

Kao group and COVID – lessons learned

Dr Anthony Wood

CERSE Meeting, Roslin Institute

31 May 2023

Me

- 2011 - 2015 **MSci Theoretical Physics** (Nottingham)
- 2015 - 2019 **PhD Theoretical Physics** (Edinburgh)
 - *Nonequilibrium statistical mechanics*
 - *Mathematica, MATLAB, pen and paper*
 - Learned: how to go about solving a problem
- 2019 - 2021 **Industry (fintech)**
 - Automating *reconciliation* of trading data from different sources (traders, banks, Bloomberg etc)
 - *SQL, Excel*
 - Learned: how to deal with high-freq/high-volume/messy/“real” data
- 2021 - **Postdoc @ Roslin, Kao group**
 - Applying data science to understand patterns in infectious diseases
 - *R*
 - Learned: how to pull insightful things from data

Kao group

Mix of vets, mathematicians, physicists, software engineers, computer scientists

Present

- Rowland Kao
- Anthony Wood
- Tijani Sulaimon
- Anna Gamza
- ★ Chris Banks
- Ewan Colman
- ★ Aeron Sanchez
- Muhammad Bilal

★ Software Engineers

Past

- Paul Bessell
- Anne-Sophie Ruget
- ★ Tom Doherty
- Sam Lycett
- Gianluigi Rossi
- Jess Enright
- Gavrila Puspitarani

Collaborators

- Anne Marie MacKintosh
- Martine Stead

Models of infectious disease transmission (often informed by “big” data)

• Networks

- Derived from populations, trading patterns, sewers, commutes, Twitter...

• Genetic sequences

- Comparing samples to infer who-infected whom

Current diseases

- Rift Valley fever
- Avian influenza
- Bovine viral diarrhoea
- “Rumours and behaviour”
- Bovine TB
- COVID-19

What happened in COVID-19?

