

# Modelling the COVID Pandemic in Ireland with Bayesian Methods in Python

Peter Nolan  
databeaker@gmail.com  
<https://github.com/dpnolan/pandemic>  
@dpnolan

Python Ireland,  
Wednesday 7 July 2025  
Workday, Dublin



From a protest at the Riksdag, the Swedish parliament, 29 April 2020

# About Me



- Born and raised in south county Dublin
- Trinity business, economics and maths
- Worked in the Dáil as a student
- London City University / Bayes Business School - Mathematical Finance M.Sc.
- Then emerging markets trading, in the middle of multiple financial crises, including Asia 1997, Russia 1998, Enron 2001, credit crisis 2007, Euro crisis 2008 - working on pricing and risk models
- National College of Ireland for M.Sc.Data Analytics last year
- Always been a writer, at college and then for the print media in Ireland, UK and elsewhere
- How information flows and is used is the commonality between the Irish Times newsroom, Leinster House, bank trading floors or social media analysis
- All this has taught me: "The future is already here. It's just not evenly-distributed yet."

[databaseaker@gmail.com](mailto:databaseaker@gmail.com)

# My Motivations

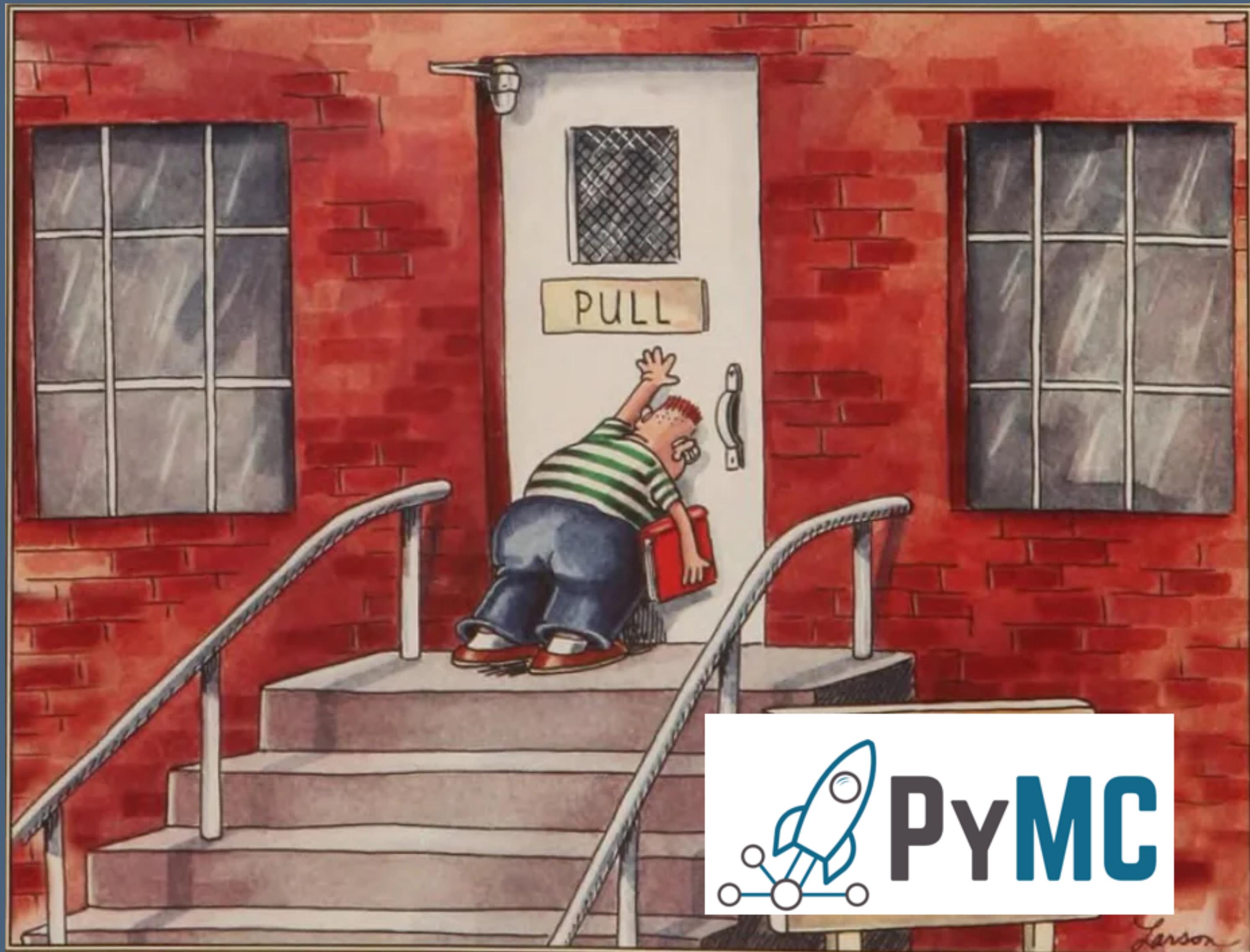


- In 2020, I was working on a major change project in Stockholm
- Nolan's Law: 'Murphy was an optimist'
- Family legacy - grandmother died of TB when my mother was 8, leaving nine children
- In Asia, bird 'flu killed an infant in Hong Kong in late 1997 and there were outbreaks throughout through Asia afterwards, now spreading in the Americas
- Fear after 9/11 and the anthrax bio-terror attacks pushed research, including on mRNA vaccines, and pandemic preparedness
- In 2005, I wrote in Magill magazine and appeared on David McWilliams' TV show to say that my 5 person software firm had done more pandemic planning than any part of the Irish government. A HSE report that autumn recommended adding a modeller
- The Swedish government were relatively relaxed in their response, private sector data analysts and traders were very risk-averse from the start
- Where and when was this photo taken?

# Research Question

- What models can we use for measuring and forecasting the evolution of epidemics and how can I understand the published model outputs?

To give a summary in one slide...

