non-Bayesian frameworks as the likelihood $\mathcal{L}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta}, \boldsymbol{\psi})$

$$\mathcal{L}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta}, \boldsymbol{\psi}) = f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta}) = \int f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{x}_{0:T}, \boldsymbol{y}_{1:T}; \boldsymbol{\theta}) d\boldsymbol{x}_{0:T}$$

- cannot be typically evaluated (apart from simpler models/data structures) - requiring high dimensional integration over the unknown model states

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:
  - Approximate Bayesian Computation (*e.g.* Kypraios et al, 2017).

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:
  - Approximate Bayesian Computation (*e.g.* Kypraios et al, 2017).
  - Synthetic likelihood (Wood, 2010) to provide proxy measures of likelihood based on summary statistics -*Heavily dependent on the choice of appropriate summary statistics*

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:
  - Approximate Bayesian Computation (*e.g.* Kypraios et al, 2017).
  - Synthetic likelihood (Wood, 2010) to provide proxy measures of likelihood based on summary statistics -*Heavily dependent on the choice of appropriate summary statistics*
- **Particle filters** to

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:
  - Approximate Bayesian Computation (*e.g.* Kypraios et al, 2017).
  - Synthetic likelihood (Wood, 2010) to provide proxy measures of likelihood based on summary statistics -*Heavily dependent on the choice of appropriate summary statistics*
- **Particle filters** to
  - *estimate* the likelihood (Arulampalam et al, 2002; Andrieu et al, 2010) - used in particle MCMC (PMCMC) and SMC^2.

Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:
  - Approximate Bayesian Computation (*e.g.* Kypraios et al, 2017).
  - Synthetic likelihood (Wood, 2010) to provide proxy measures of likelihood based on summary statistics -*Heavily dependent on the choice of appropriate summary statistics*
- **Particle filters** to
  - *estimate* the likelihood (Arulampalam et al, 2002; Andrieu et al, 2010) - used in particle MCMC (PMCMC) and SMC^2.
  - maximise the likelihood via *iterated filtering* (Ionides et al, 2006, 2015, 2017).

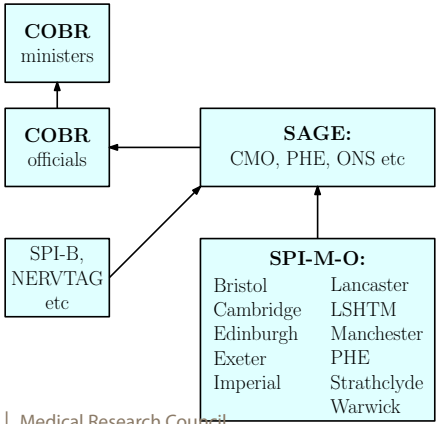
Intractable likelihoods

Active area of research in *statistics* - approaches include the use of:

- **Data augmentation** to obtain a tractable likelihood (Gibson & Renshaw, 1998; O'Neill & Roberts, 1999) - *Potentially computationally expensive*
- **Simulation methods** *e.g*:
  - Approximate Bayesian Computation (*e.g.* Kypraios et al, 2017).
  - Synthetic likelihood (Wood, 2010) to provide proxy measures of likelihood based on summary statistics -*Heavily dependent on the choice of appropriate summary statistics*
- **Particle filters** to
  - *estimate* the likelihood (Arulampalam et al, 2002; Andrieu et al, 2010) - used in particle MCMC (PMCMC) and SMC^2.
  - maximise the likelihood via *iterated filtering* (Ionides et al, 2006, 2015, 2017).
  - *use all data and retain model structure*

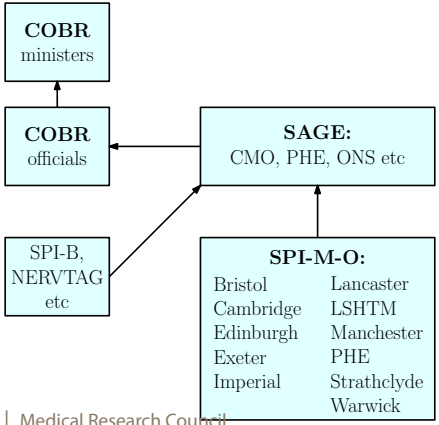
COVID19 work - involvement in governmental advisory groups

Scientific Pandemic Influenza Advisory Committee on Modelling (SPI-M-O)



COVID19 work - involvement in governmental advisory groups

Scientific Pandemic Influenza Advisory Committee on Modelling (SPI-M-O)



- Commissions received from Cabinet Office
- Expressed as a question that can be addressed through 'modelling'
- Swift answers (24/48 hours!) from a number of groups/models
- SPI-M-O discusses results - **consensus** achieved
- **Consensus** communicated to SAGE
- SAGE discussion - translation into advice

Our contribution - Nowcasting & Forecasting

- **Now-casting**: estimate of the **current** state of the epidemic
  - level of disease transmission (R numbers)
  - number of new daily infections
  - prevalence of infection
  - proportion of the population ever infected (attack rate)

- **Forecasting**: prediction of relevant quantities
  - demand on the health system (*e.g.* hospitalisations), deaths

- **Real time**: as data become available sequentially

Real-time epidemic monitoring: why it is difficult?

