Age-corrected Bayesian model of the Belgian COVID-19 epidemic

Koen Deforche

koen.deforche@gmail.com

September 17, 2020

1 Introduction

To obtain insight in the current epidemic, models are fitted to the incidence data of deaths, hospitalizations or cases. These models typically assume a infection fatality rate (IFR), hospitalization rate (HR) or case detection rate (CR) to infer from the observed incidence data the evolution of infections over time.

For COVID-19, such methods will result in wrong results since the disease severity is markedly different in younger populations versus older populations. The most studied disease severity characteristic, the IFR, has been found to increase exponentially with age. Therefore, if the infection shifts over time to different age groups, the inference of infections from observed deaths cannot be done using a constant IFR. Similar arguments can be made for hospitalization and case data. For case data, the problem with COVID-19 relates to the amount of asymptomatic or mildly symptomatic cases that go undetected.

A common solution is to use age-structured models, which model the transmission dynamics within and between age groups. These models estimate how infections may happen at different rates in different age groups, and thus also allow to relate observed deaths for each age group to infections estimated in that age group. The amount of transmission parameters, as well as the simulation time of these models, scales however with n^2, for n age groups, and estimation of these parameters from data therefore becomes intractable.

Here, we solve this issue by modeling the IFR, HR, and CR as a time-dependent function. The time-dependent IFR is estimated from the observed age-composition of deaths over time. Similarly, HR and CR are assumed to have a similar time dependency as IFR, but with a shift to offset the differences in infection-to-test, infection-to-hospitalization and infection-to-death intervals.

We model the evolution of the Belgian epidemic using an age-structured model with two age-groups: the younger group (< 65y), and the older group ($\geq 65y$). To avoid biases by changing hospitalization policies, or case detection policy/capacity, the model is estimated mainly from incidence data of deaths. To be able to capture the more recent evolution, hospitalization data, and to a much lesser extent, case data is also used. The time evolution of transmission rates (within each age group, and from the younger to the older group) is modeled using a piece-wise linear model. Some of the inflection points are fixed (such as the first lockdown in March, the second strengthening end of July, and the start of the schools on September 1), while the timing of other inflection points is co-estimated.

We present the current state of the epidemic, as well as a scenario assuming no change in behavior/policy.

2 Results

2.1 Current situation

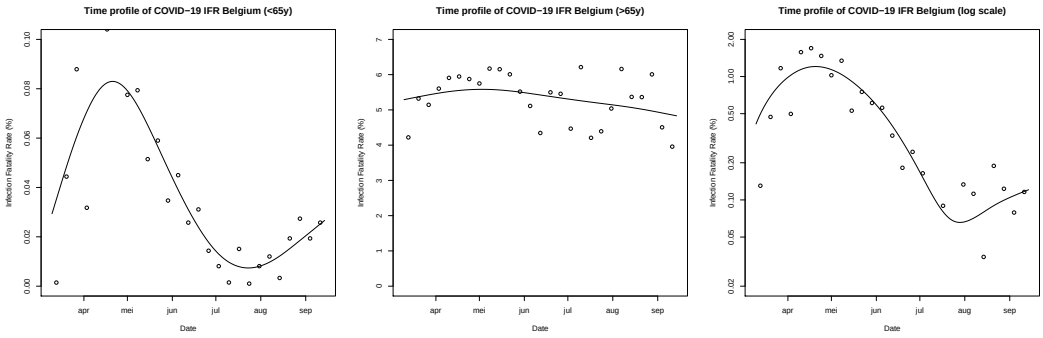

Figure 1: Estimated time profile of Infection Fatality Rate (IFR) in Belgium because of demographic shifts of age composition of infections. Left panel shows $\text{IFR}_y(t)$; middle panel shows $\text{IFR}_o(t)$; right panel shows the overall $\text{IFR}(t)$.

The estimated evolution of IFR over time is shown in Figure 1. The overall evolution of IFR shows that during during the first wave, around 1% of infected individuals died, underlining how deadly COVID-19 is when uncontrolled. This is consistent with the estimated IFR for the Belgian demography as estimated by Moolenbergs et. al from seroprevalence studies taken in April/May. These seroprevalence studies also indicated similar amounts of infection across all age groups. From May on, however, the overall IFR dropped to below 0.1% around the beginning of July, after which a slow increase can be observed. This recent increase is more evident from the estimated IFR within the younger group. On average, thus, people now infected are 10 times less likely to die from the infection than people who were infected in April, and this only based on demographic differences. Effects of improved treatment or lower viral load transmission because of the masks, could make the difference even larger.