Kao group

- Early access to detailed data
- Experience in disease modelling

Kao group

- Early access to detailed data
 - Experience in disease modelling

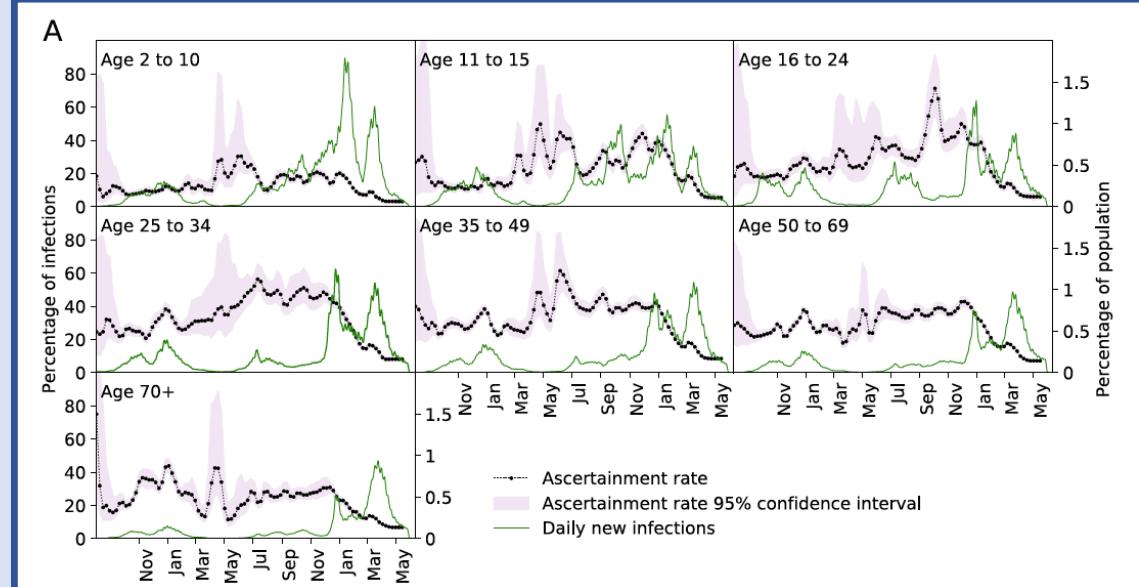
Modelling case ascertainment

Estimating true number
of infections as opposed
to cases

Colman et al. 2023

Journal of Theoretical Biology

<https://doi.org/10.1016/j.jtbi.2022.111333>

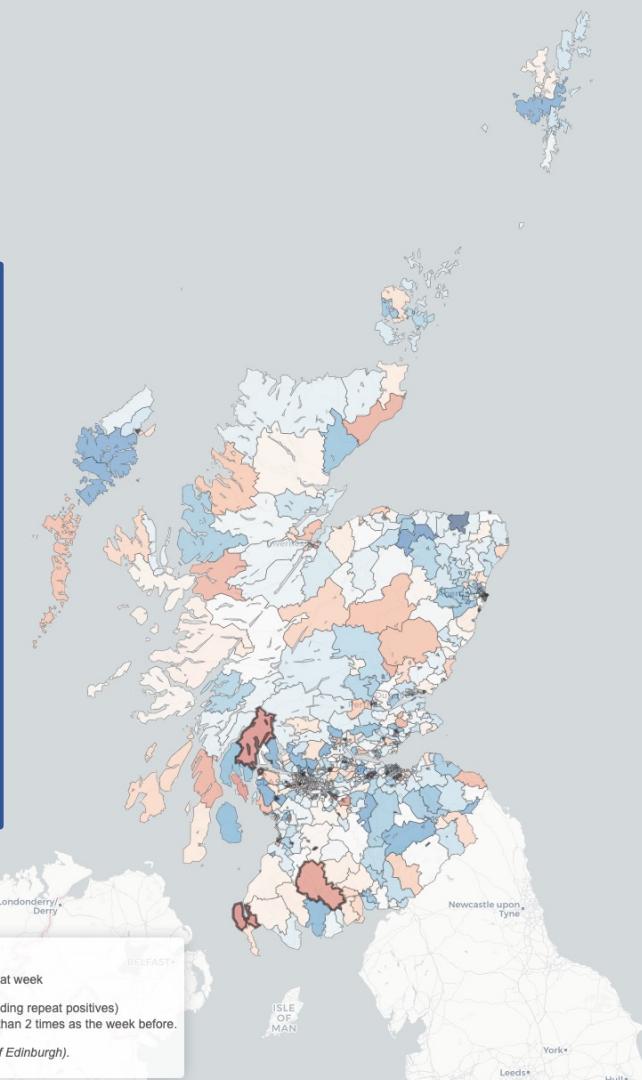
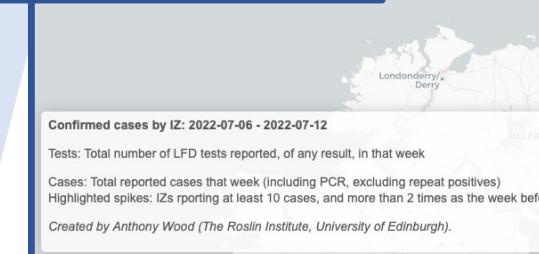
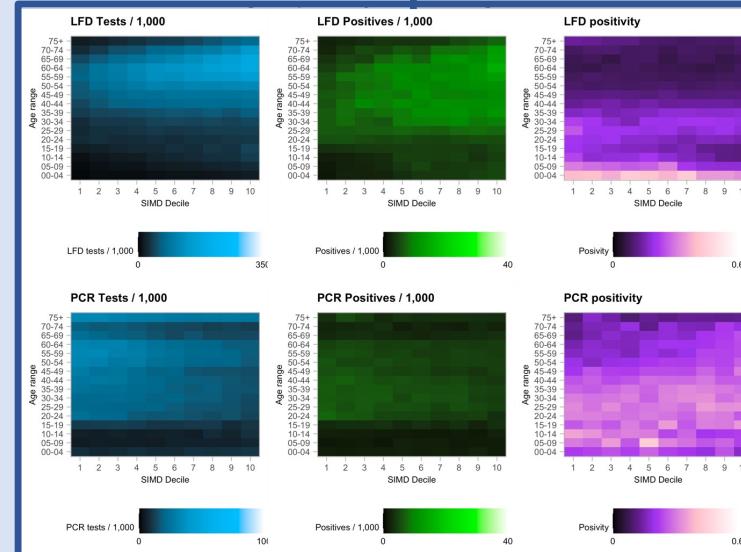


Kao group

- Early access to detailed data
- Experience in disease modelling

Modelling case ascertainment
Estimating true number of infections as opposed to cases

Real-time analysis of COVID-19 outcomes
Emphasis on variation with socioeconomic deprivation