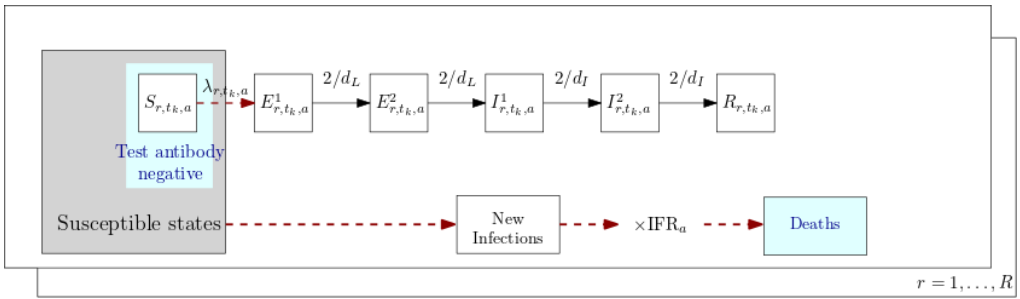
- The transmission process is latent - infections are **not observable**
- Direct data not available
- Plenty of indirect data
  - Noisy, incomplete, often biased data streams on related outcomes (e.g. time series of deaths - hospitalisation - prevalence etc.)
  - Meaningfully **integrated** in a model of disease transmission
- Analyses to be carried out in a **timely fashion** (within hours) to meet the SPI-M-O deadlines
- Model continuously adapted to tackle emergent challenges

Challenges

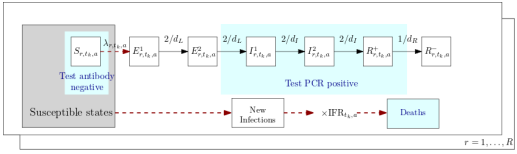
- **data integration** (Bayesian approach) (De Angelis, *et al*, 2015; De Angelis, Presanis, 2019)
- **efficient algorithms** as model becomes more complex and data accumulate

Initial model: March 16th - early July 2020 Birrell *et al.*, 2021

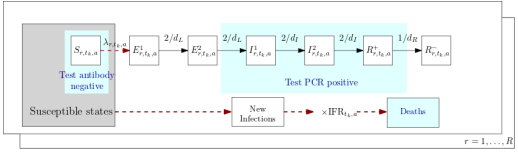
Deterministic Susceptible(S)-Exposed(E)-Infected(I)-Removed(R) model

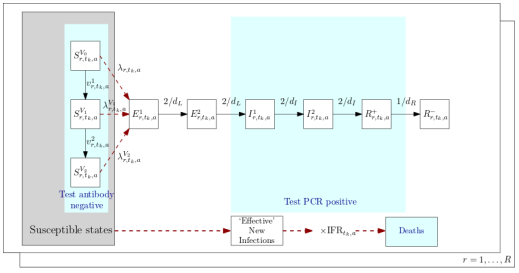


Pandemic Model Development

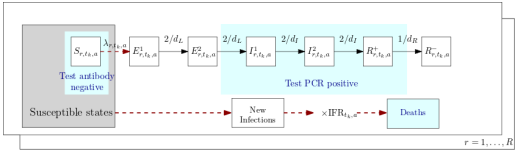


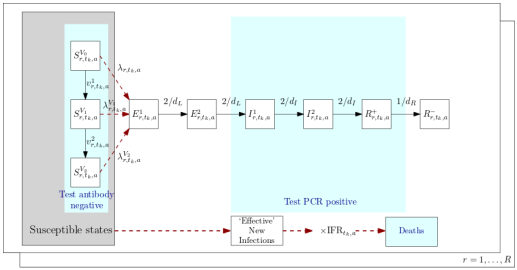
Pandemic Model Development

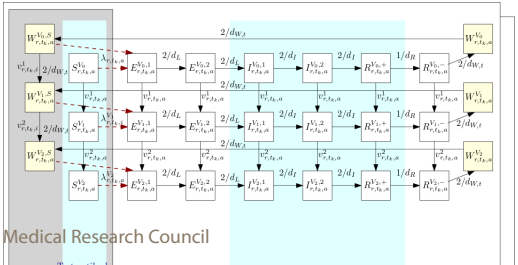




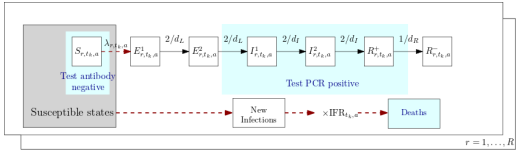
Pandemic Model Development

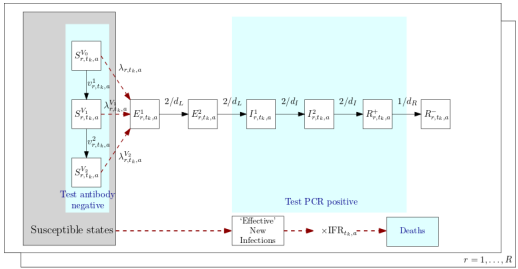


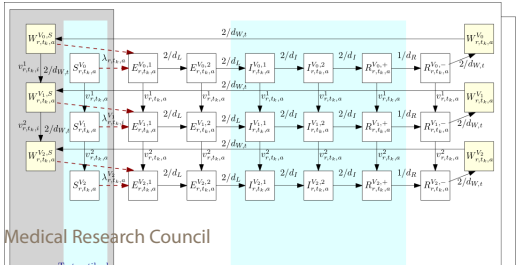


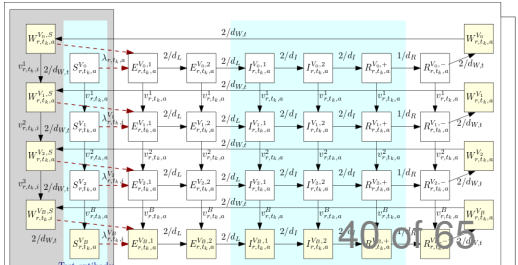


Pandemic Model Development

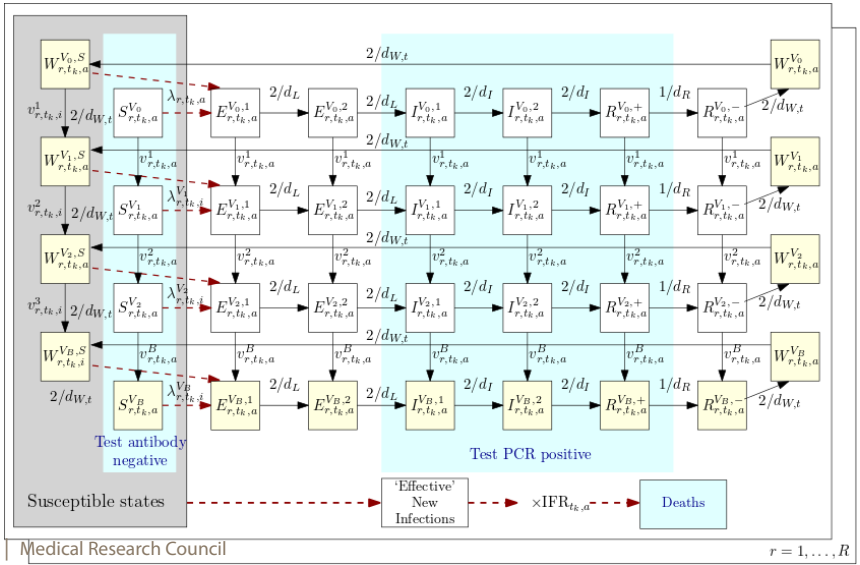








Current model: additional data - vaccination - waning



Model details - transmission dynamics

Discrete-time deterministic model for transmission governed by the system of equations