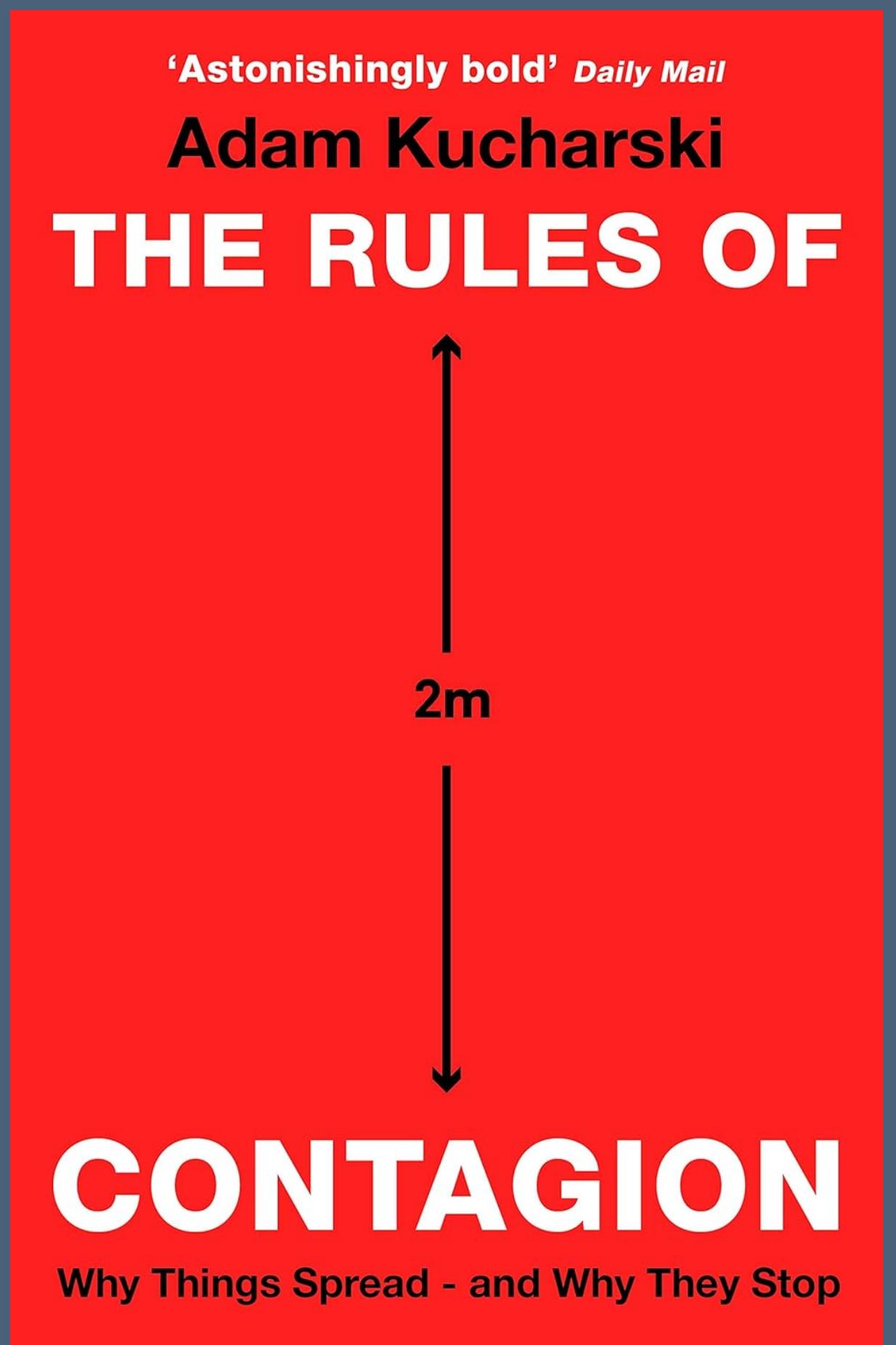


# My Approach

- Survey the literature, the models and data and policy analysis, before and during the pandemic aim to understand the terminology
- Create and run the models myself
- My experience with finance has taught me that machine learning absent domain knowledge and empirical orientation is mostly a waste of time
- Fill a gap in the Irish media and political news, as the public understanding of the modelling issues is sparse
- This analysis combines everything that I don't know about epidemiology with all I don't understand about Bayesian statistics.
- Nevertheless, our open-source community, academic publications and open data made this project possible, so show our open-source, data providers, academic and government modellers support.
- 'All models are wrong, some are useful'. Mine are more likely than most to be wrong, so feel free to comment here or later.

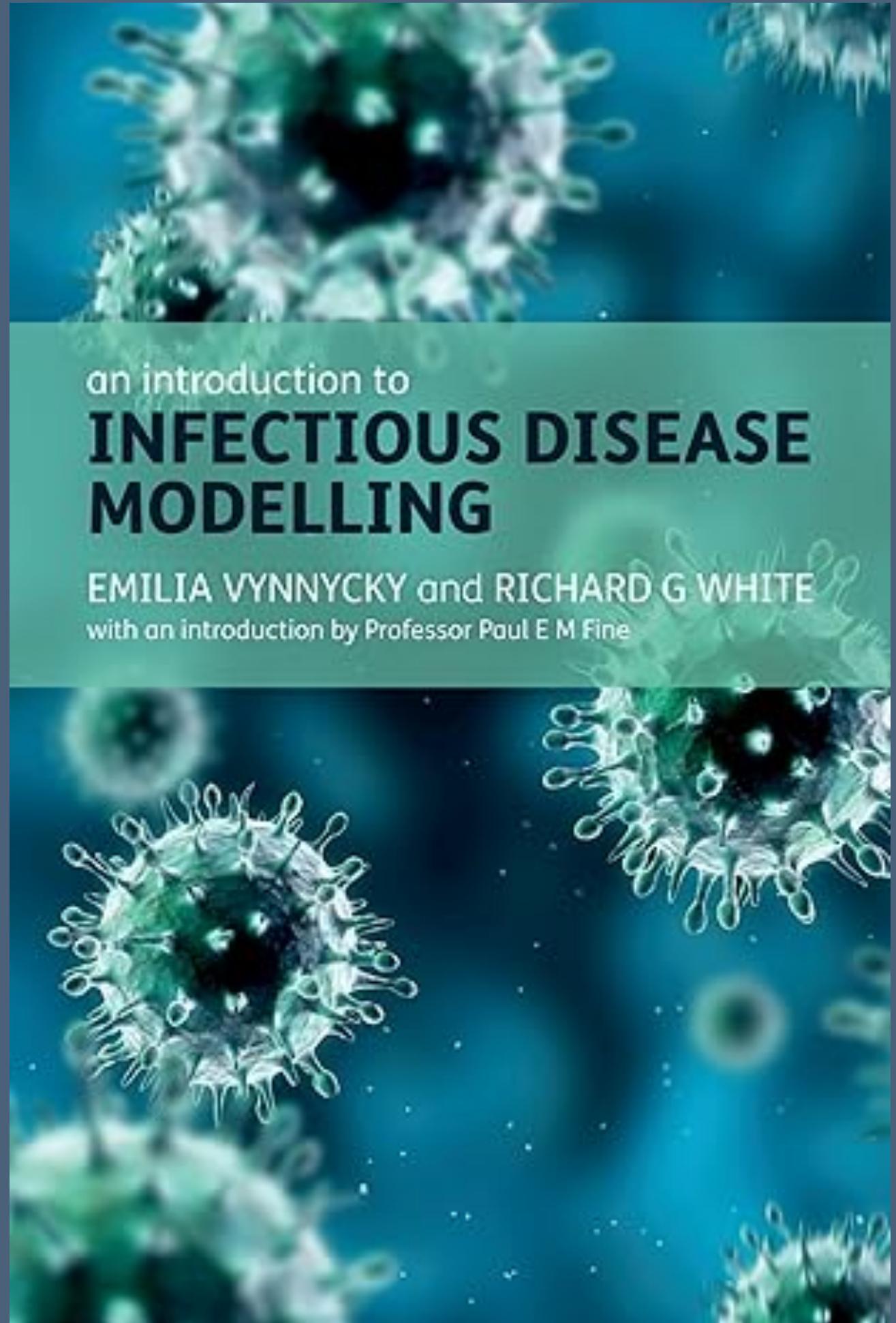
# Surveying the Literature

- Dr Adam Kucharski, mathematical epidemiologist at LSHTM
- His popular science book on epidemic models was published on 13 February 2020
- One chapter covers the century-long history of standard SIR epidemic models and he describes his work with crises like Zika virus
- Epidemic models are also applied for other purposes including forecasting shootings in inner-city America



