Excess death in relation to Covid-19

16 de May, 2020

Contents

1 Problem

As of May 15th 2020, the global death toll of the unfolding COVID-19 outbreak stands at 302 493 (*citation needed*). The global and national official death toll figures have drawn much attention and sparked vivid debate because they are at the center of the evaluation and comparison of the public health responses of national and local governments.

Among the controversies, there is first a debate on the severity of COVID-19 when compared to the yearly flu outbreak. Brazil's President Jair Bolsonaro refered the COVID-19 as a 'little flu' and refused to implement in his country the drastic lockdown measures that many other countries have enforced (*citation needed*).

Second, the limits of testing in terms of tests availability and accuracy have led many observers to point at a likely underreporting of deaths due to the novel virus. Moreover, deaths that might be indirectly due to the COVID-19 crisis because of, for example, a collapse of the health system, are not counted in official figures. On April 26th 2020, the Financial Times headlined that global coronavirus death toll could be 60% higher than reported (*citation needed*).

Finally, differences in testing and reporting policies across countries but also regions have casted more doubts on the veracity of the reporting figures. Belgium has reported the highest number of deaths per 100 000 inhabitants but Belgian officials also say they are counting in a way that no other country in the world is currently doing: counting deaths in hospitals and care homes, but also including deaths in care homes that are suspected, not confirmed, as COVID-19 cases (*citation needed*).

There is a need of a rigorous estimation of the excess mortality in the weeks of the outbreak. A direct week by week comparison of the observed number of deaths to historical averages as done by the Financial Times

Table 1: Number of deaths by week in the Netherlands

week	X2019	X2018	X2017	X2020	average
2	3262	3359	3637	3364	3637
3	3150	3364	3487	3152	3487
4	3178	3322	3626	3041	3626
5	3143	3403	3574	3158	3574
6	3185	3513	3446	3189	3446
7	3252	3660	3417	3195	3417

analysts is a first approach but is deeply limited as it fails to consider the variance of the number of deaths across years.

In this report, we propose a Bayesian approach to estimate the excess mortality in the outbreak weeks through relative risk. Our model intends to provide parts of the answer to the following questions:

- Is mortality significantly higher than usual in weeks of the outbreak?

- If confirmed, is there significant excess mortality on top of the reported COVID deaths?

2 Data

2.1 Description

Every European state has an established monitoring system of death of nationals, often centralized by the local national statistics institute. Those platforms offer a good quality and reliable source to estimate excess mortality. We gathered weekly data of the total number of deaths for weeks 2 to 17 in the following countries and following years:

- Norway: years 2014 to 2020

- Netherland: years 2017 to 2020

- Belgium: years 2009 to 2020

- Germany: years 2016 to 2020

- Switzerland: years 2015 to 2020

- Italy: years 2015 to 2020

- France: years 2010 to 2020

- England and Wales: years 2010 to 2020

The data have been retrieved from national statistic institutes of the respective countries. For each country the data were transformed to a similar structure than can be visualized below.

2.2 Exploratory data analysis

2.2.1 Visualization

Fig.1 and Fig.3 show the weekly deaths over weeks 1 to 17 for each of the countries and over the years. Most countries show the same pattern with a large peak between weeks 10 and 16 in 2020 which clearly stands out of from other years. Notably, Switzerland show a peak in the same weeks in 2020 but it is not larger than