$$
\begin{aligned}
S_{r,t_k,i} &= S_{r,t_{k-1},i}\left(1-\lambda_{r,t_{k-1},i}\delta\right)\\
E^1_{r,t_k,i} &= E^1_{r,t_{k-1},i}\left(1-\frac{2\delta}{d_L}\right)+S_{r,t_{k-1},i}\lambda_{r,t_{k-1},i}\delta\\
E^2_{r,t_k,i} &= E^2_{r,t_{k-1},i}\left(1-\frac{2\delta}{d_L}\right)+E^1_{r,t_{k-1},i}\frac{2\delta}{d_L}\\
I^1_{r,t_k,i} &= I^1_{r,t_{k-1},i}\left(1-\frac{2\delta}{d_I}\right)+E^2_{r,t_{k-1},i}\frac{2\delta}{d_L}\\
I^2_{r,t_k,i} &= I^2_{r,t_{k-1},i}\left(1-\frac{2\delta}{d_I}\right)+I^1_{r,t_{k-1},i}\frac{2\delta}{d_I},\\
R_{r,t_k,i} &= R_{r,t_{k-1},i}+I^2_{r,t_{k-1},i}\frac{2\delta}{d_I}
\end{aligned}
$$

$r = 1,\ldots,R$, $k = 1,\ldots,K$ and $i = 1,\ldots,A$.

- R **regions** (7 NHS regions or 9 ONS regions), K **time points**, A **age groups**.
- d_I: mean **infectious period**, d_L: mean latent period (**known**).
- $\lambda_{r,t_k,i}$: rate for **S** $\rightarrow$ **I**; $\delta = t_k - t_{k-1} = 0.5$.

Model details - transmission kernel

New infections are generated as

$$\begin{aligned}\Delta^{\text{infec}}_{r,t_k,i} &= S_{r,t_k,i} p_{r,t_k,i} \\ &= S_{r,t_k,i} \left(1 - \prod_{j=1}^{A} \left[\left(1 - \mathbf{b}^{t_k}_{r,ij} \right)^{I^1_{r,t_k,j} + I^2_{r,t_k,j}} \right] \right) \delta\end{aligned}$$

$$\begin{aligned}\mathbf{b}^{t_k}_{r,ij} &= \mathbb{P}(\text{Suscept. aged } i \text{ infected by infectious individual aged } j \text{ at time } t_k \text{ in region } r) \\ &= \frac{\beta_{t_k,r} R_{0,r}}{R^*_{0,r}} \tilde{C}^{t_k}_{r,ij},\end{aligned}$$

depends on $\tilde{C}^{t_k}_{r,ij} = C^{t_k}_{ij} \odot M^{t_k}_{r,ij}$ and $\beta_{t_k,r}$

- $C^{t_k}_{ij}$ - time-varying matrix of contacts between individuals in groups i and j at time t_k.

Model details - transmission kernel

New infections are generated as

$$\begin{aligned}\Delta^{\text{infec}}_{r,t_k,i} &= S_{r,t_k,i} p_{r,t_k,i} \\ &= S_{r,t_k,i} \left(1 - \prod_{j=1}^{A} \left[\left(1 - \mathbf{b}^{t_k}_{r,ij} \right)^{I^1_{r,t_k,j} + I^2_{r,t_k,j}} \right] \right) \delta\end{aligned}$$

$$\begin{aligned}\mathbf{b}^{t_k}_{r,ij} &= \mathbb{P}\big(\text{Suscept. aged } i \text{ infected by infectious individual aged } j \text{ at time } t_k \text{ in region } r\big) \\ &= \frac{\beta_{t_k,r} R_{0,r}}{R^*_{0,r}} \tilde{C}^{t_k}_{r,ij},\end{aligned}$$

depends on $\tilde{C}^{t_k}_{r,ij} = C^{t_k}_{ij} \odot M^{t_k}_{r,ij}$ and $\beta_{t_k,r}$

- $C^{t_k}_{ij}$ - time-varying matrix of contacts between individuals in groups i and j at time t_k.
- $M^{t_k}_{r,ij}$ - region-specific matrix of relative susceptibility of individual in age-group i to an infection from an infectious individual in group j given contact

Model details - time-varying stochastic transmissibility $\beta_{t_k,r}$

$\beta_{t_k,r}$ - region-time varying - encapsulate unobserved temporal fluctuations in transmission *e.g.* behavioural aspects

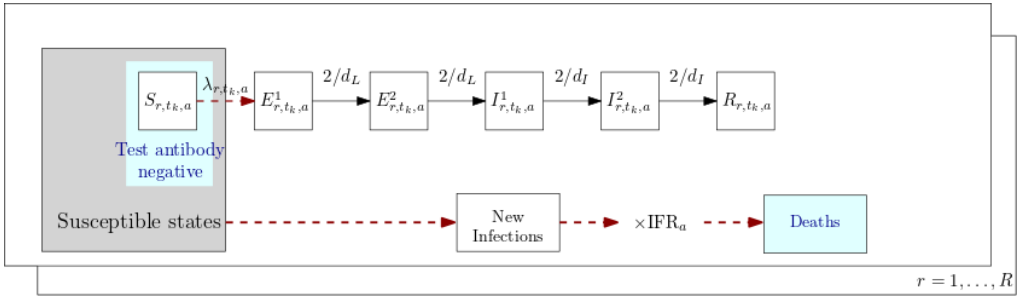
- Let $\tilde{\beta}_{t_k,r} = \log(\beta_{t_k,r})$ and assume

$$\tilde{\beta}_{t_k,r} \sim \mathrm{N}\big(\tilde{\beta}_{t_{k-1},r}, \sigma^2_{r,\beta}\big), \quad \tilde{\beta}_{\mathrm{tlock},r} = 0,$$

  $\tilde{\beta}_{\mathrm{tlock},r}$ applies in all weeks up to the first lock-down.

