**EPIC COVID19 SIMPLE MODEL OVERVIEW**

**Overview**

* Age-structured, with HCW separated population, stochastic epidemiological model
* Built for inferring parameters through an ABC-smc inference framework and predict spread within a population of a given demographic structure

**Current functionality**

Spatial structure

* No spatial structure

Demography

* 7 age groups: Under 20, 20-29,03-39, 40-49, 50-59, 60-69, 70 over
* 1 extra group to include health care worker (HCW), assumed to have similar behaviour as people between 20 and 59 (i.e. mean behaviour).

Movement

* No movement
* Contact structured informed thought contact matrices giving mean number of contacts per day between age groups.

Epidemiology

* Stochastic, compartmental SEI2HRD model (**Figure 1**).
* Flow description (see below):
  + Exposed individuals (Ea) will become infectious (Ia), then symptomatic (Isa) before seeking hospital care (Ha).
  + A portion of the asymptomatic individuals will recover (Ra) before being symptomatic
  + Recovery and hospitalisation of symptomatic individuals is based on age
  + Symptomatic individuals who are not hospitalized will recover (Ra) or die (Da) based on age.
  + Time to death in community will be twice as fast as in hospital.
  + Hospitalized individuals will either recover or die based on age
* Transmission:
  + Transmission rate between age groups as follow:

Where details and definition of parameters are given in **Table 1**.

* Control activities:
  + Control activities implemented through change in contract structure (**Figure 2**)
  + If symptomatic, infectious individuals will self-isolate
  + Self-isolated individuals will only contact a proportion q of the contacts they do at home.
  + During lock-down, a proportion *d* of individuals will restrict their contact to within home, the rest will only not make their work and school contacts.
  + Testing activities implemented by not active yet (extra epi states)
* Model computation:
  + Tau-leap (Poisson process)
  + Daily time step
* Key assumptions:
  + Time to death in community will be twice as fast as in hospital.
  + Symptomatic Infectious individuals will self-isolate.
  + Self-isolated individuals will only contact a proportion q of the contacts they do at home.
  + HCW will self-isolate perfectly once symptomatic.
  + Number of HCW is proportional to population in each HB.
  + Scots contact behaviour similar to the rest of the UK.



**Figure 1.** Epidemiological flow chart for COVID-19

* Parameters
  + Most of parameters extracted from literature and fixed
  + Inference on few, context-specific parameters involved in transmission process
  + See **Table 1** for details.
  + **Figure 3** shows some age-structured information.

**Table 1.** Description of parameters used in the model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **symbol** | **Definition** | **value** | **priors** | **ref.** |
| ps(a) | age-dependent probability of developing symptoms | 0.96 if a>0 |  | http://gabgoh.github.io/COVID/index.html |
|  |  | 0.1 if a=0 |  | from doi:10.1542/peds.2020-0702 |
| ph(a) | age-dependent probability of hospitalisation if case |  |  | Ecdc (2020) |
| p(d|h, a) | age-dependent probability of death given hospitalisation, p(d|h, a) = cfr(a)/ ph(a) |  |  |  |
| p(d|1-h, a) | age-dependent probability of death if not hospitalised, p(d|1-h, a) = cfr(a)/ (1 - ph(a)) |  |  |  |
| cfr(a) | Age-dependent, case fatality ratio |  |  | Russel el al (2020) |
| Tlat | mean latent period | 5 days |  | http://gabgoh.github.io/COVID/index.html |
| Tinf | mean asymptomatic period | 3 days |  | http://gabgoh.github.io/COVID/index.html |
| Tsym | mean symptomatic period prior hospitalization | 5 days |  | http://gabgoh.github.io/COVID/index.html |
| Thos | mean hospitalisation stay | 5 days |  | <https://www.icnarc.org/Our-Audit/Audits/Cmp/Reports> |
| Trec | mean time to recovery if symptomatic infection | 11 days |  | http://gabgoh.github.io/COVID/index.html |
| 𝚲(a,t,d,q) | age-dependent infection rate, function of time and efficacy/compliance of control activities |  |  |  |
| pi | probability of infection given contact when a≠hcw |  | Beta(3,9) |  |
| phcw | probability of infection given contact when a=hcw |  | Beta(3,3) |  |
| c(a,j|t,d,q) | average number of contacts between age groups a and j, given control strategy and their efficacy/compliance when a≠hcw |  |  | Prems et al 2017 |
| chcw | mean number of HCW-patient contacts per day |  | Poisson (42) | doi/10.1098/rsif.2012.0134 |
| q | Factor modulating number of contacts of Is, proxy of quarantine efficacy/compliance |  | Beta(3,3) |  |
| d | Factor modulating number of contacts of I, proxy of efficacy/compliance in social/physical distancing |  | Beta(3,3) |  |



**Figure 2**. Contact matrices informing on the average number of contacts between age groups for (a) normal (pre-COVID) activities, (b) for all activities NOT related to school not work, and (c) household contacts, in the UK. (source: Prem et al 2017) .