other peaks observed in previous years. Norway shows no peak, and on a contrary it shows a trough, which might be the number of deaths for an incomplete week.
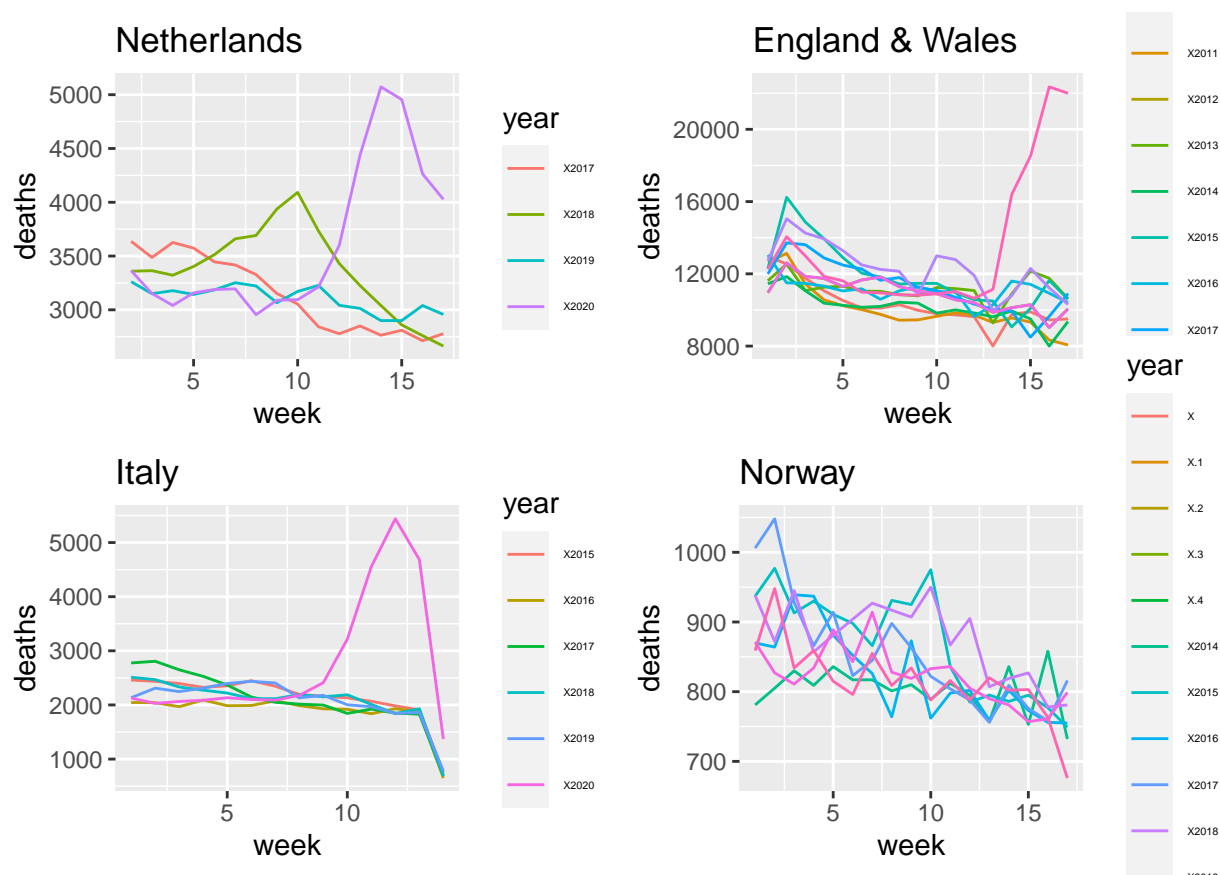

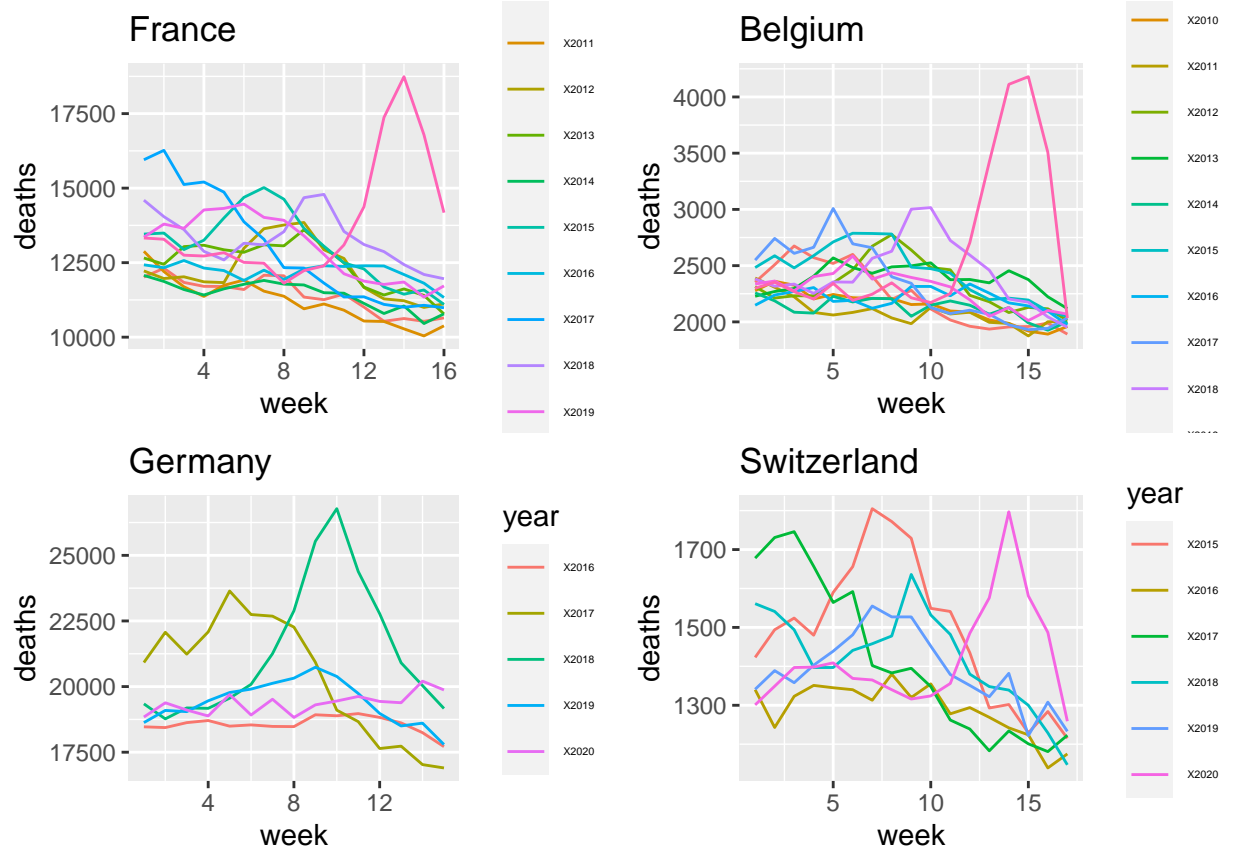Figure 1: Weekly deaths by year and country