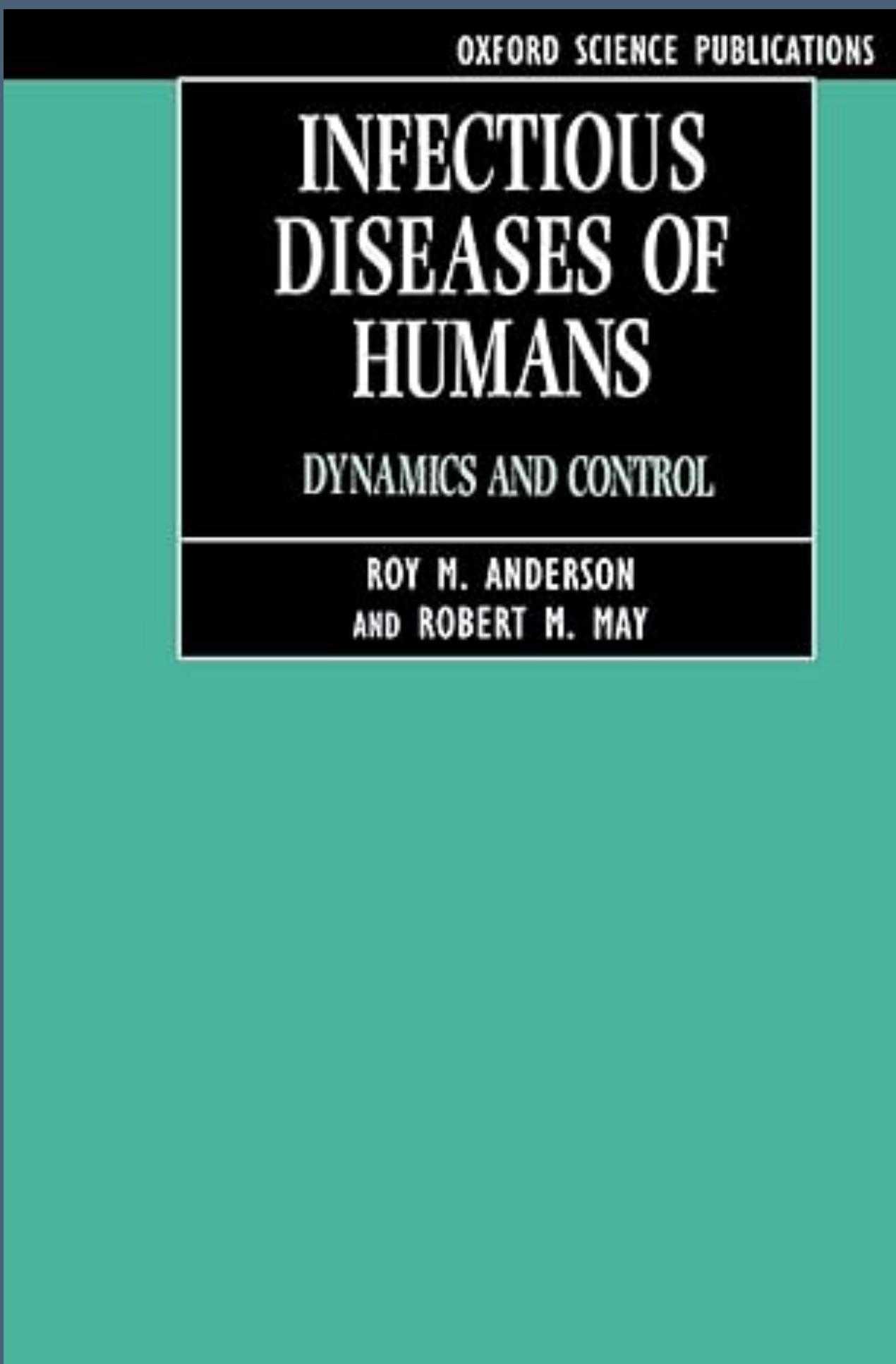
# Surveying the Literature

- Vynnycky and White give a broad survey of the mathematical models in this 2010 widely-used textbook
- Not hugely mathematical in the treatment, no code
- Starts with 'compartment' or 'population' models
- Difference and differential equations underly the population models, first used more than 100 years ago to describe disease dynamics as a function of time by Kermack-McKendrick, two Scottish public health doctors working in India
- Agent-based model individuals, simulate behaviour of individuals, but the theories are more recent, and more computationally intensive,
- Social networks models look at interactions, but are also more recent mathematics
- Personal characteristics - age, gender, recent travel, location, can dealt with by separate models, or hierarchical model, but can be included as a model variable



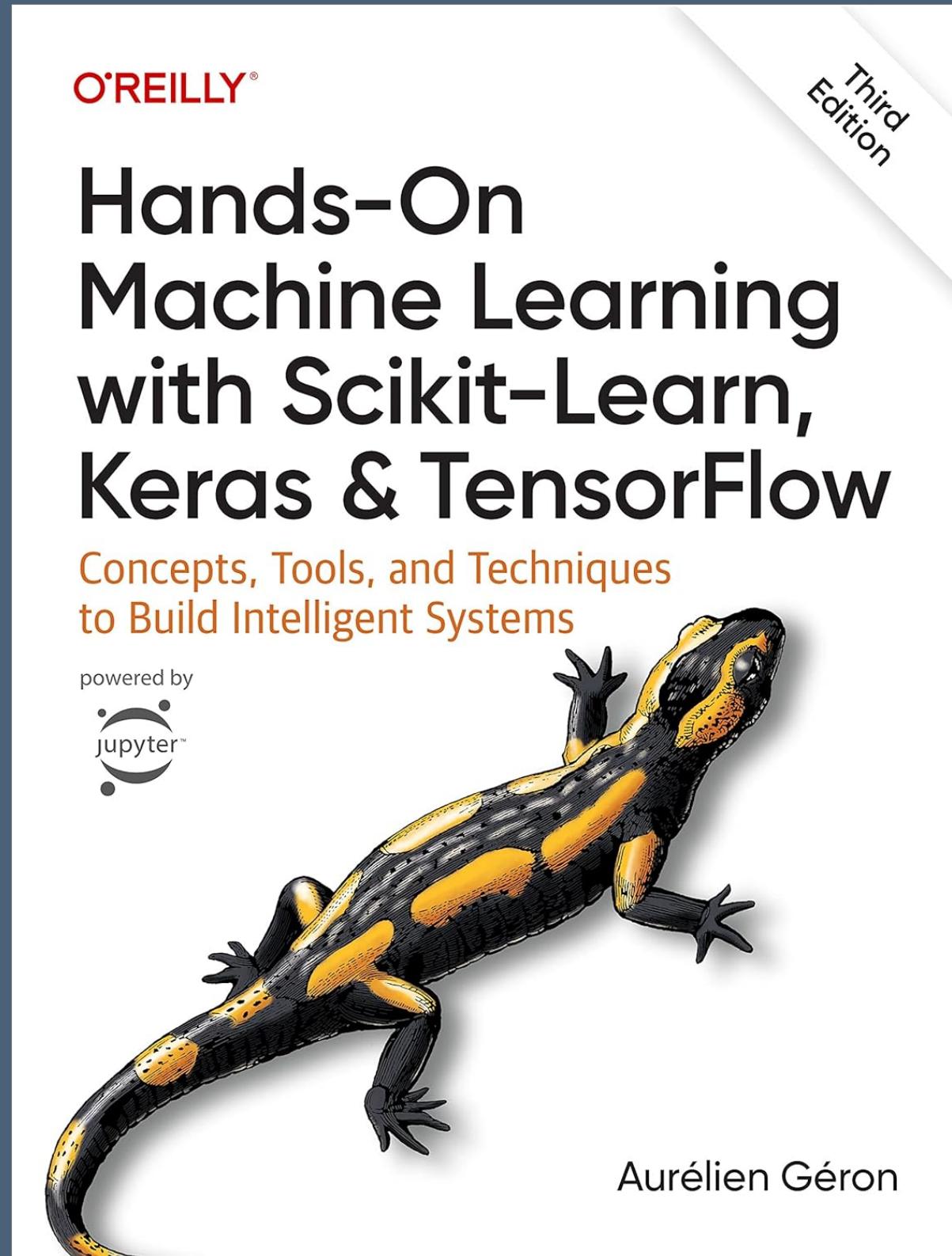
# SIR Model Dynamics

- The population models assume otherwise homogenous populations with stable parameters - population, no births, deaths, migration or travel. Also:
- No reinfection - otherwise SIRS, SIS etc.
- Disease characteristics - no drift or variants
- No changes in any vaccination
- No Non-Pharmaceutical Interventions (NPI) like social distancing
- Dynamics can be more complex: Work by Robert May et al. (1994) showed how chaotic, unpredictable behaviour arose within simple, models without many inputs, a lesson all modellers need to take on board