- Currently a **piecewise** constant process, e.g. fortnightly:
  - Let $w_k \equiv w(t_k)$ indicate the week in which time t_k falls.
  - Then $\tilde{\beta}_{w_k,r} \sim \mathrm{N}\big(\tilde{\beta}_{w_{k-1},r}, \sigma^2_{r,\beta}\big)$.

Initial model: March 16th - early July 2020 Birrell *et al.*, 2021



- time series of **COVID19-confirmed deaths**
- **serology data** (NHS Blood&Transplant)
- **contact rates** POLYMOD/Google mobility/ONS UK Time-Use Survey/DfE schools attendance
- prior information (*e.g.* age-specific IFR)

The likelihood(s)

- **Deaths**[1]: $\mathbf{X_{r,t_k,i}}$ number of deaths, p_i age-specific infection-fatality ratio, $F_{k-\ell}$ (given) prob. of ℓ days from infection to death

$$\mathbf{X_{r,t_k,i}} \sim \text{NegBin}\left(p_i \sum_{\ell=0}^{k} F_{k-\ell} \Delta^{\text{infec}}_{r,t_\ell,i}, \eta\right), \quad \eta \text{ to measure over-dispersion}$$

[1] Similarly we incorporate hospitalisation data to make projections of hospital burden.

The likelihood(s)

- **Deaths**[1]: $\mathbf{X_{r,t_k,i}}$ number of deaths, p_i age-specific infection-fatality ratio, $F_{k-\ell}$ (given) prob. of ℓ days from infection to death

$$\mathbf{X_{r,t_k,i}} \sim \text{NegBin}\left(p_i \sum_{\ell=0}^{k} F_{k-\ell} \Delta^{\text{infec}}_{r,t_\ell,i}, \eta\right), \quad \eta \text{ to measure over-dispersion}$$

- **Serological data**: k_{sens} and k_{spec} denote sens. and spec. of blood tests, $n_{r,t_k,i}$ test samples, $\mathbf{Y_{r,t_k,i}}$ serologically positive tests

$$\mathbf{Y_{r,t_k,i}} \sim Bin\left(n_{r,t_k,i}, k_{\text{ksens}}\left(1 - \frac{S_{r,t_k,i}}{N_{r,i}}\right) + (1 - k_{\text{kspec}})\frac{S_{r,t_k,i}}{N_{r,i}}\right).$$

[1] Similarly we incorporate hospitalisation data to make projections of hospital burden.

Parameters

static parameters

- $\theta_r = (I_{0,r}, \psi_r, m_{1,r}, m_{2,r}, m_{3,r})$ are **region-specific** parameters
  - $I_{0,r}$: initialization of the ODE system.
  - ψ_r: epidemic growth rate parameters
  - $m_{1,r}, m_{2,r}, m_{3,r}$: parameters of the contact matrices $\tilde{C}_r^{t_k}$
- $\theta_g = (\eta, \sigma_\beta^2, k_{\text{sens}}, k_{\text{spec}}, d_I, p_{1:n_A})$ are **global** (common across regions) parameters

dynamic parameters - realisation of a stochastic process

- $\tilde{\beta}_r$ (region-specific) parameters; $\tilde{K}$**-dimensional vector** where $\tilde{K}$ is the length (in weeks) of the data time series

Posterior of interest

- We denote by D all the **data**.
- We use **highly informative priors** for the parameters we have information on (details in Birrell et al., (2021))

- **Posterior of interest**:

$$\pi(\theta_g, \theta_{1:R}, \tilde{\beta}_{1:R}|D) \propto p(\theta_g)p(\tilde{\beta}_{1:R}|\theta_{1:R})p(\theta_{1:R})p(D|\theta, \theta_{1:R}, \tilde{\beta}_{1:R})\}$$

Posterior of interest

- We denote by D all the **data**.
- We use **highly informative priors** for the parameters we have information on (details in Birrell et al., (2021))
- **Posterior of interest**:

$$\pi(\theta_g, \theta_{1:R}, \tilde{\beta}_{1:R}|D) \propto p(\theta_g)p(\tilde{\beta}_{1:R}|\theta_{1:R})p(\theta_{1:R})p(D|\theta, \theta_{1:R}, \tilde{\beta}_{1:R})\}$$

- **High-dimensional** problem - dimension increasing over time: **from initial 50 parameters to $>$800** currently in the model

Posterior of interest

- We denote by D all the **data**.
- We use **highly informative priors** for the parameters we have information on (details in Birrell et al., (2021))
- **Posterior of interest**:

$$\pi(\theta_g, \theta_{1:R}, \tilde{\beta}_{1:R}|D) \propto p(\theta_g)p(\tilde{\beta}_{1:R}|\theta_{1:R})p(\theta_{1:R})p(D|\theta, \theta_{1:R}, \tilde{\beta}_{1:R})\}$$

- **High-dimensional** problem - dimension increasing over time: **from initial 50 parameters to $>$800** currently in the model
- **Weakly-identifiable parameters** partially address through prior information

Posterior of interest

- We denote by D all the **data**.
- We use **highly informative priors** for the parameters we have information on (details in Birrell et al., (2021))

- **Posterior of interest**:

$$\pi(\theta_g, \theta_{1:R}, \tilde{\beta}_{1:R}|D) \propto p(\theta_g)p(\tilde{\beta}_{1:R}|\theta_{1:R})p(\theta_{1:R})p(D|\theta, \theta_{1:R}, \tilde{\beta}_{1:R})\}$$

- **High-dimensional** problem - dimension increasing over time: **from initial 50 parameters to $>$800** currently in the model
- **Weakly-identifiable parameters** partially address through prior information
- **Bespoke** sampling algorithms needed to produce results **on time**