The estimated model epicurves are shown in Figure 2. The model fits well the evolution of deaths and hospitalizations over time. Case data only matches the model around August. The predicted case data curve reflects that if we would have been finding cases during the first wave with the same rigor as is applied now, a much larger number of cases would have been detected. Infected people peaked during the first wave at around 1.5% of the Belgian population, and initially all age-groups were infected at a similar rate. As a result of the lockdown, infections in the older age group declined much faster than infections in the young age group. In fact, at no time, infections in the younger age group had a prevalence less than 0.05%, and prevalence started to increase again in June. As a result, at the beginning of August, over 0.5% of the Belgian population was again infected. Today, a prevalence of 0.7% is being estimated. The evolution of the effective reproduction number over time indicates that after the first lockdown, transmission started to increase around the end of

May. The estimated effective reproduction number today, including the effect of reopening of the schools, is 1.3 (1.1 – 1.5).

The evolution of the test detection rate is shown in Figure 3. Although we are now testing much better than we were in March, still, we are now finding only 5% of infections, compared to over 10% during the first wave. This seems a contradiction, but can be understood from the different age-composition of the infections: today most infections are in young people that are in many cases asymptomatic or only mildly symptomatic.

2.2 Prediction assuming no change in behavior

The predicted evolution of the epidemic until the end of the year is shown in Figure 4. This prediction assumes no change in behavior compared to today, and no change in average IFR compared to today. Despite an increase of number of infections to a peak in the second half of October of over 2% of the Belgian population, the peak of hospitalizations is predicted at around 200 per day, and deaths around 30 per day. Because of an increase in immunity, the effective reproduction number Re will drop under 1 in the first half of November.

3 Methods

Figure 5 illustrates the three steps of the method. These steps are discussed individually below.

3.1 Estimating the time profile of IFR

To estimate the time profile of IFR, death incidence data $d_h(t)$ is used, with h each of the 6 age-groups (0-24, 25-44, 45-64, 65-74, 75-84, 85+). For each age-group, data is aggregated on a weekly basis, to obtain $d_h(w)$ with w the year week number. From $d_h(w)$, the estimated amount of infected people is derived $i_h(w)$ as:

$$i_h(w) = \frac{d_h(w)}{\text{IFR}_h}$$

From this, $\text{IFR}_g(w)$, for the two age-groups g used in the compartment model, is derived as:

$$\text{IFR}_g(w) = \frac{\epsilon + \sum_{h \in g} d_h(w)}{\sum_{h \in g} i_h(w)}$$

The addition of ϵ is to overcome a difficulty inherent to estimating the time profile of infections in the younger populations based on very low overall death counts. The exact date of these deaths are less informative for changes in trends. Without the addition of ϵ, the subsequent spline function fitted through the data points would also tend to consider these deaths as rare outliers. The current value of ϵ was chosen experimentally so that the model estimated total accumulative number of infections in each age-group g is close to the straight forward estimate based on total accumulative deaths D_h in each age-group:

$$I_g = \sum_{h \in g} \frac{D_h}{\text{IFR}_h}$$

$\text{IFR}_g(t)$ is then defined as a spline function fitted through $\text{IFR}_g(w)$.

The hospitalization rate HR and case detection rate CR are assumed to have a similar time relation as IFR, but they are shifted in time to account for the offset caused by different infection-to-death μ_d, infection-to-hospitalization μ_h and infection-to-test interval μ_c.

$$\begin{aligned}\mathrm{HR}_g(t) &= h_g \mathrm{IFR}_g(t + \mu_{d,g} - \mu_{h,g}) \\ \mathrm{cR}_g(t) &= c_g \mathrm{IFR}_g(t + \mu_{d,g} - \mu_{c,g})\end{aligned}$$