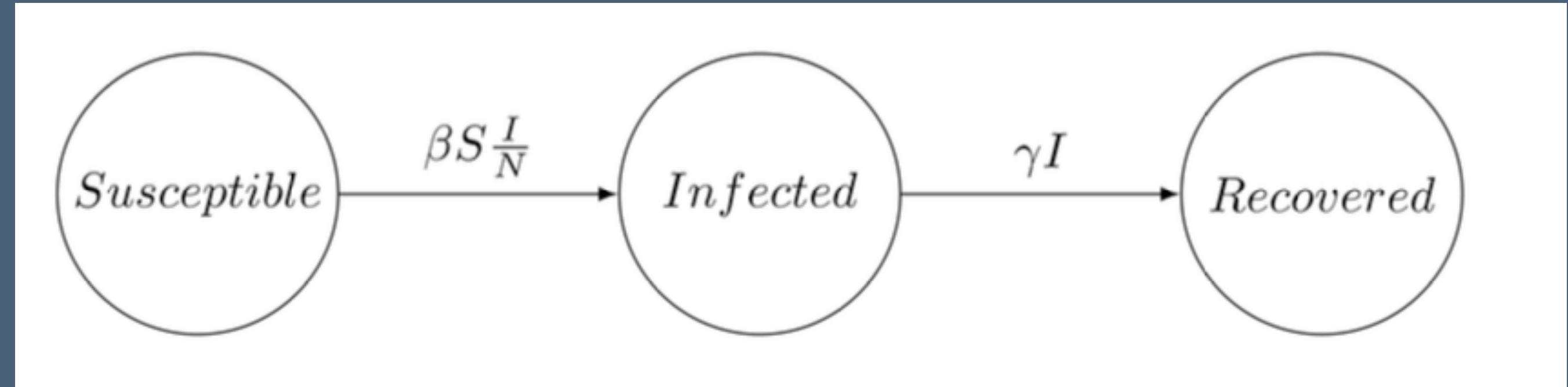


# Literature Review - Machine Learning

- The Irish Epidemiological Modelling Advisory Group (IEMAG) was an independent team advising National Public Health Emergency Team and the Chief Medical Officer. They independently produced the analysis for the government to act on.
- There were considerable differences in opinion: The government was surprised by the winter surges in infections 2020 and 2021 and reluctant to lock down further
- Some politicians gathered other data and models to challenge or ‘peer review’ IEMAG’s work. FF health minister Stephen Donnelly wanted more of a ‘big data’ and ‘artificial intelligence’.
- External consulting firms supported this under the “1 Government Centre” programme. They wanted to access many forms of data including VISA card payments, CCTV and mobile phone network mobility data.
- In spite of support from the politicians, IEMAG did not back down and the alternatives supported its scenarios
- No real Dail, press or academic scrutiny of the Irish models



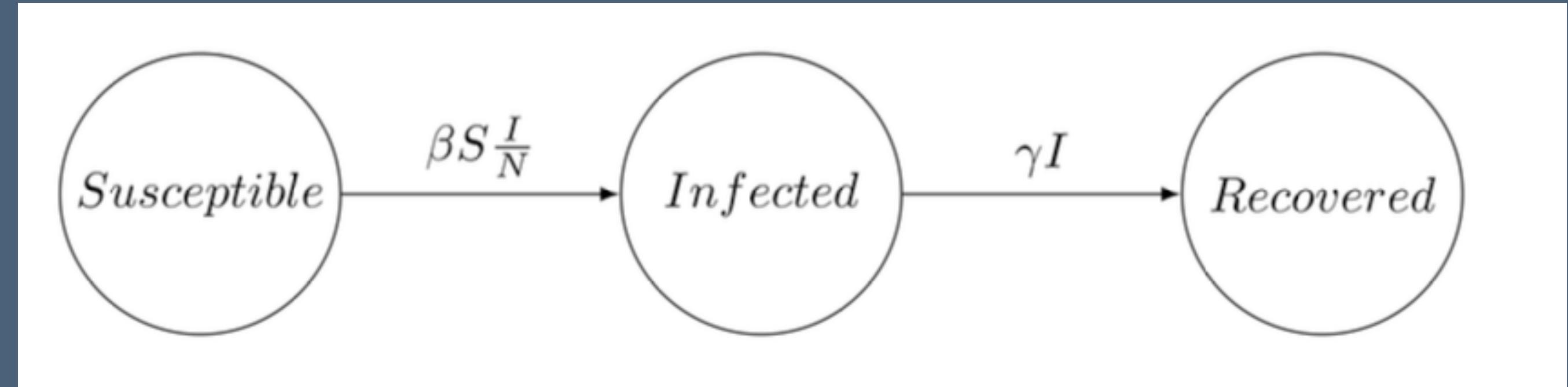
# The SIR Model - Specification



$$\begin{aligned}\frac{dS}{dt} &= -\beta S \frac{I}{N} \\ \frac{dI}{dt} &= \beta S \frac{I}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

- These divide the population into homogenous groups or compartments representing stages in the infection process
- SIR is commonly used for 'flu, stages are different for each disease e.g. SEIR, SIS, etc.
- Use difference and differential equations, math tools were available 100 years ago
- N is total population modelled, for example the population of the Republic of Ireland
- $S_t$ , Susceptible may be infected, a portion become so based on beta,  $\beta$ , the transmission rate, the constant rate of infectious contact between individuals
- $I_t$ , Infected have caught the disease and show symptoms and are infectious
- $R_t$ , Recovered, have got better or die at the recovery rate gamma,  $\gamma$ , from among the Infected  $I(t)$ .