Posterior of interest

- We denote by D all the **data**.
- We use **highly informative priors** for the parameters we have information on (details in Birrell et al., (2021))

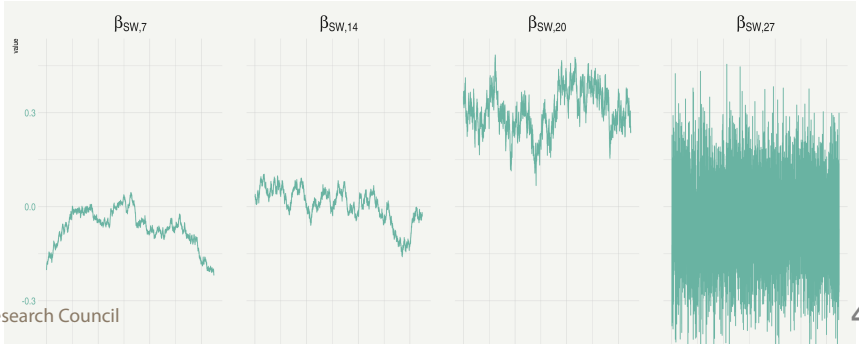
- **Posterior of interest**:

$$\pi(\theta_g, \theta_{1:R}, \tilde{\beta}_{1:R}|D) \propto p(\theta_g)p(\tilde{\beta}_{1:R}|\theta_{1:R})p(\theta_{1:R})p(D|\theta, \theta_{1:R}, \tilde{\beta}_{1:R})\}$$

- **High-dimensional** problem - dimension increasing over time: **from initial 50 parameters to $>$800** currently in the model
- **Weakly-identifiable parameters** partially address through prior information
- **Bespoke** sampling algorithms needed to produce results **on time**
  - quantities improving performance of Bayesian methods are **not always tractable** or **cheap to compute**, e.g. likelihood gradients

MCMC: Random Walk Metropolis Hastings (RW-MH)

- **February 2020 - April 2021**: simple RW-MH with block updates.
  - Consider M blocks of the parameters where τ_m is the m-th d_m-dimensional block.
  - At the i-th MCMC iteration propose $\tau_m^{i+1} = \tau_m^i + \sqrt{\lambda_m} N(0, I_d)$.
  - We learn $\sqrt{\lambda_m}$ during the burn-in period to achieve acceptance rate for each block in a desired level (e.g. 0.234)
- **16/04/2021** Estimating $\sim$ 300 parameters in 25 hours



MCMC: Adaptive Metropolis with Global Scaling (AMGS)

- **March 2021 - January 2022**:
  - We now consider **region specific blocks** updated **in parallel** and a single block for the global parameters [H] MCMC that targets the posterior of interest. [1] Set the number of iterations ν. $i = 1, \ldots, \nu \quad r = 1, \ldots, R$ Draw (θ_r, β_r) from $p(\theta_r, \beta_r | \theta, D_r)$. Draw θ from $p(\theta | \theta_{1:R}, \tilde{\beta}_{1:R}, D)$.
  - We have $M = R + 1$ blocks and we apply the AMGS algorithm (see e.g. Andrieu and Thom, 2008) to update the parameters in τ_m.
  - We have to learn the **covariance matrix** of the proposal distribution.

AMGS: naive implementation

- Let τ the parameters in the *m*-th block at the *i*-th iteration; **Propose** $\tau^{i+1} \sim N(\tau^i, \lambda^i \Sigma_i)$ to accept/reject with MH probability $\alpha(\tau^i, \tau^{i+1})$ and obtain τ^{i+1}.

AMGS: naive implementation

- Let τ the parameters in the *m*-th block at the *i*-th iteration; **Propose** $\tau^{i+1} \sim N(\tau^i, \lambda^i \Sigma_i)$ to accept/reject with MH probability $\alpha(\tau^i, \tau^{i+1})$ and obtain τ^{i+1}.

- **Update** λ^i and Σ_i using RM recursions:
  - $\mu_{i+1} = \mu_i + \gamma_{i+1}(\tau^{i+1} - \mu_i)$
  - $\Sigma_{i+1} = \Sigma_i + \gamma_{i+1}[(\tau^{i+1} - \mu_i)(\tau^{i+1} - \mu_i)^\top - \Sigma_i]$
  - $\log(\lambda^{i+1}) = \log(\lambda^i) + \gamma_{i+1}[\alpha(\tau^i, \tau^{i+1}) - \alpha^\star]$, $\alpha^\star$ desired acceptance rate (0.234).

AMGS: naive implementation

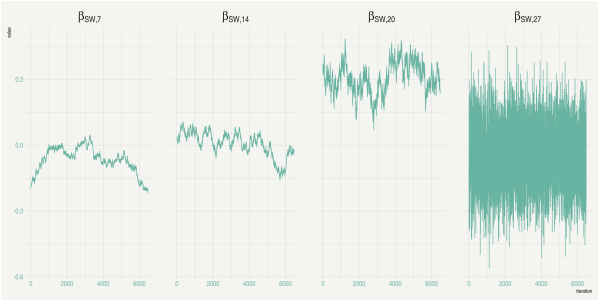
- Let τ the parameters in the *m*-th block at the *i*-th iteration; **Propose** $\tau^{i+1} \sim N(\tau^i, \lambda^i \Sigma_i)$ to accept/reject with MH probability $\alpha(\tau^i, \tau^{i+1})$ and obtain τ^{i+1}.

- **Update** λ^i and Σ_i using RM recursions:
  - $\mu_{i+1} = \mu_i + \gamma_{i+1}(\tau^{i+1} - \mu_i)$
  - $\Sigma_{i+1} = \Sigma_i + \gamma_{i+1}[(\tau^{i+1} - \mu_i)(\tau^{i+1} - \mu_i)^\top - \Sigma_i]$
  - $\log(\lambda^{i+1}) = \log(\lambda^i) + \gamma_{i+1}[\alpha(\tau^i, \tau^{i+1}) - \alpha^\star]$, $\alpha^\star$ desired acceptance rate (0.234).