Kao group

- Early access to detailed data
- Experience in disease modelling

Modelling case ascertainment

Estimating true number of infections as opposed to cases

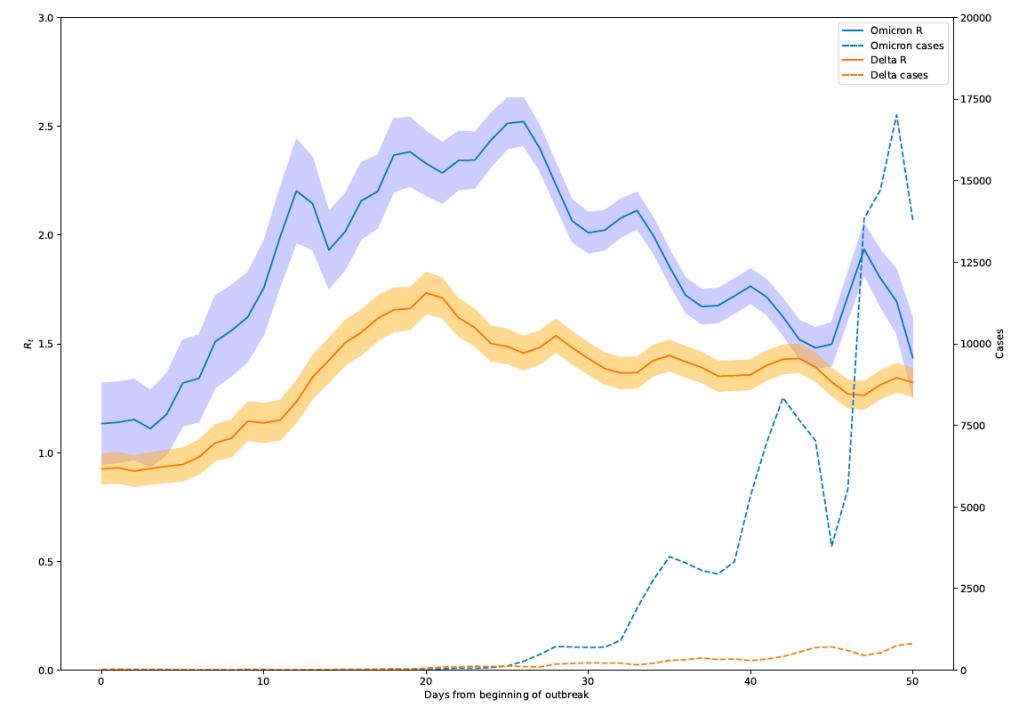
R-value estimates

Using existing case numbers to estimate past + present R-value locally

Real-time analysis of COVID-19 outcomes

Emphasis on variation with socioeconomic deprivation

Chris Banks, Aeron Sanchez



Kao group

- Early access to detailed data
- Experience in disease modelling

Modelling case ascertainment

Estimating true number of infections as opposed to cases

Real-time analysis of COVID-19 outcomes

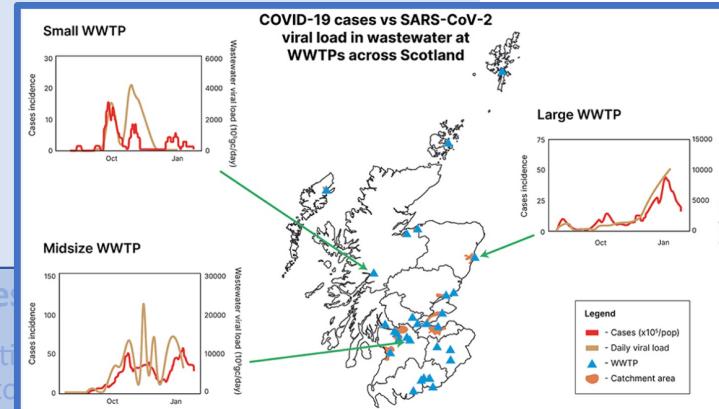
Emphasis on variation with socioeconomic deprivation