Figure 2: Weekly deaths by year and country bis

2.2.2 Overdispersion

Our data is characterized by much larger variance than mean, it is overdispersed. For example, Table 2 shows the weekly means and variance for Belgium.

Table 2: Means and variances of weekly deaths in Belgium

	8	9	10	11	12	13	14	15	16	17
Mean	2216	2206	2182	2156	2096	2087	2078	2108	2069	1996
Variance	488398	490557	496106	479915	431053	431961	531369	718078	742182	526573

3 Model

3.1 Excess mortality

3.1.1 Statistical model

We approach the modelling of excess mortality by estimating the relative risk θ understood as the ratio of weekly mortality during the COVID-19 outbreak and the weekly mortality in non-outbreak times.

$$RelativeRisk_i = \theta_i = \frac{mortality_{COV,i}}{mortality_{noCOV,i}} = \frac{\frac{O_{COV,i}}{N}}{\frac{O_{noCOV,i}}{N}} = \frac{O_{COV,i}}{O_{noCOV,i}}$$

where $O_{COV,i}$ is the observed number of deaths in week i of COVID outbreak, $O_{noCOV,i}$ the number of deaths for the same week in non-outbreak times and N is the population assumed stable accross years.

The number of deaths O is a count variable. For such variable, two commonly used statistical models are the Poisson model (1) and the negative binomial model (2) that we parametrize with location and scale parameters.

$$(1) \quad O_{COV,i} \sim \mathcal{P}(\lambda), \ E(O_{COV,i}) = Var(O_{COV,i}) = \lambda$$

$$(2) \quad O_{COV,i} \sim \mathcal{NB}(\mu, \phi), \ E(O_{COV,i}) = \mu, \ V(O_{COV,i}) = \mu + \frac{\mu^2}{\phi}$$

A major difference between the two is that the Poisson model assumes the mean is equal to the variance while the negative binomial model has one more parameter that allows for overdispersion.

We have:

$$O_{COV,i} = O_{noCOV,i}\theta_i \Rightarrow E(O_{COV,i}) = E(O_{noCOV,i})\theta_i = E_i\theta_i$$

where E_i is the expected number of deaths in week i in non-outbreak times. For each week i, we estimate E_i as the historical average of number of deaths in year prior to 2020.

We therefore define the following statistical models:

$$(1) \quad O_{COV,i} \sim \mathcal{P}(E_i\theta_i)$$

$$(2) \quad O_{COV,i} \sim \mathcal{NB}(E_i\theta_i, \phi)$$

In thoses models, a value of θ_i larger than 1 can be interpreted as excess mortality in week i with respect to non-outbreak times and a value smaller than 1 as reduced mortality. This approach to modelling mortality is handy as the estimated value is standardized to the expected number of deaths and can therefore be compared between weeks and countries.