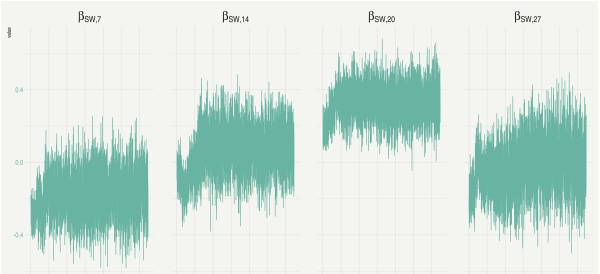
- $\{\gamma_i\}$ is a sequence of stepsizes ensuring variations of λ^i, Σ_i **vanish** wrt MCMC iterations; **Vanishing adaptation** is required for π-ergodicity of the algorithm.

AMGS: naive implementation

RW-MH traceplots
16/04/2021



AMGS traceplots
30/04/2021



AMGS: naive implementation

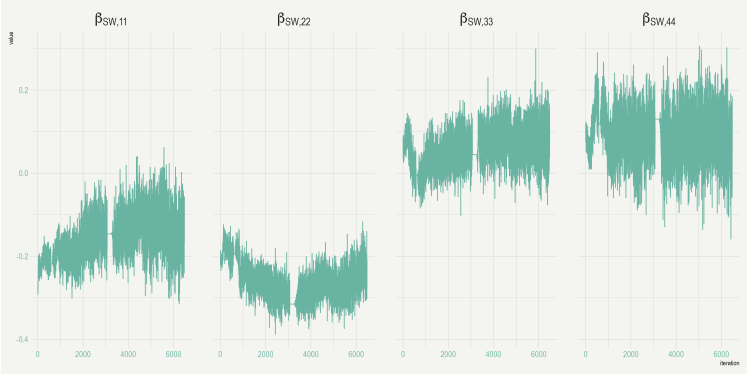
27/11/2021

- **“Over-adapting”** proves to be a source of **“bad”** values for λ and Σ
- Can lead to **‘sticky’** chains

AMGS: naive implementation

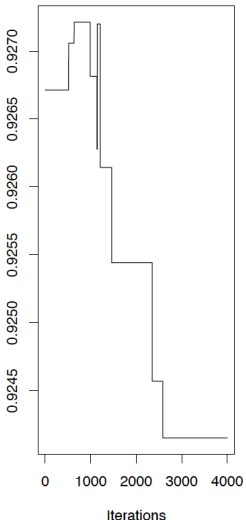
27/11/2021

- **"Over-adapting"** proves to be a source of **"bad"** values for λ and Σ
- Can lead to **'sticky'** chains



Badly tuned λ_i and Σ_i

η over-dispersion parameter (static, global parameter)



Adaptive MCMC: improved adaptation

- **Prevent over-adapting** by borrowing the **early stopping** regularization from ML applications (e.g. Zhang and Yu, 2005).
  1. Calculate the **sample average acceptance ratio** $\hat{S}_{\lambda,\Sigma}$ based on m consecutive MCMC iterations

Adaptive MCMC: improved adaptation

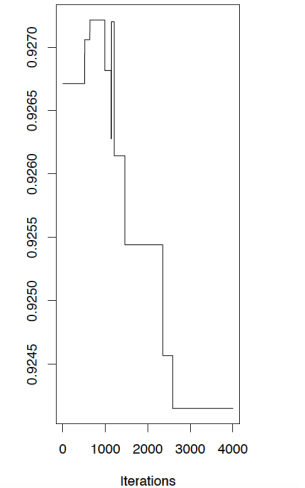
- **Prevent over-adapting** by borrowing the **early stopping** regularization from ML applications (e.g. Zhang and Yu, 2005).
  1. Calculate the **sample average acceptance ratio** $\hat{S}_{\lambda,\Sigma}$ based on m consecutive MCMC iterations
  2. If $\hat{S}_{\lambda,\Sigma} \in (\alpha_-^\star, \alpha_+^\star)$ do **not** update the RM recursions

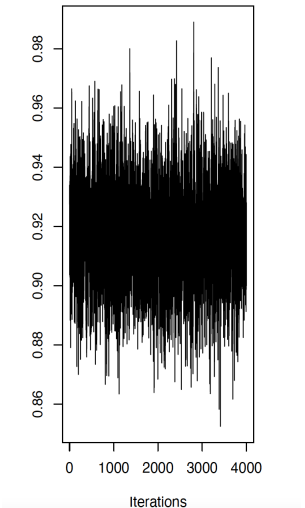
Adaptive MCMC: improved adaptation

- **Prevent over-adapting** by borrowing the **early stopping** regularization from ML applications (e.g. Zhang and Yu, 2005).
  1. Calculate the **sample average acceptance ratio** $\hat{S}_{\lambda,\Sigma}$ based on m consecutive MCMC iterations
  2. If $\hat{S}_{\lambda,\Sigma} \in (\alpha_-^\star, \alpha_+^\star)$ do **not** update the RM recursions
  3. **Stop** the RM recursions at the **end of the burn-in** period to **ensure** π-ergodicity

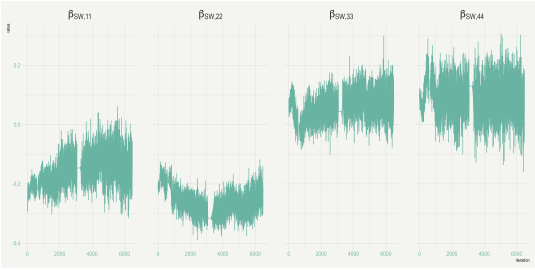
Improved tuning of λ_i and Σ_i

η over-dispersion parameter (static, global parameter)