Models of wastewater shedding

Prevalence vs shedding concentration

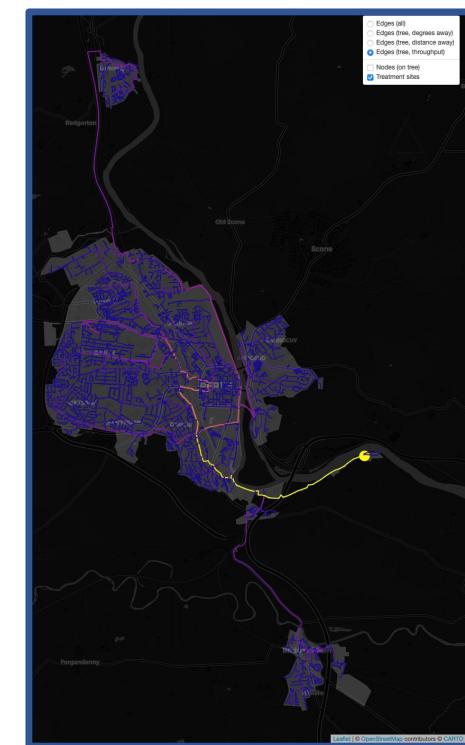
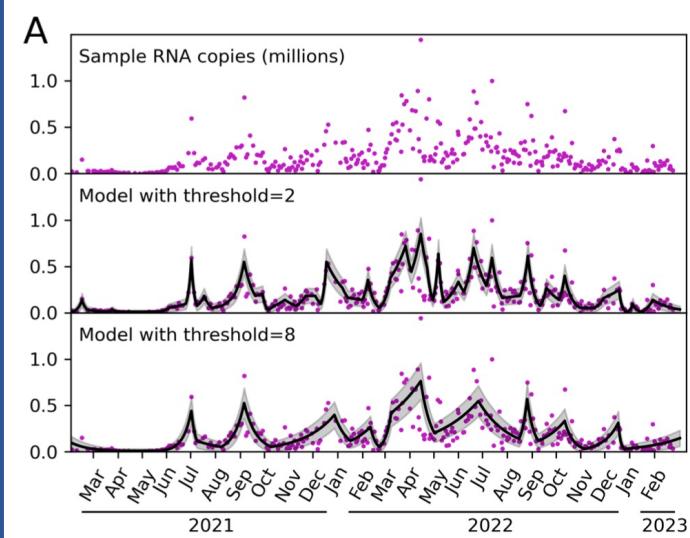
Optimal monitoring strategies with limited budget

Fitzgerald, Rossi et al
Environmental Science & Technology
<https://doi.org/10.1021/acs.est.1c05029>



Colman, Kao 2022
Medrxiv preprint

<https://doi.org/10.1101/2023.03.07.23286904>



Anthony Wood, Jess Enright, Rowland Kao
Ongoing work

Kao group

- Early access to detailed data
- Experience in disease modelling

Modelling case ascertainment

Estimating true number of infections as opposed to cases

Real-time analysis of COVID-19 outcomes

Emphasis on variation with socioeconomic deprivation

Models of wastewater shedding

Prevalence vs shedding concentration

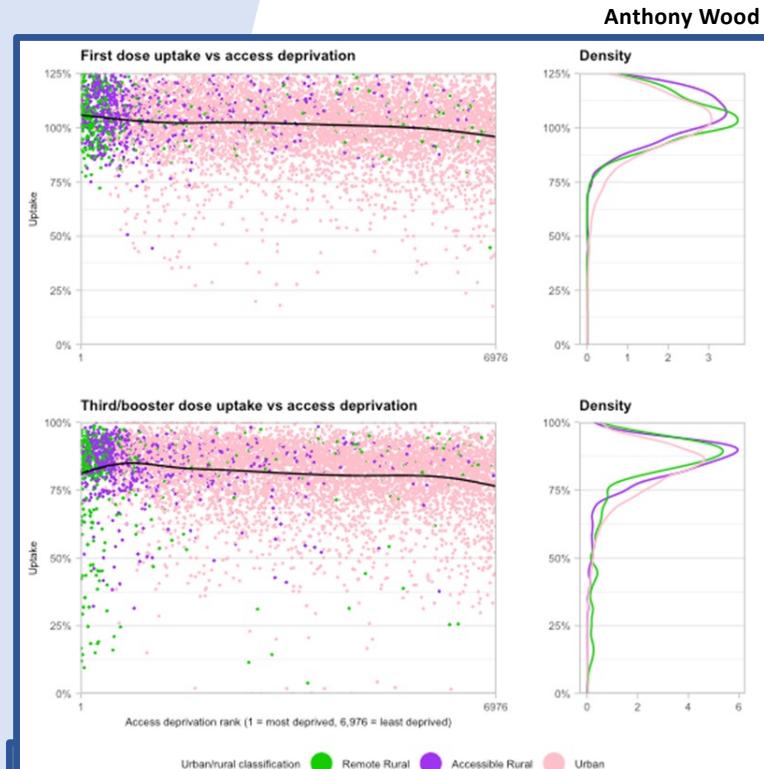
Optimal monitoring strategies with limited budget