**Figure 3.** Age-structured information. (a) Population structure (%N, source: Scottish census 2011), (b) age-specific hospitalisation probability (ph, source: ECDC 2020), and (c) case fatality ratio (CFR, source: Russell et al 2020).

* Model inference
  + Model is integrated within an ABC-smc inference framework.
  + Inference based on the daily reported incident cases (**Figure 4**).
  + Eligible particles were selected based on the normalised sum of squares of residuals. Specifically, we measured the deviation D of each simulated trajectories constructed by each particle such as defined by

where and denote the simulated and observed number of hospitalised individuals (aggregated over all age groups) due to COVID-19 on day t, respectively, with the observation period for the herd running from day t0 to day tn.

* + Currently 5 parameters fitted: pi, phcw, chcw, q and d
  + Priors:
    - pi: Beta distribution with mean of 0.25
    - phcw, q and d: Beta distribution with mean of 0.5 (Uninformative)
    - chcw: Poisson distribution with mean of 42 contacts (see Table 1)
  + Key assumption for inference
    - All reported cases are hospitalised (not yet testing)
    - All regions of interest (e.g. Health Boards) are independent to each other (closed syst.)
    - Infection seeded with 1 infectious individual 8 days prior detection of index case.



**Figure 4.** Model fit. (a) Posterior distribution of 4 of 5 parameters inferred in the ABC-smc inference framework and (b) goodness of fit over all cases reported in Lothian Health Board until April,13 2020. Solid black dots and line in (b) are the daily reported incident cases. Dark and light blue shaded areas represent the 80% and 50% simulation envelops of the simulations accepted during the last selection step of the ABC-smc framework. Dark blue line represents the median trajectory of the simulations accepted during the last selection step of the ABC-smc framework

Parallelisation

* The inference framework can be parallelised. It actually works using OpenMP on Windows environment but not working on macOS.
* Although optimisations on the basis of streamlining memory allocation was made as much as possible (given limited programming skills), it will most likely be improved with better programming.
* Currently working on laptop fairly fast, we inference for each Health Board within few hours

**Modifications necessary**

Epidemiology

* Transmission procedures
  + Improving infection process for HCW
    - to account infection of HCW from older age group in nursing home
    - To account transmission from HCW to population
  + Integration of the time-dependant activity data (google-like activity data) to modulate transmission rate.
* Flow control
  + Introduce an Erland process for recovery of symptomatic individuals (increase infection potential)
  + Implementation of the health status (frailty) information to modulate triage at hospital and improve inferences from death records.
* Mitigation activities
  + Implementation of the testing component
  + Implementation of the progressive lift of restrictions (e.g. time-dependent modulator)

Inference

* Need fitting on both cases and deaths
  + Issues with model structure intrinsically underpredicting
  + Need fitting more parameters (more computationally challenging)
* Use age-structured data on cases and deaths to inferred the model.
* Issues of summary statistics for 2 or more observations.

Programming

* Verification of the code + speed up + bug fixing + design (structure vs classes)
* Currently, data manipulation in R, inference in C++, projection + visualisation in R
  + Might need Rcpp procedure to avoid multiplatform procedure
  + Benefit: integration of the model in Rshiny -> improve communication, + less prone to errors

**Modifications appreciated but not urgent**

Spatial structure

* model will need to be modified to account for the household structure of the community.

**GENERAL EXPERTISE REQUIREMENT**

**Specific technical skills**

Necessary:

* C++ programmers.
* Parallel programming – especially OpenMP, but anyone with C++-specific parallel experience would also be ideal.
* Programmers with Cross platform development (Windows, macOS)
* Experience in model inference, mainly ABC-like framework or other more likelihood based (MCMC) or history matching.

Potential:

* (GP)GPU experience for speeding up optimisation / inference.
* Data processing and visualisation – C++ specific GUI or Rcpp to integrate with web-based dashboard.

**Domain expertise**

* Epidemiology – covered by current consortium set up, but knowledge of COVID-specific epidemiology may need to be sought from CVR, literature searches etc.
* Demographics & movement – covered by current consortium set up, but knowledge on how to integrate new dataset within simple modelling framework.
* …

**Other skills**

* Project / team management
* Code management
* GitHub automation and testing
* Web development to present results – particularly important for transparency.