3.2 An age-structured SEIR model predicting deaths, hospitalizations, and cases

Figure 6 shows the structure of the compartment model, which is a straightforward adaptation of a SEIR model to two age-groups. The equations for this model are:

$$\begin{aligned}\frac{dS_y}{dt} &= -\frac{\beta_y I_y}{N_y} S_y \\ \frac{dE_y}{dt} &= \frac{\beta_y I_y S_y}{N_y} - \sigma E_y \\ \frac{dI_y}{dt} &= \sigma E_y - \gamma I_y \\ \frac{dR_y}{dt} &= \gamma I_y \\ \frac{dS_o}{dt} &= -\left(\frac{\beta_o I_o}{N_o} + \frac{\beta_{yo} I_y}{N_y}\right) S_o \\ \frac{dE_o}{dt} &= \left(\frac{\beta_o I_o}{N_o} + \frac{\beta_{yo} I_y}{N_y}\right) S_o - \sigma E_o \\ \frac{dI_o}{dt} &= \sigma E_o - \gamma I_o \\ \frac{dR_o}{dt} &= \gamma I_o\end{aligned}$$

Constant values $\sigma = 1/4$ and $\gamma = 1/1.4$ are used, resulting in an average generation time of 4.7.

From the time evolution of the SEIR model, the estimated incidence of deaths $d_{g,e}(t)$, hospitalizations $h_{g,e}(t)$ and cases $c_{g,e}(t)$ are derived from $S(t)$ using convolution with a gamma distribution to reflect the distribution of the interval for each of these 3 processes.

$$\begin{aligned}d_{g,e}(t) &= -\mathrm{IFR}_g(t) \frac{dS_g(t)}{dt} * \gamma(\mu_{d,g}, \sigma_{d,g}) \\ h_{g,e}(t) &= -\mathrm{HR}_g(t) \frac{dS_g(t)}{dt} * \gamma(\mu_{h,g}, \sigma_{h,g}) \\ c_{g,e}(t) &= -\mathrm{CR}_g(t) \frac{dS_g(t)}{dt} * \gamma(\mu_{c,g}, \sigma_{c,g})\end{aligned}$$

The transmission rates β_y, β_o, and β_{yo} are allowed to change over times, according to a mechanistic model assuming a piece-wise linear model. At the moment the epidemic is fitted assuming 8 change points for transmission rate, chosen to reflect known and unknown changes in policy and behavior. Table 1 lists how the 8 dates are defined. The transmission rates themselves are all estimated from the data.

3.3 Parameter estimation using MCMC

Table 1 lists all parameters and their values (either a constant, or a prior distribution for parameters that are estimated).

The parameters are estimated using MCMC with Metropolis coupling (20 chains at different temperatures).

Table 1: Parameters of model

Parameter	Description	Value
N_y	Population size, younger group (< 65)	8.55×10^6
N_o	Population size, older group (≥ 65)	2.95×10^6
$1/\sigma$	Latent period	4 days
G	Generation time	4.7 days
$\beta_{y,i}$	Transmission rates ($i = 0..7$) within younger group	estimated
$\beta_{o,i}$	Transmission rates ($i = 0..7$) within older group	estimated
$\beta_{yo,i}$	Transmission rates ($i = 0..7$) from younger to older group	estimated
c_y	Case rate, younger group	estimated
c_o	Case rate, older group	estimated
$\mu_{c,y}$	Average infection-to-case interval, younger group	estimated
$\mu_{c,o}$	Average infection-to-case interval, older group	estimated
σ_c^2	Variance infection-to-case interval	estimated
h_y	Hospitalization rate, younger group	estimated
h_o	Hospitalization rate, older group	estimated
$\mu_{h,y}$	Average infection-to-hospitalization interval, younger group	estimated
$\mu_{h,o}$	Average infection-to-hospitalization interval, older group	estimated
σ_h^2	Variance infection-to-hospitalization interval	estimated
$\mu_{d,y}$	Average infection-to-death interval, younger group	estimated
$\mu_{d,o}$	Average infection-to-death interval, older group	estimated
σ_d^2	Variance infection-to-death interval	estimated
d_0	Date for decrease in transmission prior to lockdown	estimated
d_1	Start of lockdown transition	March 13
d_2	End of lockdown transition	March 20
d_3	Start of increase after lockdown	estimated
d_4	End of increase after lockdown	estimated
d_5	New measures for second wave	July 27
d_6	Effect of measures for second wave	estimated
d_7	Start of school year	September 1
ϵ	Spread-out factor for deaths	0.07 (estimated)

3.4 Code availability

The scripts for this model are available at `https://github.com/kdeforche/epi-mcmc`. The setup for this analysis described in `analyses/be/age/`.