R-value estimates

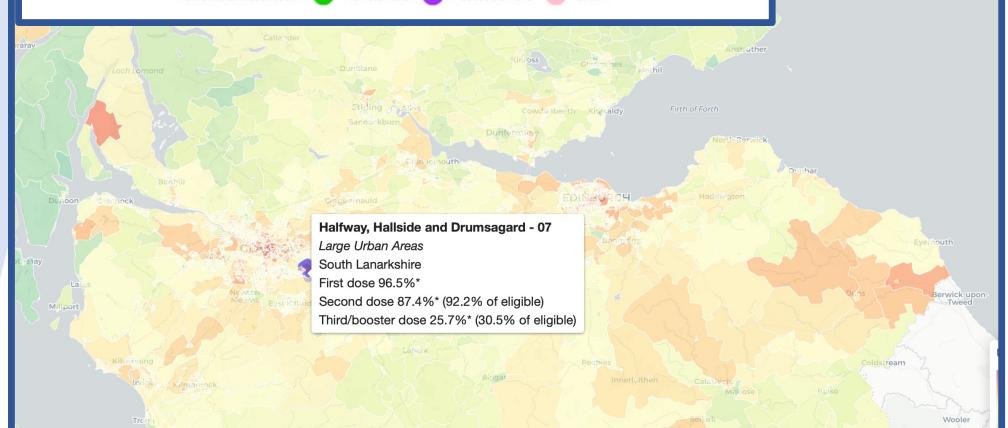
Using existing case numbers to estimate past + present R-value locally

Real-time monitoring of vaccine uptake

Highlighting regions of “low uptake” (hesitancy) and “slow uptake” (logistics)



Anna Gamza (in prep.)



Kao group

- Early access to detailed data
- Experience in disease modelling

Modelling case ascertainment
Estimating true number of infections as opposed to cases

R-value estimates
Using existing case numbers to estimate past + present R-value locally

Real-time analysis of COVID-19 outcomes
Emphasis on variation with socioeconomic deprivation

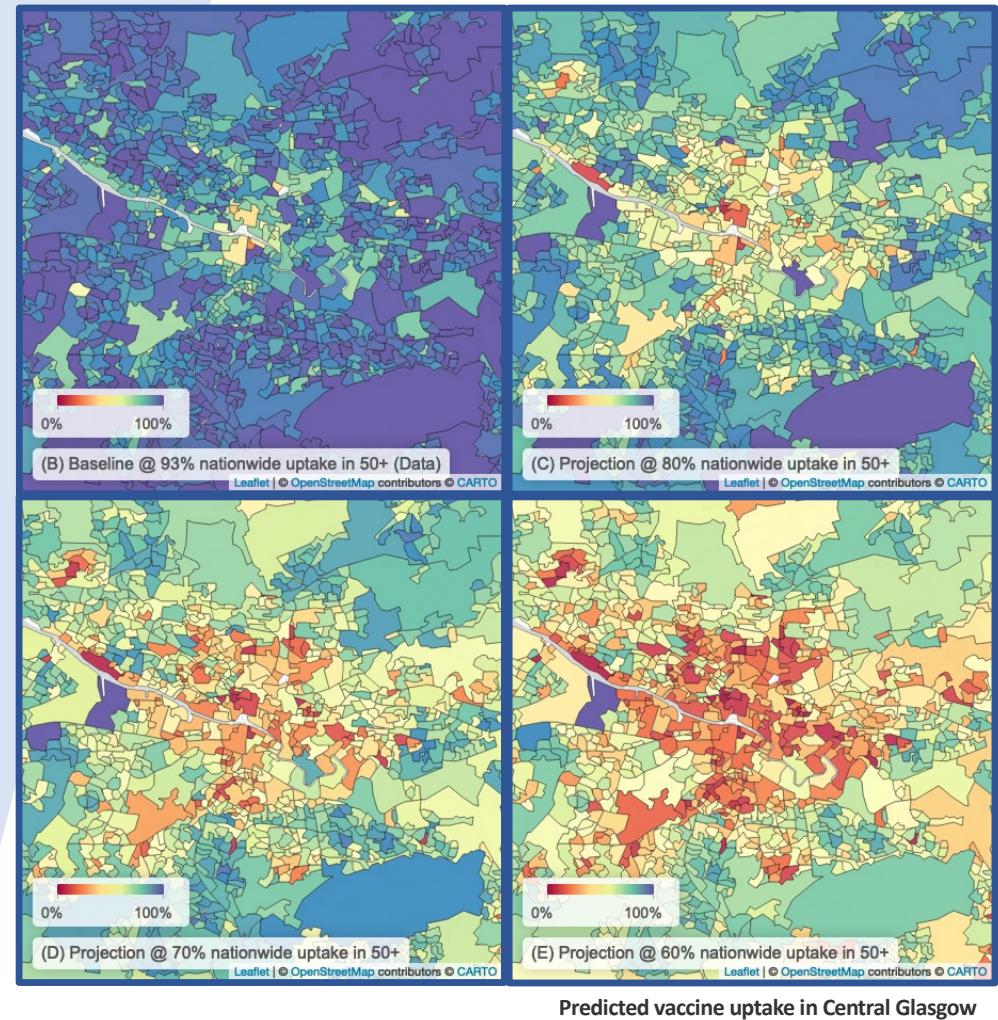
Real-time monitoring of vaccine uptake
Highlighting regions of “low uptake” (hesitancy) and “slow uptake” (logistics)

