

# Inference tools for infectious diseases

Authors Chris Pooley<sup>1,3</sup>, Andrea Doeschl-Wilson<sup>2</sup>, and Glenn Marion<sup>1,3</sup>

<sup>1</sup> Biomathematics and Statistics Scotland, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK

<sup>2</sup> The Roslin Institute, The University of Edinburgh, Midlothian, EH25 9RG, UK.

<sup>3</sup> Scottish COVID-19 Response Consortium, [www.gla.ac.uk/scrc/](http://www.gla.ac.uk/scrc/)



[epicscotland.org](http://epicscotland.org)

## INTRODUCTION

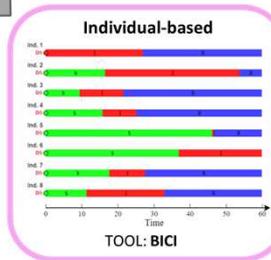
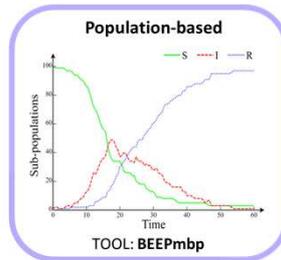
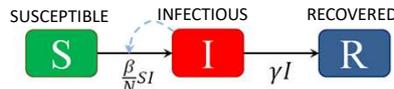
### Why do mathematical modelling?

- To understand dynamical systems.
- To look into the future.
- To test possible control strategies before applying them in reality.

**Mathematical models are only useful if they are well parameterised.** The process of obtaining parameter values from data is known as "inference".

**Bayesian inference provides the best evidence-based advice to inform policy** because it:

- Accounts for uncertainties in data and any prior knowledge regarding plausible parameter values.
- Outputs estimates for model parameters and system behaviour INCLUDING uncertainties.



**Compartmental models divide the population into subgroups depending on their disease status.**

Two contrasting approaches can be taken to inference depending on the type of data available and the system being studied: Population-based methods and individual-based methods.

We have been developing inference tools for each of these possibilities, as discussed below.

## POPULATION-BASED MODELS

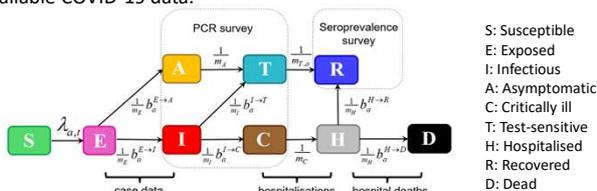


Underlying (for the most part unobserved) system dynamics are represented by variation in the compartmental populations as a function of time.

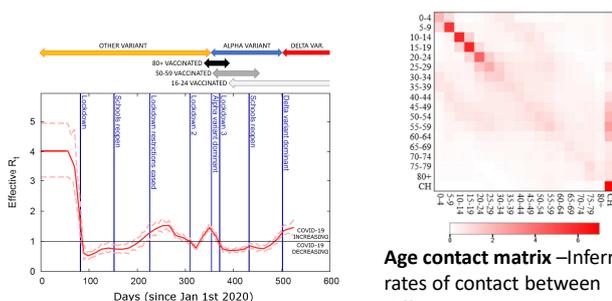
- ✓ Appropriate for large populations or when only population-level data exists.
- ✗ Cannot capture individual-based variability. The times individuals spend in compartments must be exponentially distributed. Information is lost in the case of individual-based data.

**Software tool** – A new inference tool has been developed called "Bayesian estimation of epidemiological parameters using model-based proposals" (BEEPmbp). One way to achieve inference would be to try out lots of different parameter sets, and for each set simulate from the model to see how well the resulting dynamics fit the observed data. In practice this approach tends to be very computationally inefficient, so many alternative types of algorithm have been developed to make inference faster. We have introduced a new method called ABC-MBP which is found to work faster than other methods in a variety of scenarios.

**Application** – Here we show results of applying BEEPmbp to publicly available COVID-19 data.



The model captures different levels of disease severity. Inference is informed by data from cases, hospitalisations, deaths and PCR and antibody test results from randomly selected individuals in the population.



**Reproduction number** - Shows the inferred time variation in  $R_t$  since the beginning of the pandemic

**Age contact matrix** – Inferred rates of contact between different age groups. Relatively more mixing is found in younger age groups compared to pre-pandemic studies.

## INDIVIDUAL-BASED MODELS



Here the underlying dynamics are represented by timelines for each of the individuals in the system.

- ✓ Used when either individual-level data exists or individual variability needs to be accounted for. The probability distributions for the times individuals spend in the different compartments can be flexibly specified.
- ✗ Becomes computationally slow for large populations.

**Software tool** – A software application called "Bayesian individual-based compartmental inference" (BICI) has been developed.



BICI features an easy-to-use point and click interface that has been designed for flexible specification of compartmental models (e.g. allowing for demographic/spatial stratification). Outputs include posterior distributions for model parameters as well as estimates for disease progression along individual timelines and overall population dynamics.

**Application** – So far BICI has been used to analyse individual-level data from disease transmission experiments. These experiments have been aimed at understanding the effect of vaccination and the potential benefits of disease control through selective breeding (e.g. to reduce the overall levels of susceptibility to disease in the population as well as diminish the propensity for infected individuals to pass on their infection to others).

Many animal diseases, such as avian influenza (AI) and foot and mouth disease (FMD), can be modelled using an individual-based model, where farms take the place of "individuals" in the system (e.g. classified into disease-free/incubating/infectious/under-controls). Adaptation of BICI to tackle these problems is underway.

## CONCLUSIONS

In the era of increasingly complex models being applied to larger and larger datasets, development of computationally efficient inference algorithms is vital for scientists to inform policy makers in disease outbreak scenarios.

This poster has introduced two contrasting inference tools which are currently being developed: BEEPmbp for population-based models (e.g. appropriate for COVID-19) and BICI for individual-based models (e.g. used on farm-level data in FMD/AI).

**FUNDING** - CMP and GM were funded by Scottish Government under the RESAS (Rural and Environment Science and Analytical Services Division) Strategic Research Programme and EPIC (Epidemiology, Population health and Infectious disease Control). ADW was funded by the BBSRC Institute Strategic Programme Grants (BBS/E/D/20002172-4 (ISP2)).