# The SIR Ordinary Differential Equations

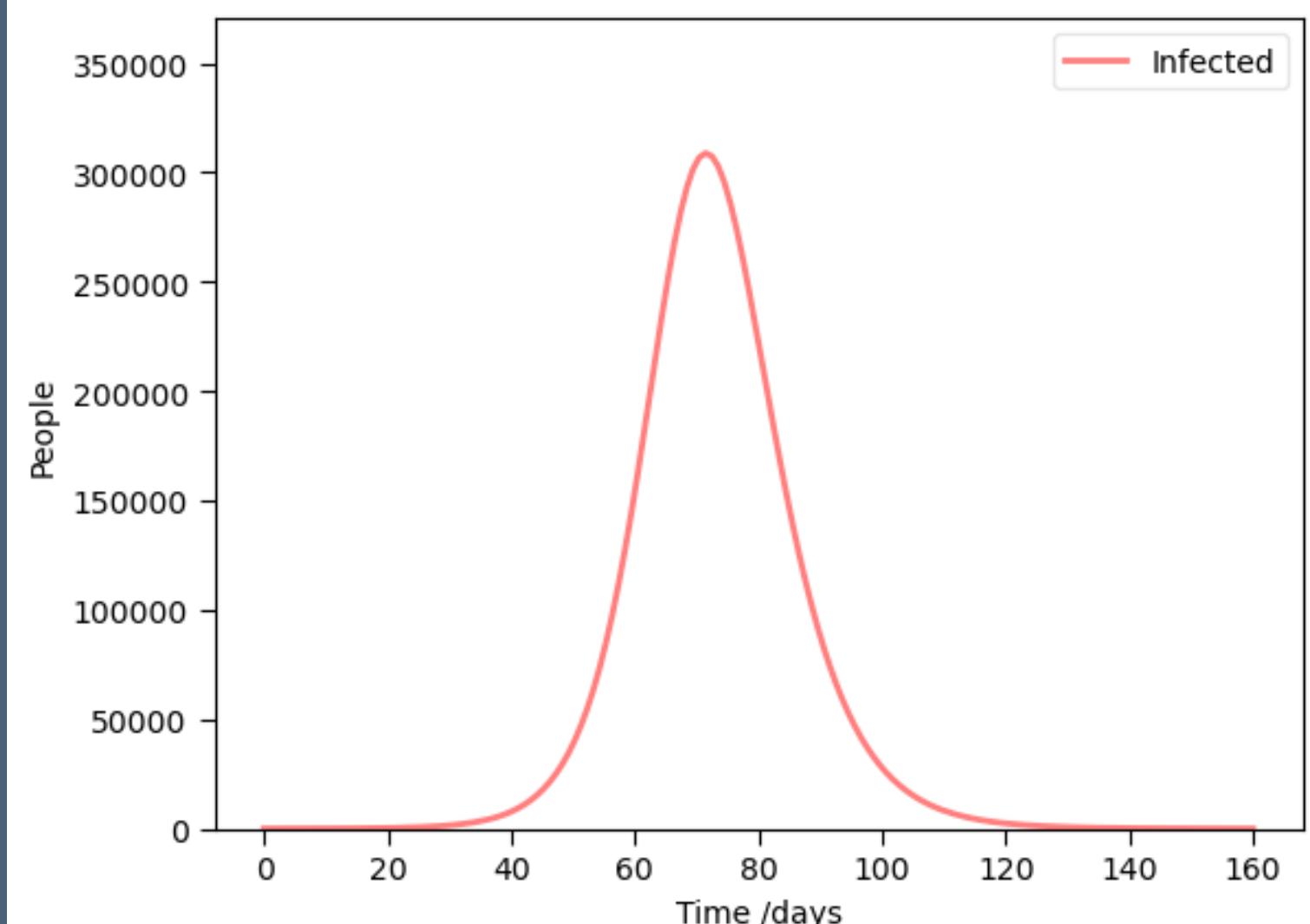
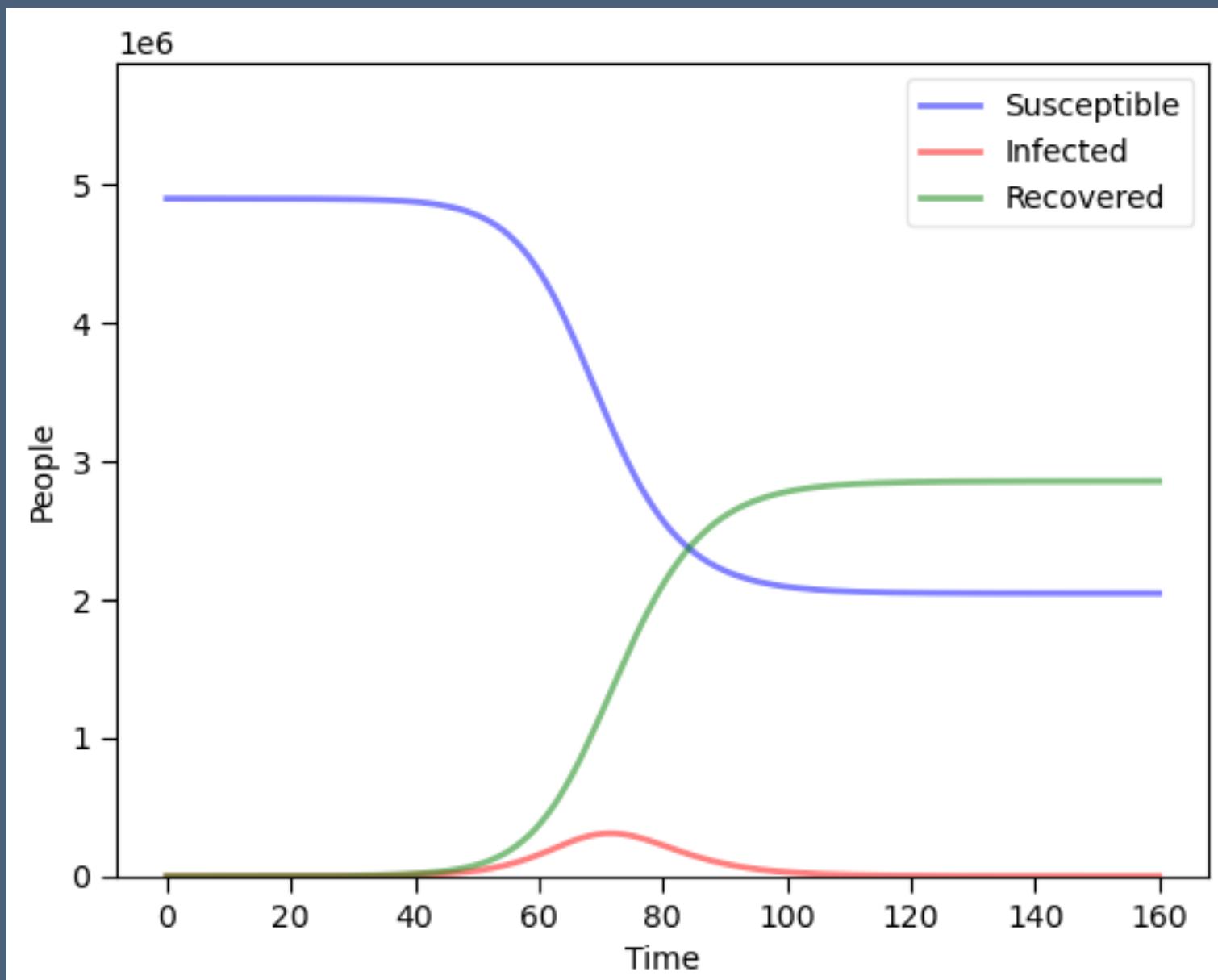


$$\begin{aligned}\frac{dS}{dt} &= -\beta S \frac{I}{N} \\ \frac{dI}{dt} &= \beta S \frac{I}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

- $S$  reduced by new infections, calculated by infected as proportion of Beta,  $\beta$ , the transmission rate, the constant rate of infectious contact between individuals, scaled by  $S(t)$  and  $I(t)$
- $I_t$ , Infected increase with new infections, transmit the infection in turn, reduced by recovery
- $R_t$ , Recovered, have got better or die at a rate gamma,  $\gamma$ , among the Infected
- $N$ , the total population at start time, and  $N = S + I + R$
- Personal characteristics - age, gender, recent travel, location, usually dealt with by separate models, or hierarchical model, but can be included as a model variable