Models of wastewater shedding
Prevalence vs shedding concentration
Optimal monitoring strategies with limited budget

Predictive models of future vaccine uptake

Predicting variation in uptake “fall-off” with future boosters

Wood et al.
Medrxiv preprint
<https://doi.org/10.1101/2022.08.30.22279415>



Kao group

- Early access to detailed data
- Experience in disease modelling

SCoVMod: Scotland's Coronavirus Model

Individual-level model of disease spread across Scotland

Forecasts under different restrictions

Modelling case ascertainment

Estimating true number of infections as opposed to cases

R-value estimates

Using existing case numbers to estimate past + present R-value locally

Real-time analysis of COVID-19 outcomes

Emphasis on variation with socioeconomic deprivation

Real-time monitoring of vaccine uptake

Highlighting regions of “low uptake” (hesitancy) and “slow uptake” (logistics)

Models of wastewater shedding

Prevalence vs shedding concentration

Optimal monitoring strategies with limited budget

Predictive models of future vaccine uptake

Predicting variation in uptake “fall-off” with future boosters

Cattle disease model, rapidly adapted (early 2020) to COVID/humans by RSE

- Test driven development
- Paired programming

Banks et al. 2022

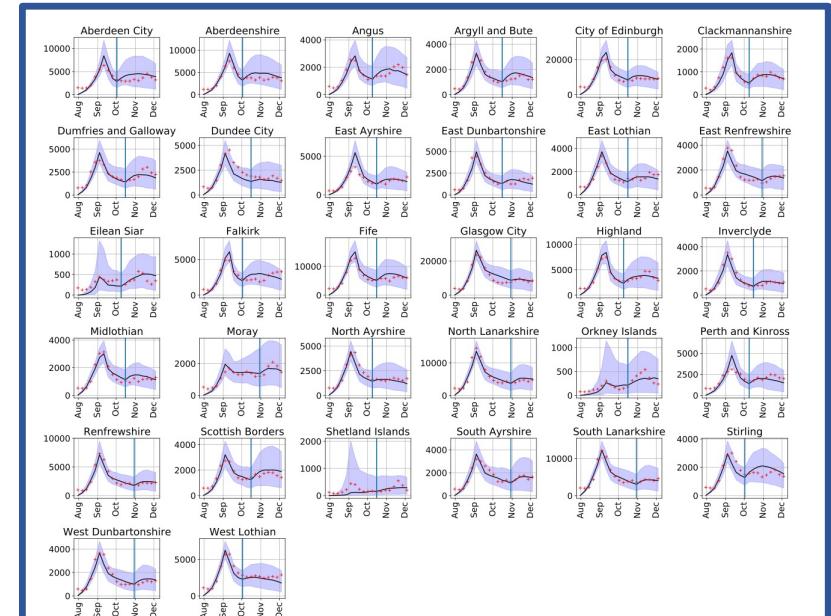
Wellcome Open Research

<https://doi.org/10.21956%2Fwellcomeopenres.19604.r50789>

Banks et al. 2022

arXiv preprint

<https://doi.org/10.48550/arXiv.2211.13704>



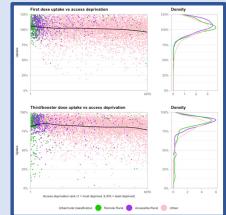
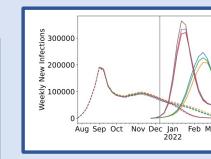
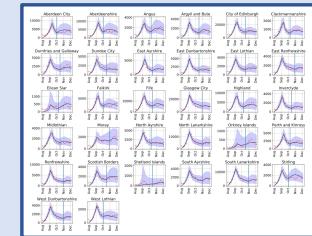
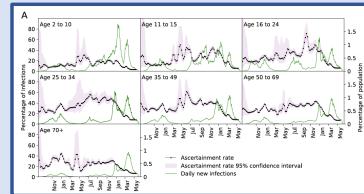
Kao group

