

How can we predict the evolution of COVID-19 in Belgium?

Antoine Soetewey

April 24, 2020

Table of contents

- Modeling COVID-19 Infectious cases
 - ▶ *SIR* model
 - ▶ Fitting a *SIR* model
 - ▶ Reproduction number R_0
 - ▶ Predictions
 - ▶ More summary statistics
 - ▶ Additional considerations and improvements

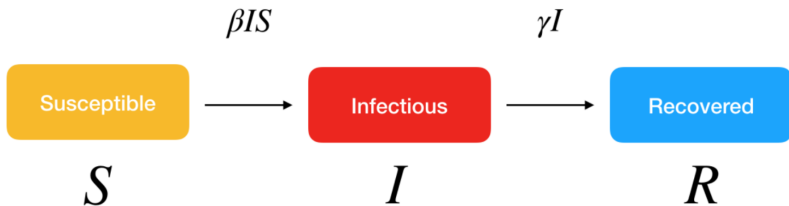
SIR model

- ▶ A classic epidemiological model
- ▶ Applicable to many disease outbreaks
- ▶ 3 groups of individuals:
 1. **S**usceptible: healthy individuals but susceptible to the disease.
At t_0 , S = entire population since no one is immune to the virus
 2. **I**nfectious
 3. **R**ecovered (or removed): contaminated individuals but who have either recovered or died. They are not infectious anymore

SIR model

As the virus progresses in the population:

- ▶ S decreases when individuals are contaminated and move to I
- ▶ As people recover or die, they go from I to R



SIR model

To model the outbreak we need to describe the change in each group, parameterised by:

- ▶ β (infection rate) which controls $S \rightarrow I$
- ▶ γ (removal rate) which controls $I \rightarrow R$

SIR model

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- ▶ Eq. 1: S decreases with newly infected individuals
- ▶ Eq. 2: I increases with newly infected individuals, minus infected people who recovered
- ▶ Eq. 3: R increases with the number of individuals who were infectious and who either recovered or died

SIR model

In R:

```
SIR <- function(time, state, parameters) {  
  par <- as.list(c(state, parameters))  
  with(par, {  
    dS <- -beta * I * S / N  
    dI <- beta * I * S / N - gamma * I  
    dR <- gamma * I  
    list(c(dS, dI, dR))  
  })  
}
```

Fitting a *SIR* model

To fit the model to the data we need to find the optimal values of our parameters that minimise the sum of the squared differences between $I(t)$ and the corresponding number of cases as predicted by our model $\hat{I}(t)$:

$$RSS(\beta, \gamma) = \sum_t (I(t) - \hat{I}(t))^2$$

Fitting a *SIR* model

In R, with `ode()` (for ordinary differential equations) and `optim()`:

```
library(deSolve)
```

```
RSS <- function(parameters) {  
  names(parameters) <- c("beta", "gamma")  
  out <- ode(y = init, times = Day,  
            func = SIR, parms = parameters)  
  fit <- out[, 3]  
  sum((Infected - fit)^2)  
}
```

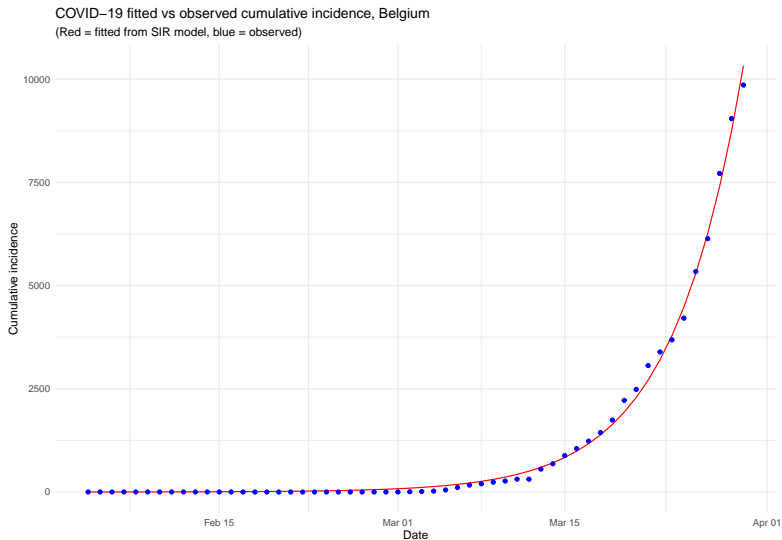
```
Opt <- optim(c(0.5, 0.5), # find the optimal values  
            RSS,          # that give the smallest RSS  
            method = "L-BFGS-B", # start with values of 0.5  
            lower = c(0, 0),      # and constrain them to  
            upper = c(1, 1)      # the interval 0 to 1.0  
)
```

Data

- ▶ Dataset of John Hopkins (collection of 12 resources), via {coronavirus} R package
- ▶ Data from Feb. 4 (1st confirmed case) until March 30 because:
 - ▶ What is needed are currently infected persons (cumulative infected minus the removed, i.e. recovered or dead)
 - ▶ But numbers of recovered persons are hard to obtain and probably **underestimated** (underreporting bias)
 - ▶ We thus consider the cumulative number of infected people until the number of recovered individuals becomes non-negligible
 - ▶ Which I assumed was ± 14 days¹ after lockdown
- ▶ Analyses done here are still valuable to see how the virus would have evolved

¹Average duration after which COVID-19 patients are considered as cured.