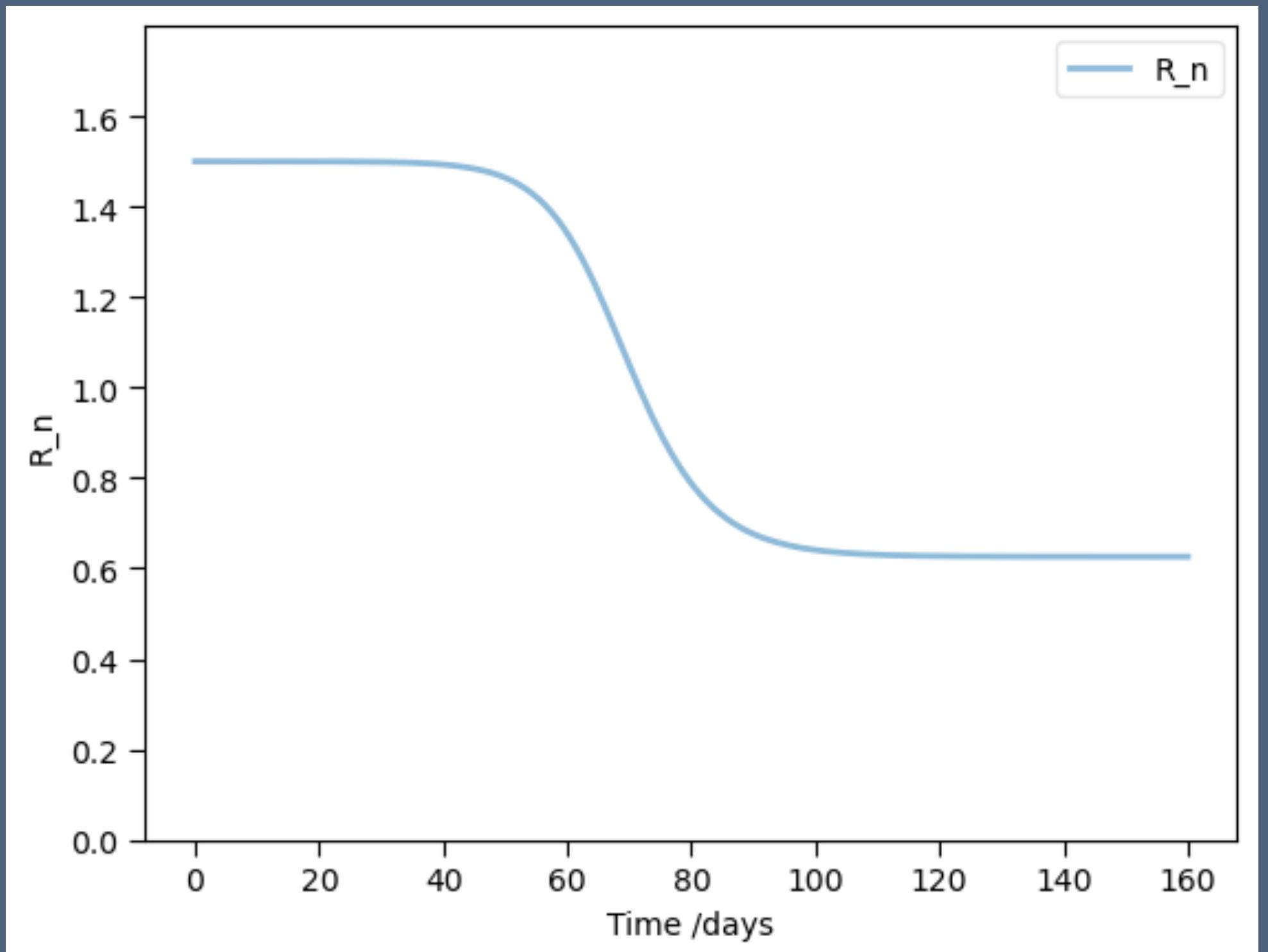
# SIR Model Dynamics

- Dynamics - typical changes over time, growth, rise and fall in infections
- See Jupyter workbook 1
- Flu example based on 'flu in Ireland,  $\beta = 0.5$ ,  $\gamma = 1/3$ ,  $N = 4.9$  million people
- Susceptible become Infected, the Infected Recover or die
- Infections peak, but as a small part of the total population at any one time



# SIR Model Dynamics

- $R_0$  ('R zero') is the average number of infections from the first infected person in the population
- $R_0 = \beta/\gamma = 0.5/(1/3) * (S/N) = 1.5$
- COVID began around 4, 1.5 to 2 is typical for seasonal 'flu, measles around 20
- $R_t$  or 'R number' is the variable used in forecasting to alter change
- IF  $R_t < 1$ , then the infection is not growing, if  $R_t > 1$ , it will grow
- Over time,  $R_t$  declines as more of the population pass through infection, as S declines
- In our example, when  $S/N = 2/3$  then,  $R_t=1$ , called the Herd Immunity Threshold. The epidemic will no longer grow, which vaccines or reinfections. Vaccination strategy targets this.
- 'Contagion' movie explainer: <https://www.youtube.com/watch?v=X-YR3UlH3aA>



# Bayesian Estimation

- Model the parameters as random variables to be estimated
- From the perspective of early to mid 2020, the underlying parameters for COVID were only going to become clear over time as evidence was gathered and published
- Making assumptions about distributions would be premature
- Some informative data is available early on and we best incorporate it into our analysis
- Expert opinion should be incorporated where useful
- Also, our uncertainty and updates to it as more data is observed should be updated over time
- All these argue for a Bayesian rather than a simple ODE approach

# SIR Model Estimation

- Textbooks such as Vynnycky & White (2010) and a shelf full of others in Trinity's library do not cover estimation, to my surprise
- Searching during 2020 for an estimation of a simple population model, I found almost nothing published
- One simple approach was to solve the Ordinary Differential Equation (ODE) system
- One non-specialist, some Russian developers at Microsoft and students, implemented this in Python using `scipy.integrate.odeint` to model and `scipy.optimize` to fit the data up to June 2020:  
See Petrova et al. at <https://github.com/shwars/SlidingSIR/tree/master>

# Surveying the Literature

- Grinsztajn et al. (2021) work in mathematical epidemiology and published a model estimating population models for COVID, although in 2021
  - Paper at <https://doi.org/10.1002/sim.9164>
  - Code + data [https://github.com/charlesm93/disease\\_transmission\\_workflow/tree/main](https://github.com/charlesm93/disease_transmission_workflow/tree/main)
  - also in the STAN documentation [https://mc-stan.org/learn-stan/case-studies/boarding\\_school\\_case\\_study.html](https://mc-stan.org/learn-stan/case-studies/boarding_school_case_study.html)
- A small, simple dataset on a 1978 'flu outbreak in an English school is commonly used in this literature, including by Grinsztajin
  - Paper <https://pmc.ncbi.nlm.nih.gov/articles/PMC1603269/pdf/brmedj00115-0064.pdf>
  - Data <https://www.kaggle.com/datasets/antonymgitau/influenza-england-1978-school-csv>
- IEMAG published a second edition of their model methodology in October 2021, updating an earlier model from May 2020.
  - Gleeson JP, Murphy TB, O'Brien J, O'Sullivan D. (2021) "A population-level SEIR model for COVID-19 scenarios (updated)," <https://assets.gov.ie/122667/8379f0cc-5be3-4c89-9a1e-3b7328ae03af.pdf>
  - Code and data [https://github.com/obrienjoey/ireland\\_covid\\_modelling](https://github.com/obrienjoey/ireland_covid_modelling)
  - Full CIDR dataset [19.geohive.ie/datasets/d8eb52d56273413b84b0187a4e9117be\\_0/about](https://19.geohive.ie/datasets/d8eb52d56273413b84b0187a4e9117be_0/about)
- A wider historical literature review is in Lit Review Epidemic Models.pdf in the GitHub at <https://github.com/dpnolan/pandemic>

# A Model to Answer the Research Question

- What is the past and present and possible future state of the epidemic?
- Infections can be observed by symptoms, a patient needing treatment, or a test result
- Inputs are the known observed infected numbers,  $I(t)$
- Candidate models will be the simple SIR population model
- Task is estimation, output for decision-makers are the historical  $R_t$  and the future cases
- Estimation done with Bayesian methods