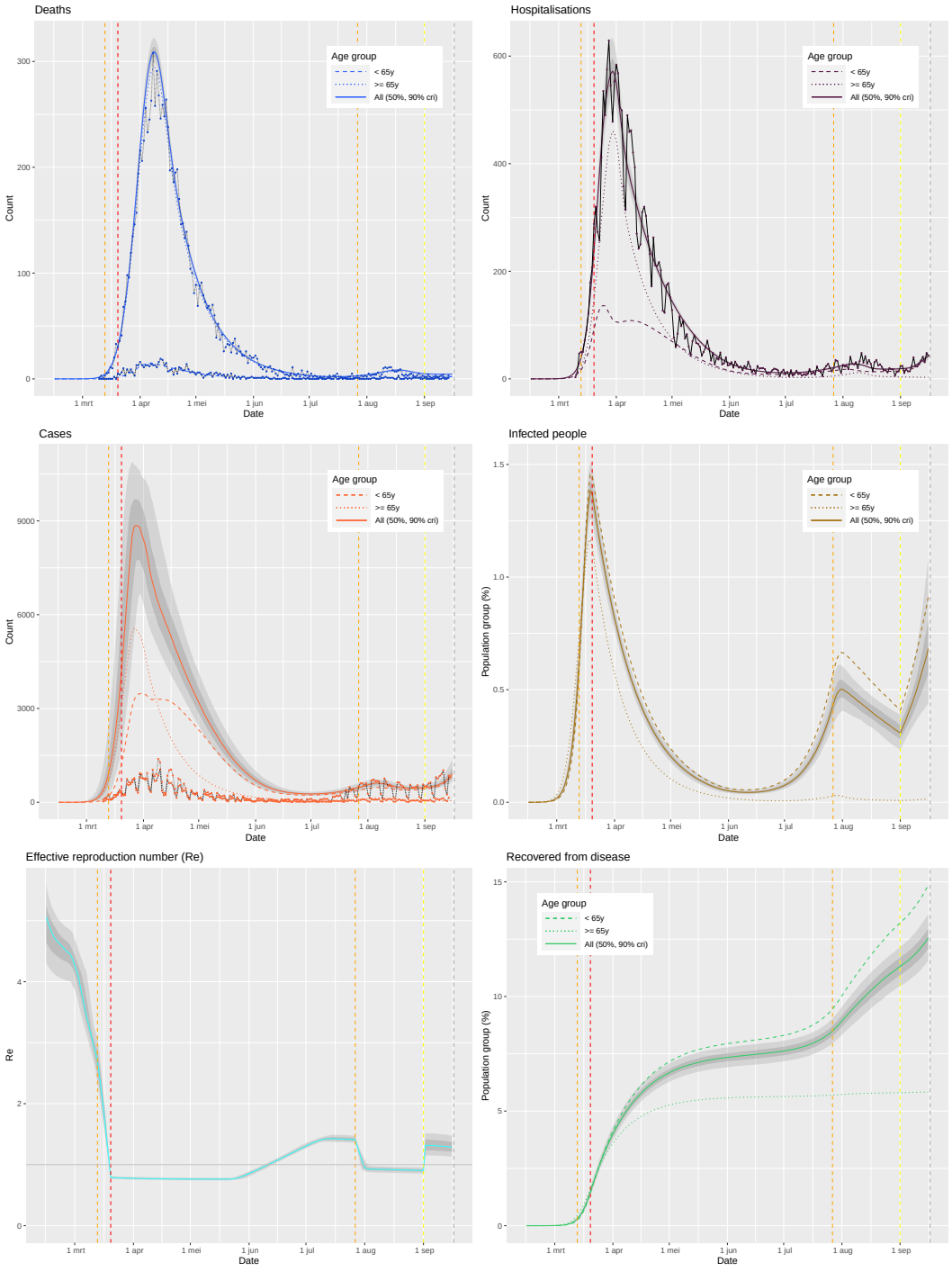

Figure 2: Estimated current state of the epidemic in Belgium, based on an age-corrected model fitted on incidence data of deaths, hospitalizations and cases.