Application to Belgium



Reproduction number R_0

- ▶ Model fits well to the observed data, so we can compute the *reproduction number* R_0 as

$$R_0 = \frac{\beta}{\gamma}$$

- ▶ Gives the average **number of healthy people that get infected per number of sick** (infectious) people
- ▶ The larger the R_0 , the harder it is to control the epidemic and the higher the probability of a pandemic

Reproduction number R_0

In R:

```
Opt_par <- setNames(Opt$par, c("beta", "gamma"))  
Opt_par
```

```
##      beta      gamma  
## 0.5841185 0.4158816
```

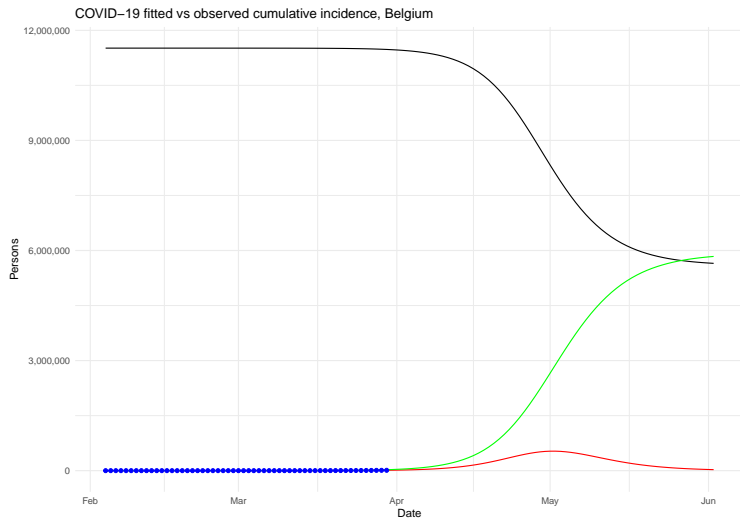
```
R0 <- as.numeric(Opt_par[1] / Opt_par[2])  
R0
```

```
## [1] 1.404531
```

- On average in Belgium, 1.4 persons were contaminated for each infected person for the period considered

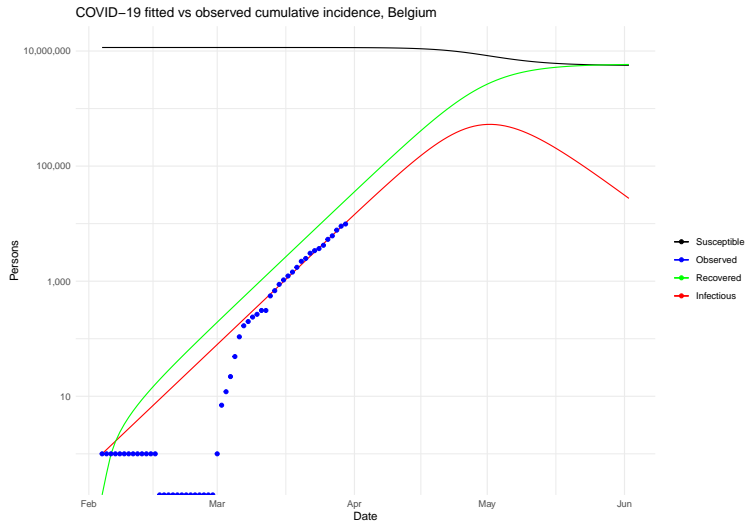
Predictions

► No health intervention and fixed R_0 :



Predictions

► In log scale:



More summary statistics

```
# peak of pandemic
```

```
fit[fit$I == max(fit$I), c("Date", "I")]
```

```
##           Date           I  
## 89 2020-05-02 531000.4
```

```
# severe cases
```

```
max(fit$I) * 0.2
```

```
## [1] 106200.1
```

```
# cases with need for intensive care
```

```
max(fit$I) * 0.06
```

```
## [1] 31860.03
```

```
# deaths with supposed 4.5% fatality rate
```

```
max(fit$I) * 0.045
```

```
## [1] 23895.02
```


Additional considerations

- ▶ Previous figures must be taken with extreme **caution**:
 - ▶ Based on rather unrealistic assumptions:
 - ▶ no public health interventions
 - ▶ fixed reproduction number R_0
 - ▶ Other assumptions (more realistic?) for severe cases, ICU and fatality rates
 - ▶ Data quality
- ▶ **BUT** previous pandemics (e.g., Spanish & swine flu) showed that high number are not impossible...

Improvements

- ▶ **SEIR** model: \approx *SIR* but infected people *I* are divided into:
 1. *E* for *Exposed*/infected but asymptomatic
 2. *I* for *Infected* and symptomatic
- ▶ Modelling the epidemic trajectory using **2 log-linear models**:²
 1. one to the growth phase (before the peak)
 2. one to the decay phase (after the peak)allowing to estimate doubling and halving times
- ▶ Estimate the current **effective reproduction number** R_e on a day-by-day basis³
- ▶ More sophisticated projections⁴

²See {incidence} R package.

³See {EpiEstim} R package.

⁴See {projections} R package.

This talk is based on & complements:

- ▶ LIDAM Report ([link](#))
- ▶ Blog ([link](#))

Thanks!

Questions?