# Bayesian Estimation

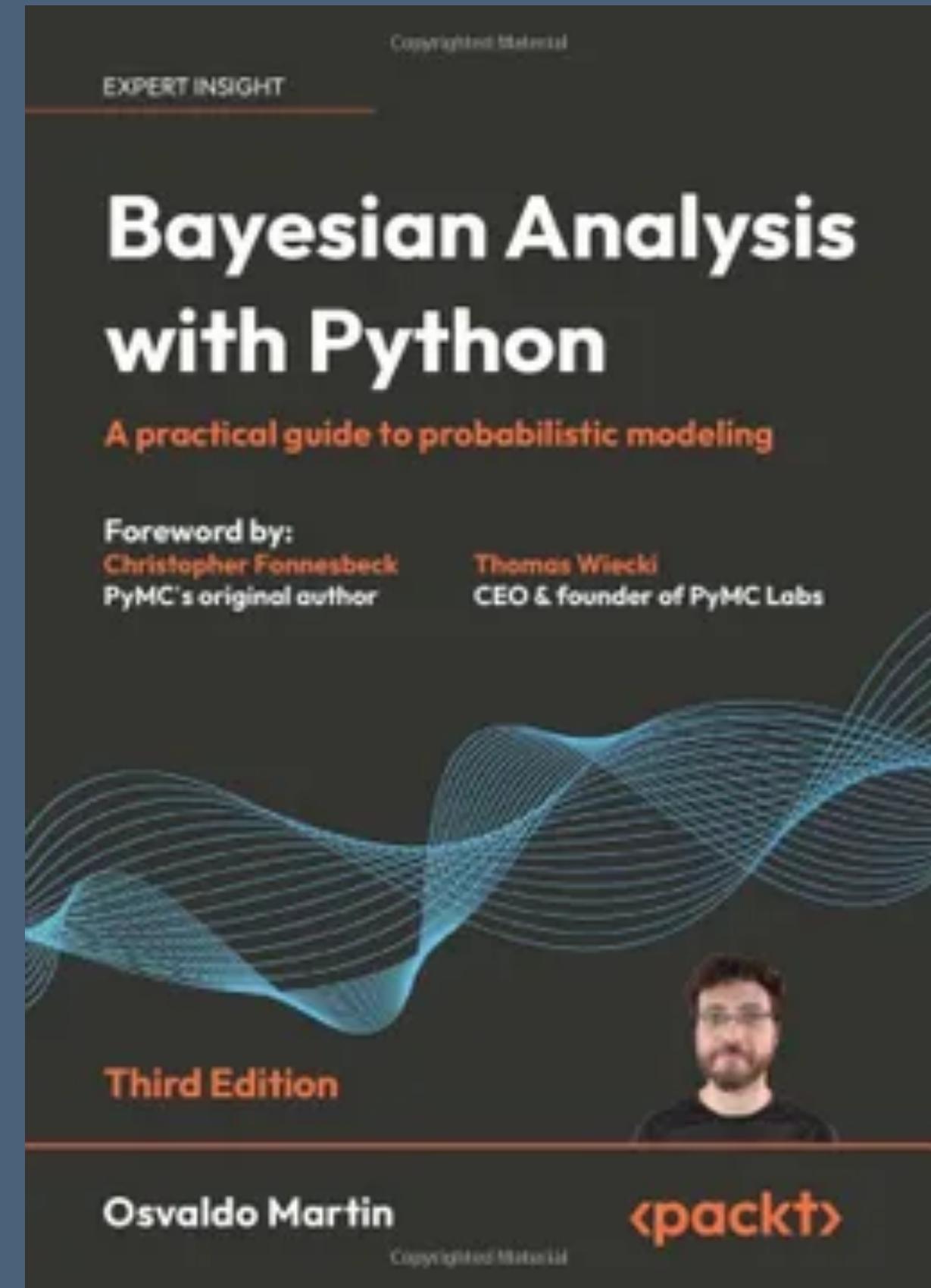
- Count discrete, non-independent events, which occur in clusters in time or space
- The first version of the model in May 2020 used the Poisson distribution.
- Exponential distribution is often used model time between outbreaks, where  $\lambda$  parameter is the time between events and also the mean of the Poisson distribution of events per year.
- Negative Binomial distribution was used instead to count the cases: was used instead in the data and model from November 2020, probably to deal with greater variation in the pandemic case numbers than the Poisson, which assumes mean and variance are the same value.
- Inverse overdispersion with exponential distribution with mean  $1/\gamma$ .
- NB uses  $I(t)$  as the mean and a parameter for inverse overdispersion phi,  $\phi$  i.e. smaller phi means more dispersion
- $\beta$  and  $\gamma$  are assumed non-negative. Not strong assumptions in these priors, so used both Half-Normal and Wald distributions for this.

$$\begin{aligned}\frac{dS}{dt} &= -\beta S \frac{I}{N} \\ \frac{dI}{dt} &= \beta S \frac{I}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

# Workflow

- Martin (2024) and PyMC core developer Thomas Wiecki suggest something like  
<https://gist.github.com/twiecki/43def0aa16ee6a23f822a124eb429958>

1. Plot the data
2. Build model
3. Check prior predictions
4. Fit models
5. Assess convergence
6. Run posterior predictive check
7. Improve model



# Markov Chain Monte Carlo

- Without a simple arithmetic Conjugate Prior from our chosen distributions, we estimate using simulation with MCMC
- Monte Carlo - simulate by random draws
- Markov Chain - future state depends on the current state only
- We sample from  $P(\theta | \text{Data})$  using the prior and likelihood distributions to set the frequency with which we sample from each region of the posterior
- From this we calculate the characteristics of our posterior distribution
- PyMC uses No U-Turns (NUTS) version of the Hamiltonian sampling algorithm, searching parameter space using gradient, then stopping search when the sample search turns back to its previous steps on the path

$$P(\theta | D) = \frac{P(D | \theta) \cdot P(\theta)}{P(D)}$$

$$P(\theta | D) \propto P(D | \theta) \cdot P(\theta)$$



A Student's Guide to  
**BAYESIAN STATISTICS**

Ben Lambert



# Python Libraries

- Usual data analysis libraries
- SciPy for the old Fortran odeint() solver
- PyMC for running the MCMC
- STAN language was used with R for Bayesian models
- ArviZ for MCMC diagnostics and metrics
- However, much more of the published models run on R, which seems favoured by epidemiologists, and often uses the STAN language