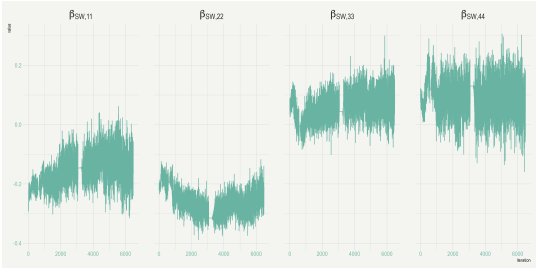




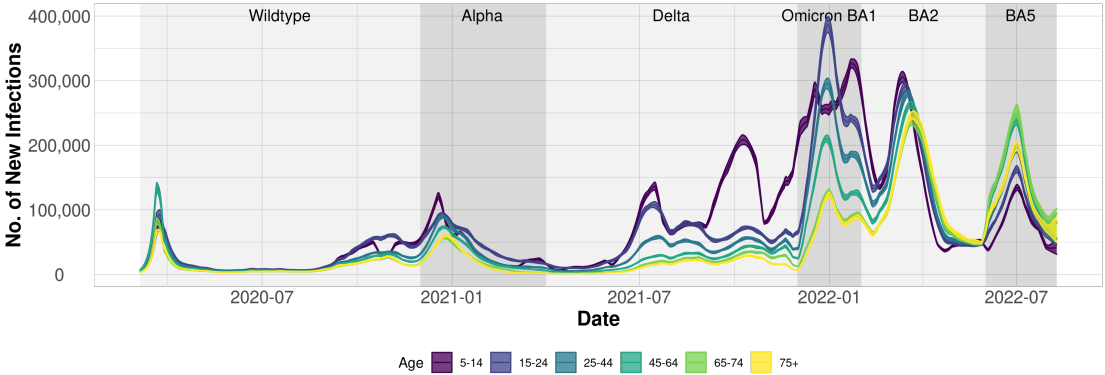
Improved tuning of λ_i and Σ_i



Improved tuning of λ_i and Σ_i

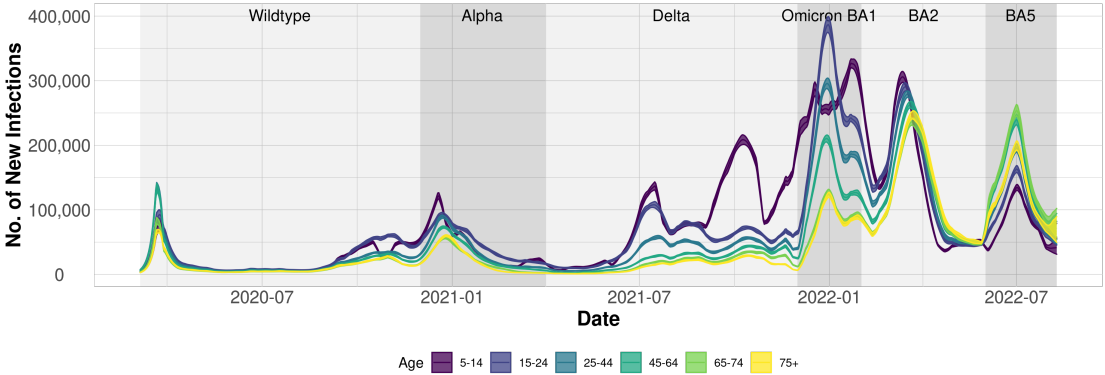


Incidence of infection



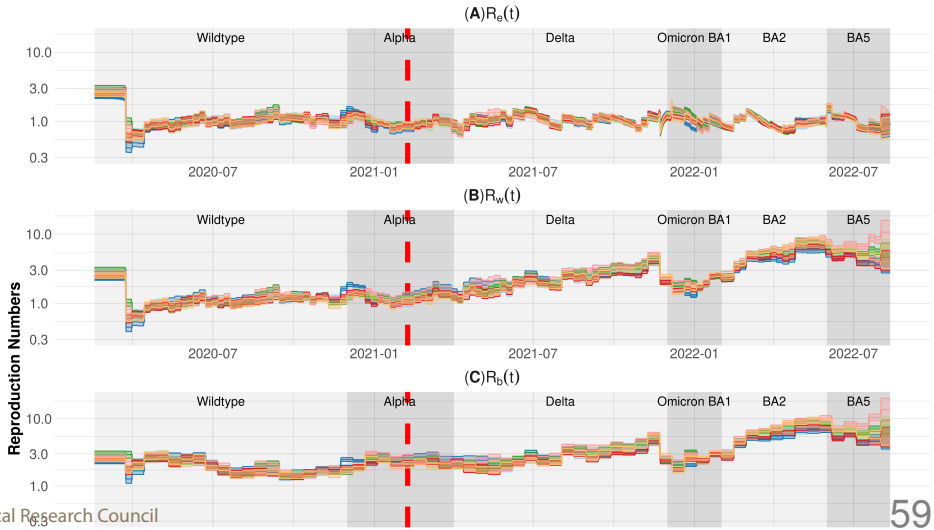
- Estimation based on over 930 days of data from: **hospital admissions**, **serological surveys** and **ONS prevalence surveys**

Incidence of infection



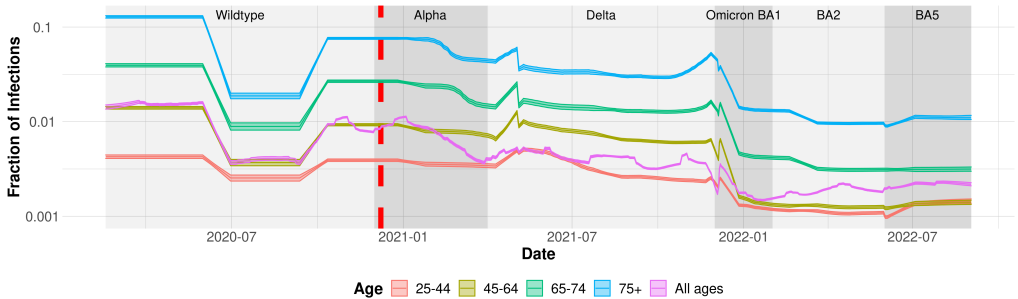
- Estimation based on over 930 days of data from: **hospital admissions**, **serological surveys** and **ONS prevalence surveys**
- Note huge impact of children in late stage of the delta wave