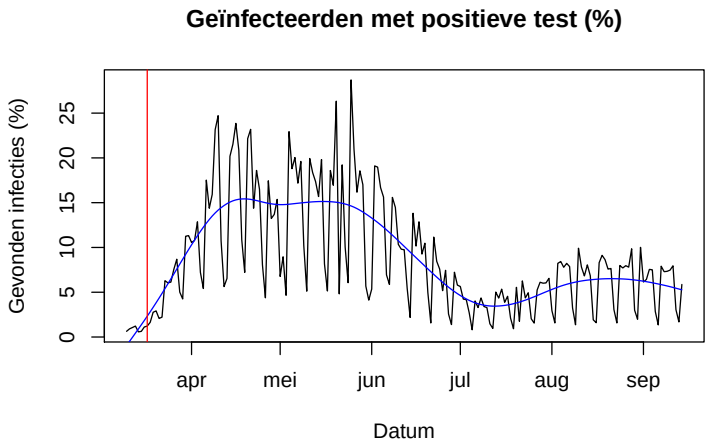

Figure 3: Evolution of estimated infection detection over time, showing the percentage of infections that are diagnosed using a positive test (PCR).

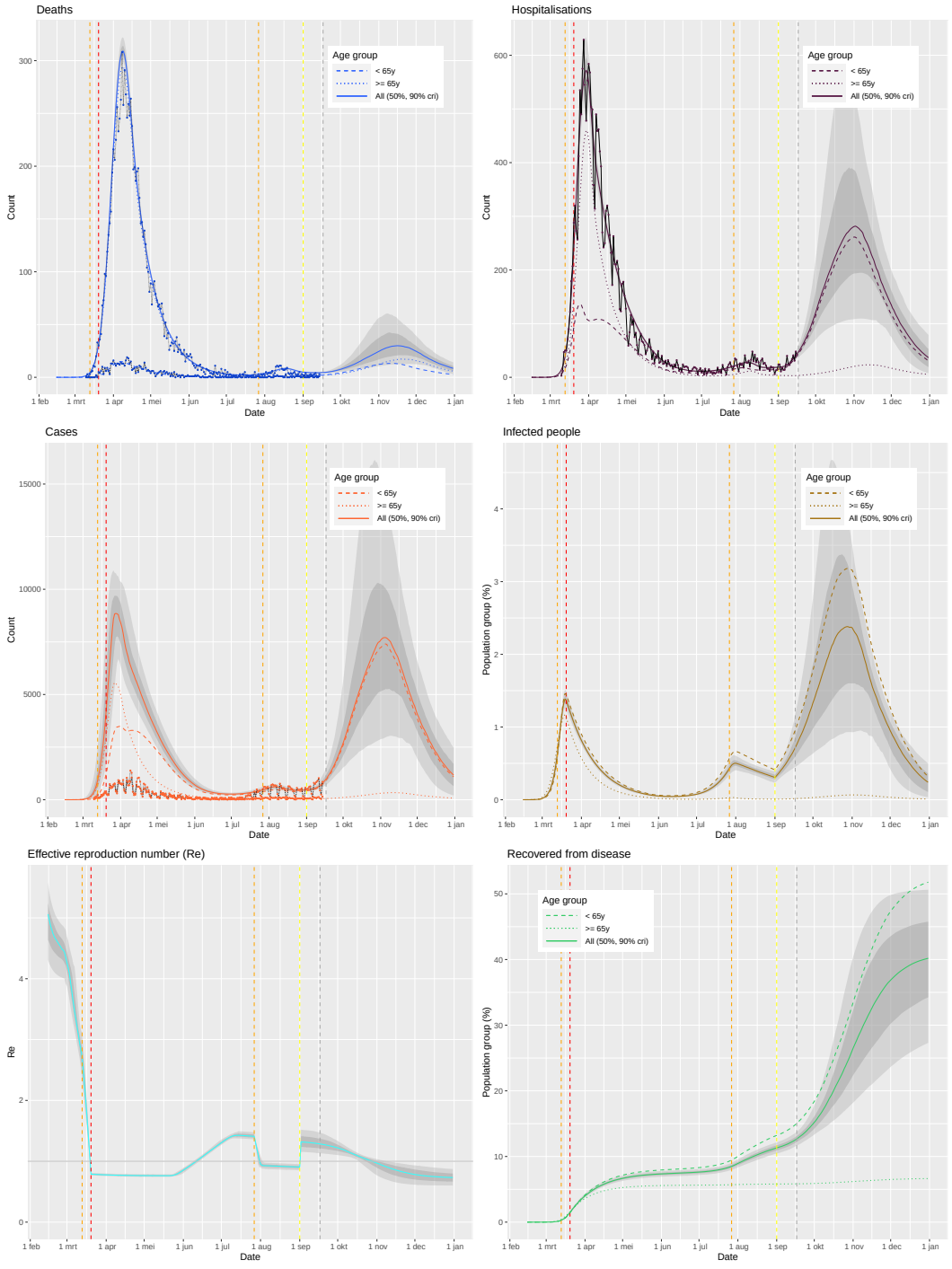

Figure 4: Predicted evolution of the Belgian epidemic, provided the current behavior/measures are continued.

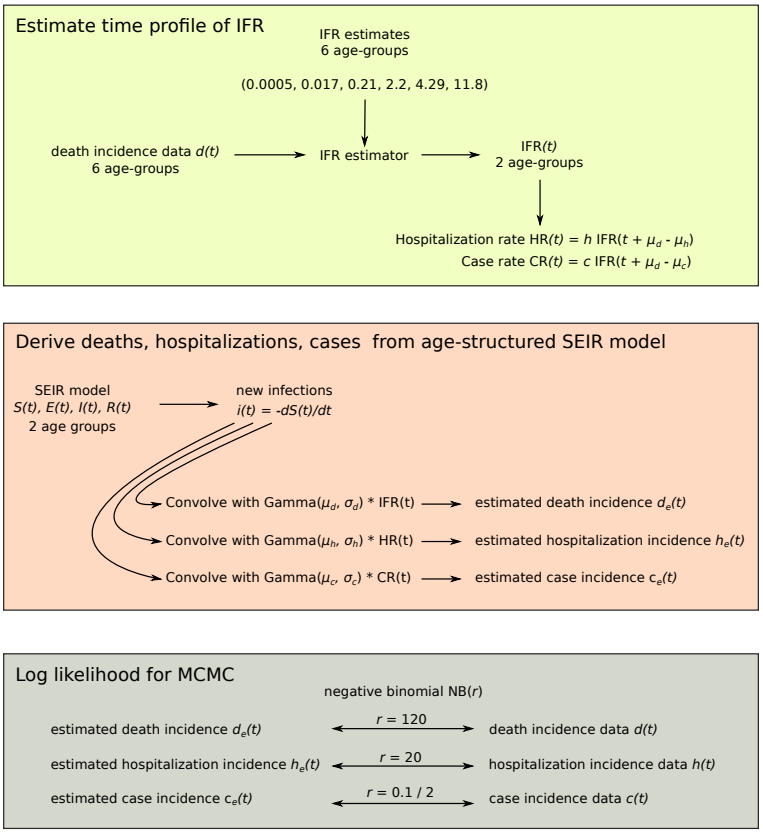

Figure 5: Overview of the building blocks of the model: (1) time profile of IFR is estimated directly from death incidence data, and the time dependence of hospitalizations and cases is assumed similar but shifted in time; (2) death incidence, hospitalization incidence and case incidence is estimated from infections in the SEIR model using a convolution with a Gamma distribution, and multiplication with IFR(t); (3) log likelihood function uses a negative binomial to compare model predictions with observed data.

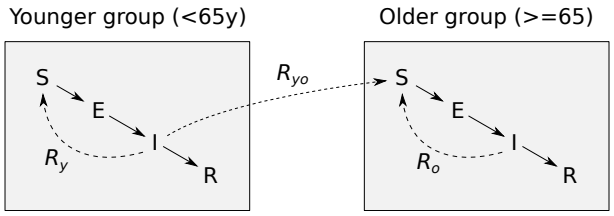

Figure 6: Model structure. Within each age group, individuals are considered to be in one of four states: susceptible (S), infected but latent (E), infectious (I) or recovered/death (R). Within each age group, infectious people can infect susceptible people, modulated by reproduction number R_y for the younger group, and R_o for the older group. The younger infectious people can also infect older people at rate R_{yo}.