# Simulated Dataset

- See Jupyter Notebook 2

# School Dataset

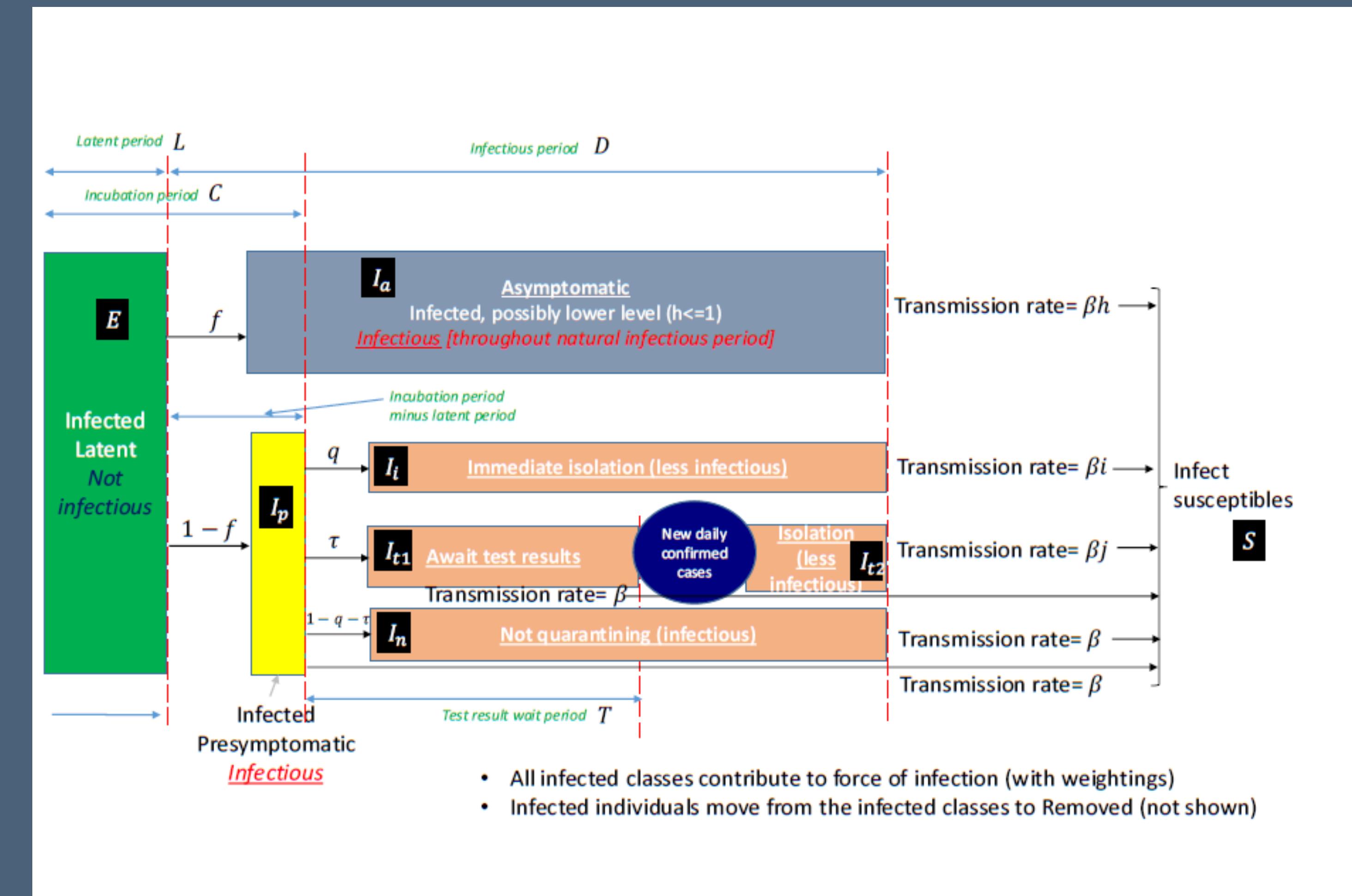
- See Jupyter Notebook 3

# IEMAG Data

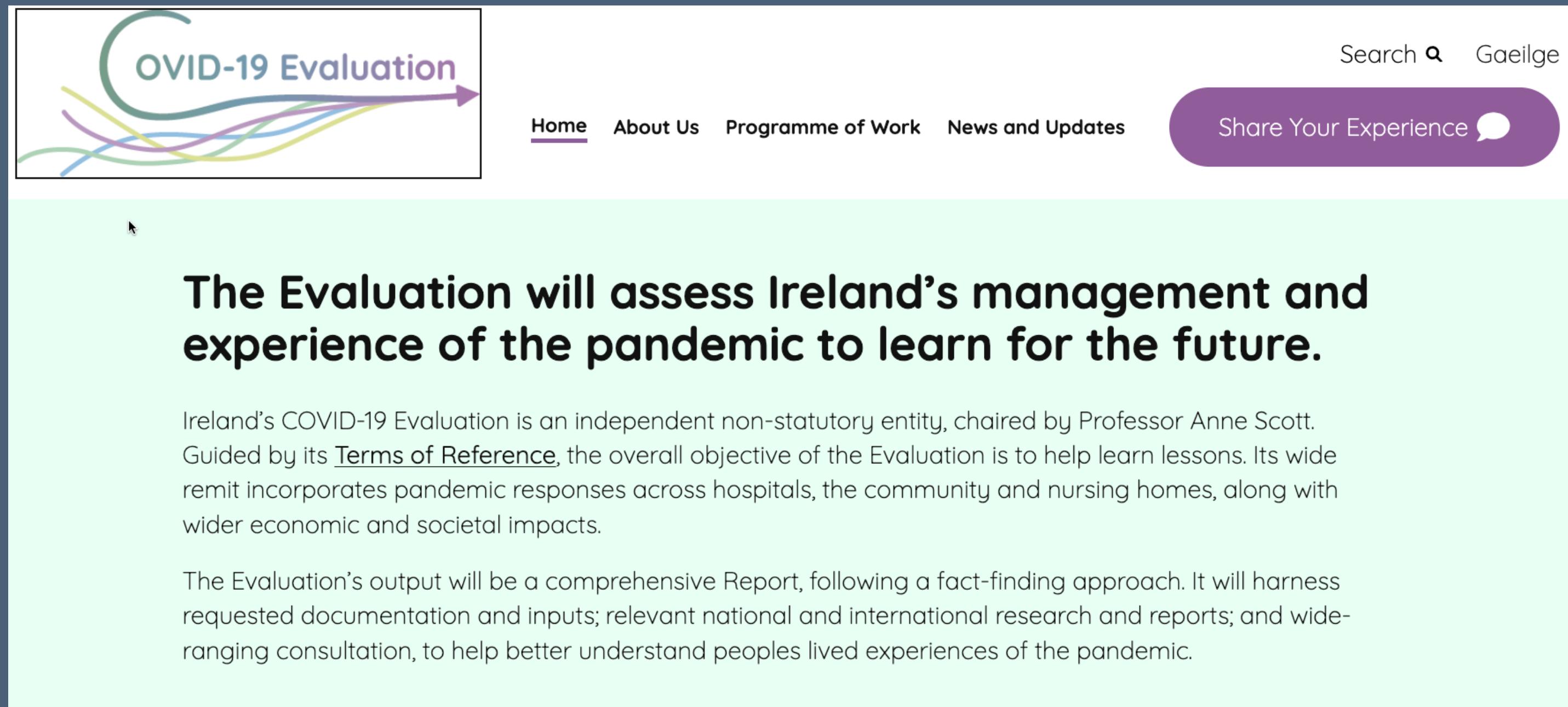
- See Jupyter Notebook 4

# Next Steps

- Estimate full IEMAG model: Production model uses SEIR, but 5 different populations for I, 4 transmission rates
- Compare with Northern Ireland's similar epidemic model, perhaps in one hierarchical model
- Compare with UK model population <https://www.gov.uk/government/publications/reproduction-number-r-and-growth-rate-methodology/reproduction-number-r-and-growth-rate-methodology#fnref:1>
- Implement age-based model
- Add economic effects



# The Irish COVID Enquiry



The screenshot shows the homepage of the COVID-19 Evaluation website. At the top left is the logo 'COVID-19 Evaluation' with stylized wavy lines. The top right features a search bar, a 'Gaeilge' link, and a purple button labeled 'Share Your Experience'. A navigation menu below the logo includes 'Home' (underlined), 'About Us', 'Programme of Work', and 'News and Updates'. The main content area has a light green background and displays the text: 'The Evaluation will assess Ireland's management and experience of the pandemic to learn for the future.' Below this, a paragraph explains the Evaluation's purpose and scope, mentioning its independence, chairperson Professor Anne Scott, and its focus on learning lessons from pandemic responses across various sectors. Another paragraph describes the output as a comprehensive Report using a fact-finding approach, incorporating documentation, research, and consultation.

**The Evaluation will assess Ireland's management and experience of the pandemic to learn for the future.**

Ireland's COVID-19 Evaluation is an independent non-statutory entity, chaired by Professor Anne Scott. Guided by its [Terms of Reference](#), the overall objective of the Evaluation is to help learn lessons. Its wide remit incorporates pandemic responses across hospitals, the community and nursing homes, along with wider economic and societal impacts.

The Evaluation's output will be a comprehensive Report, following a fact-finding approach. It will harness requested documentation and inputs; relevant national and international research and reports; and wide-ranging consultation, to help better understand peoples lived experiences of the pandemic.

- The Irish Epidemiological Modelling Advisory Group (IEMAG) team head Prof. Philip Nolan discusses his experience at an RIA panel, commenting that the 'follow the science' approach was authoritarian <https://www.youtube.com/watch?v=Estl2VBSKcc>
- Irish enquiry into the pandemic is now under way by a panel of academics scrutinising the plans and responses, but with relatively limited powers, no cross-examination and no power of compel appearance. Contact details are online at <https://www.covid19evaluation.ie/>
- My own view is that the analysis needs prior preparation, probably with nominated specialists and a menu of data sources, models and tools, with the data infrastructure set up in advance

# Conclusions

- PyMC can estimate simple SIR models, coming close to published results or simulations
- Ecosystem of libraries and published work for epidemiology in R and STAN is larger
- IEMAG2 uses measurements from health system processes and the SIR models
- SIR models are the transparent and useful tool for measuring and forecasting the pandemic, perhaps THE most useful
- What comes next? Bird 'flu remains a threat and spreads through farmed and wild animal populations, with more than half the Common Terns breeders dying off here in Dublin Port <https://birdwatchireland.ie/devastating-bird-flu-impacts-on-irish-seabirds-revealed-in-new-study/>



# Modelling the COVID Pandemic in Ireland with Bayesian Methods in Python

Peter Nolan  
databeaker@gmail.com  
<https://github.com/dpnolan/pandemic>  
@dpnolan

Python Ireland,  
Wednesday 7 July 2025  
Workday, Dublin



From a protest at the Swedish Parliament, 29 April 2020