- Early access to detailed data
- Experience in disease modelling

SCoVMod: Scotland's Coronavirus Model

Individual-level model of disease spread across Scotland

Forecasts under different restrictions

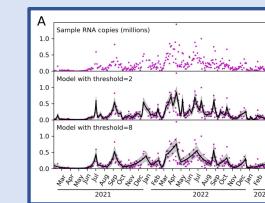
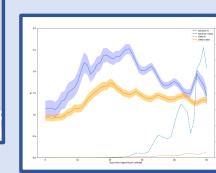
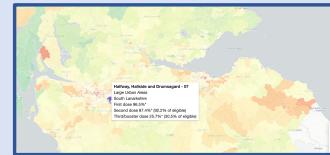


Modelling case ascertainment

Estimating true number of infections as opposed to cases

R-value estimates

Using existing case numbers to estimate past + present R-value locally



Real-time analysis of COVID-19 outcomes

Emphasis on variation with socioeconomic deprivation

Real-time monitoring of vaccine uptake

Highlighting regions of “low uptake” (hesitancy) and “slow uptake” (logistics)

Models of wastewater shedding

Prevalence vs shedding concentration

Optimal monitoring strategies with limited budget

Predictive models of future vaccine uptake

Predicting variation in uptake “fall-off” with future boosters

Weekly
Daily
Ad-hoc

SPI-M / SAGE

Public Health Scotland

Scottish Government

Coronavirus lockdown 'avoided thousands of deaths' in Scotland

© 30 June 2020



By David Cowan

BBC Scotland

A study has suggested between 7,000 and 40,000 people could have died from coronavirus in Scotland if the country had not gone into lockdown.

Scientists have tried to estimate what would have happened if Scotland had copied the approach taken in Sweden, where a lockdown was not imposed.

Rowland Kao, a professor of epidemiology and data science at Edinburgh University, led the team which carried out the research for BBC Scotland.

He said: "An obvious question to ask is if Scotland had done something similar to Sweden, would we have had a similar outcome without all the restrictions."

Although there was no lockdown, Sweden relied on voluntary social distancing, banning gatherings of more than 50 people and halting visits to elderly care homes.

"We took the relative amount of transmission going on in Sweden, translated that over to Scotland and looked at what the resultant number of deaths would have been had we taken that approach."



About Topics News Publications Statistics and research Consultations Blogs

Home > Publications

Publication - Research and analysis

Coronavirus (COVID-19): modelling the epidemic (Issue No.107)

Published: 1 September 2022
From: Director-General Health and Social Care
Directorate: Constitution Directorate
Part of: Coronavirus in Scotland
ISBN: 9781804359105

Latest findings in modelling the COVID-19 epidemic in Scotland, both in terms of the spread of the disease through the population (epidemiological modelling) and of the demands it will place on the system, for example in terms of health care requirement.



This document is part of a collection

one Disclosure

Home Episodes Clips Information and Support



Sorry, this episode is not currently available

Scotland's Lockdown

Series 2

29 minutes

Investigation into government handling of the Covid 19 pandemic. Reporter Mark Daly investigates the decisions made and asks whether Scotland could or should have locked down sooner. Disclosure speaks to scientists desperate to keep us safe and meets the families hardest hit by the virus.

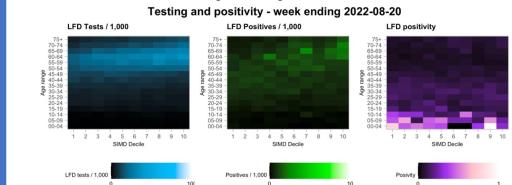
Summary of spatial analysis of Covid-19 spread in Scotland

Researchers at the Edinburgh Roslin Institute have conducted spatial analysis of COVID-19 Spread in Scotland. A summary of findings from data up to 25th August 2022 is included here.

Rates of LFD and PCR testing are stabilising, albeit at very low levels (1-2 LFD tests reported, per week, per 1,000).

The distribution of lateral flow/LFD tests being reported varies substantially by both age and deprivation status, with many fewer tests reported in younger adults across all deciles of deprivation, and for children in more deprived deciles. The high level of LFD positivity in the latter category is marked, and when compared to the high number of positives amongst the least deprived, suggests that ascertainment may be lower in younger people in deprived areas.