Transmission: effective reproduction number $R_e(t)$



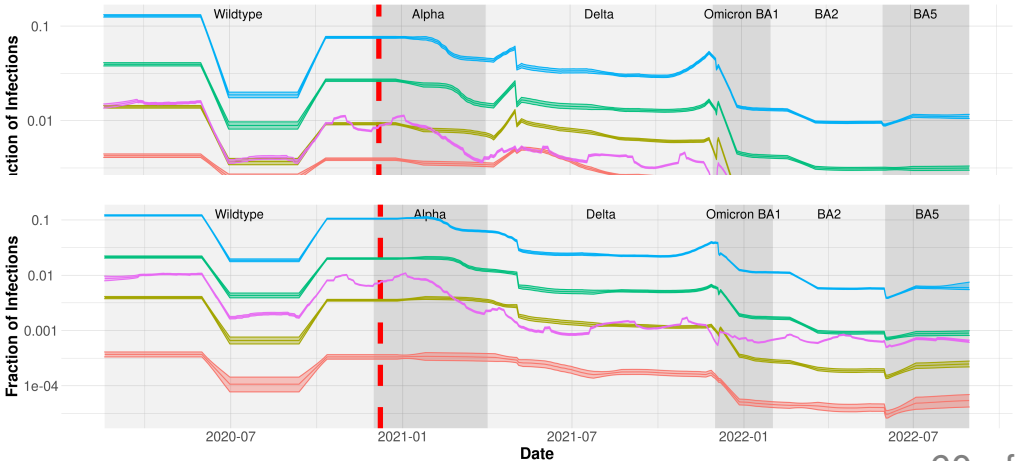
Risk of severe event - hospitalisation - mortality

IHR

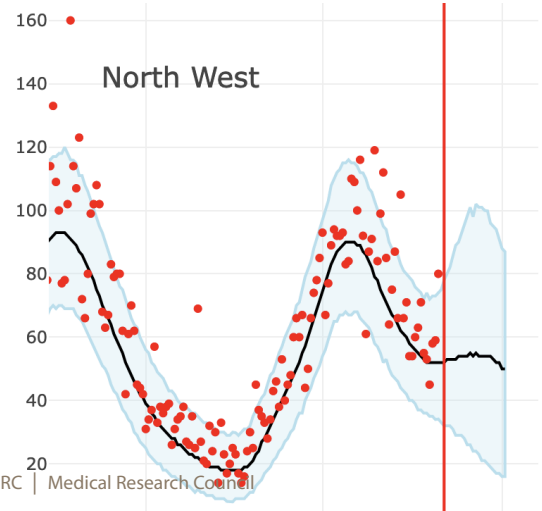


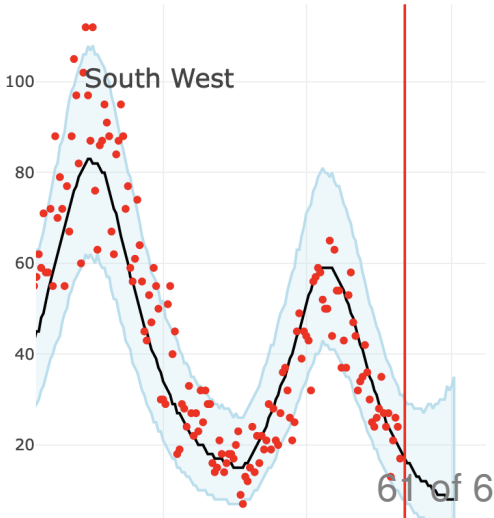
Risk of severe event - hospitalisation - mortality

IHR and IFR



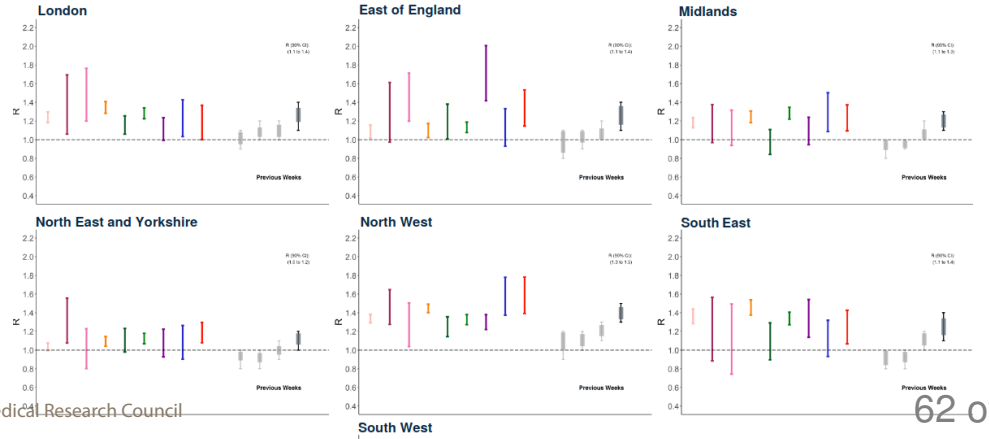
Short-term predictions: hospital admissions





Contribution to SPI-M-O consensus as the official Public Health England's model

- Government consensus on relevant indicators (*e.g.* $R(t)$) - combination of models from across academic institutions



In conclusion

- Official model of Public Health England (UKHSA) did not have luxury/resources to stop providing results and developing sophisticated MCMC
- Clever tweaking of existing algorithms **'on the fly'** - never abandoned exact inference - produce robust results under extreme time pressure
- Highlights importance of work on reliable/scalable algorithms that can be adapted swiftly in an emergency situation

Peace time

- Relaxation of non-realistic assumptions e.g. piece-wise constant $\tilde{\beta}_{t_k,r}$ /independence across regions - complicates model/inference
  - Approximating the model (e.g. spectral approximation of diffusion processes)
  - Designing more sophisticated MCMC and/or SMC methods to conduct exact inference without making non-realistic assumptions
- Automatic differentiation to employ likelihood derivatives
- Addition of **demographic stochasticity** - approximate inference

Acknowledgement

Funding from

- UKRI (MRC)
- NIHR
- UKHSA
- Bayes4Health

Analyses regularly published at https://www.mrc-bsu.cam.ac.uk/now-casting/