Figure 11: Variation in testing outcomes comparing Lateral Flow and PCR testing considering age and deprivation status of the data zone of record based on data in week ending 20th August 2022



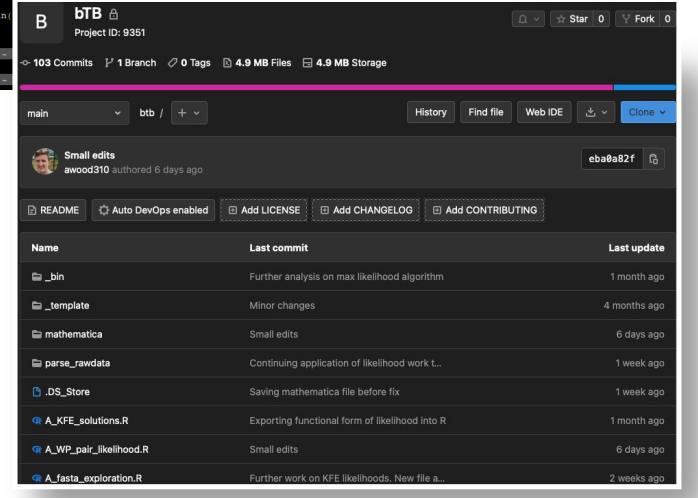
My experience

Notable development in my programming

- from smaller, self-contained code to perform simpler tasks
to
- “large”, complex scripts that
 - Are central to work that ends up in peer-review
 - Require extreme care with storage of data
 - Have spin-offs with short turnarounds
- Essential that the work is reproducible
 - My process for COVID was good *enough*, but very manual and caused some headaches later on

Resolutions

- At the start of a new project, take time to set up processes for **version control, sharing, standardising of code structure**
- Essential for my next project: The *aim* is to develop methods to be used by others!
- 2-3 days spent at the start of a project learning some basic git commands saves weeks later on!



A screenshot of a GitHub repository titled "btB". The repository has a Project ID of 9351, 103 commits, 1 branch, 0 tags, 4.9 MB files, and 4.9 MB storage. The main branch is selected. The commit history shows several recent edits by user "awood310" made 6 days ago, including "Small edits" and "eba0a82f". The repository includes files like README, .DS_Store, A_KFE_solutions.R, A_WP_pair_likelihood.R, and A.fasta_exploration.R. The repository is set up with Auto DevOps enabled and has options to Add LICENSE, CHANGELOG, and CONTRIBUTING.

```
A_stochastic_solutions.R
1 # -----
2 # Script
3 # - A stochastic solutions.R
4 #
5 # Purposes
6 # - Initial exploration of master equations and trajectories for dual processes for epidemiology and substitution.
7 # - This is the stochastic element - the analytic solution was in A_KFE_solutions.R. 2023-03-16: This no longer replicates the deterministic solution.
8 # - and a new piece of code would need writing to verify A_KFE_solutions.R output.
9 #
10 #
11 # - 2023-01-13: Addition of transmission trees
12 # - 2023-01-17: Removal of substitution (can be added later to the trees as an independent process), addition of heterogeneous infection rates
13 # - 2023-03-16: This no longer replicating the KFE solutions
14 #
15 # Authors
16 # - Anthony Wood, Roslin Institute (Kao group), University of Edinburgh
17 #
18 # Email
19 # - anthony.wood@ed.ac.uk
20 #
21 # Date created
22 # - 2022-10-31
23 #
24 #
25 # -----
26 # (0) Establish working directory, required packages, source files
27 # -----
28 #
29 #
30 # Working directory
31 setwd("~/Users/awood310/Desktop/bTB")
32 #
33 # Required packages, and why I'm using them
34 library(tidyverse) # Tidyverse
35 options(dplyr.summarise.inform = FALSE) # Avoid message when summarising by multiple groups
36 library(RColorBrewer) # Colour palettes
37 library(gtools) # For permutations with replacement
38 library(zoo) # For rollmean
39 library(crayon) # Print output text in different colours
40 library(igraph) # Network building
41 library(GGally) # Plot networks
42 library(gridExtra) # grid.arrange function
43 #
44 # Clear all vars
45 rm(list = ls())
46 #
47 # Source files (if required)
48 source("code/f_state_details.R") # Function for pulling the details of all possible microstates - used in both the analytic and numerical code
49 #
50 # Full outer join function
51 full_outer_join = function(x,y, nameX = "x", nameY = "y"){tibble(x) %>% full_join(tibble(y), by = c(nameX, nameY)) %>% select(-nameX, -nameY)}
52 #
53 #
54 # -----
55 # (1) Defining parameters, and the state
56 # -----
```