From the exploratory data analysis, we see that there is a clear time dependence in the weekly mortality. We then introduce a time structured random effect θ_t in our models. We use a autoregressive structure of order 1 as time strcuture. Our models are then as follows:

$$(1) \quad O_{COV,i} \sim \mathcal{P}(E_i\theta_i\theta_{t,i})$$

$$(2) \quad O_{COV,i} \sim \mathcal{NB}(E_i\theta_i\theta_{t,i}, \phi)$$

where:

$$\log(\theta_{t,i}) = \alpha + \beta\log(\theta_{t,i-1})$$

3.1.2 Bayesian model

We propose hierarchichal models for our Bayesian estimation of θ and θ_t.

Poisson model (1):

$$(O_{COV,2}, ..., O_{COV,17} | (\theta_2, \theta_{t,2}), ..., (\theta_{17}, \theta_{t,17})) \sim \prod_{i=2}^{17} \mathcal{P}(E_i \theta_i \theta_{t,i})$$

with parameters priors:

$$(\log(\theta_2), ... \log(\theta_{17})) \sim \mathcal{N}(0, \sigma)$$

$$\forall i, \log(\theta_{t,i}) = \alpha + \beta \log(\theta_{t,i-1}) + \epsilon, \ \epsilon \sim \mathcal{N}(0, \sigma_t)$$

where σ, α, β and σ_t are hyperparameters with hyperpriors:

$$\sigma \sim \mathcal{U}(a_\sigma, b_\sigma)$$
$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha)$$
$$\beta \sim \mathcal{U}(a_\beta, b_\beta)$$
$$\sigma_t \sim \mathcal{N}(\mu_{\sigma_t}, \sigma_{\sigma_t})$$

Similarly the negative binomial model (2) is:

$$(O_{COV,2}, ..., O_{COV,17} | (\theta_2, \theta_{t,2}), .., (\theta_{17}, \theta_{t,17})) \sim \prod_{i=2}^{17} \mathcal{NB}(E_i (\theta_i \theta_{t,i})$$

with parameters priors:

$$(\log(\theta_2), ..., \log(\theta_{17})) \sim \mathcal{N}(0, \sigma)$$

$$\forall i, \log(\theta_{t,i}) = \alpha + \beta \log(\theta_{t,i-1}) + \epsilon, \ \epsilon \sim \mathcal{N}(E_i \theta_i \theta_{t,i}, \phi)$$

$$\phi \sim \mathcal{U}(a_\phi, b_\phi)$$

where σ, α, β and σ_t are hyperparameters with hyperpriors:

$$\sigma \sim \mathcal{U}(a_\sigma, b_\sigma)$$
$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha)$$
$$\beta \sim \mathcal{U}(a_\beta, b_\beta)$$
$$\sigma_t \sim \mathcal{N}(\mu_{\sigma_t}, \sigma_{\sigma_t})$$

3.2 Excess mortality on top of the reported COVID deaths

To model excess mortality on top of reported COVID deaths, we use the same approach as for modelling of excess mortality during the COVID-19 outbreak with respect to non-outbreak times. We previously estimated the relative risk θ understood as the ratio of weekly mortality during the COVID-19 outbreak and the weekly expected mortality in non-outbreak times. We now adjust the exptected mortality E_i to account for COVID deaths.

$$RelativeRisk_i = \theta_i = \frac{O_{COV,i}}{O_{noCOV,i} + D_i}$$

where $O_{COV,i}$ is the observed number of deaths in week i of COVID outbreak, $O_{noCOV,i}$ the number of deaths for the same week in non-outbreak times and D_i is the reported number of deaths due to COVID in week i.

We have:

$$\tilde{E}_i = E_i + D_i$$

4 Results

4.1 Poisson model

We started by estimating the above Poisson model in a simpler form, without the time dependence. The model resulted insuitable despite having tried numerous variations of the hierarchical model: changing the priors and levels of the multilevel model. Indeed, as can be observed in Fig. 3 for Belgium, the model adequately captured the change in relative risk in peak weeks but failed to adequately reflect variance with extremely narrow confidence intervals. We obtained similar results for the Netherlands, England & Wales, Italy and Germany. As a consequence, we abandonned the model.

```
knitr::include_graphics("Poisson_belgium.png")
```
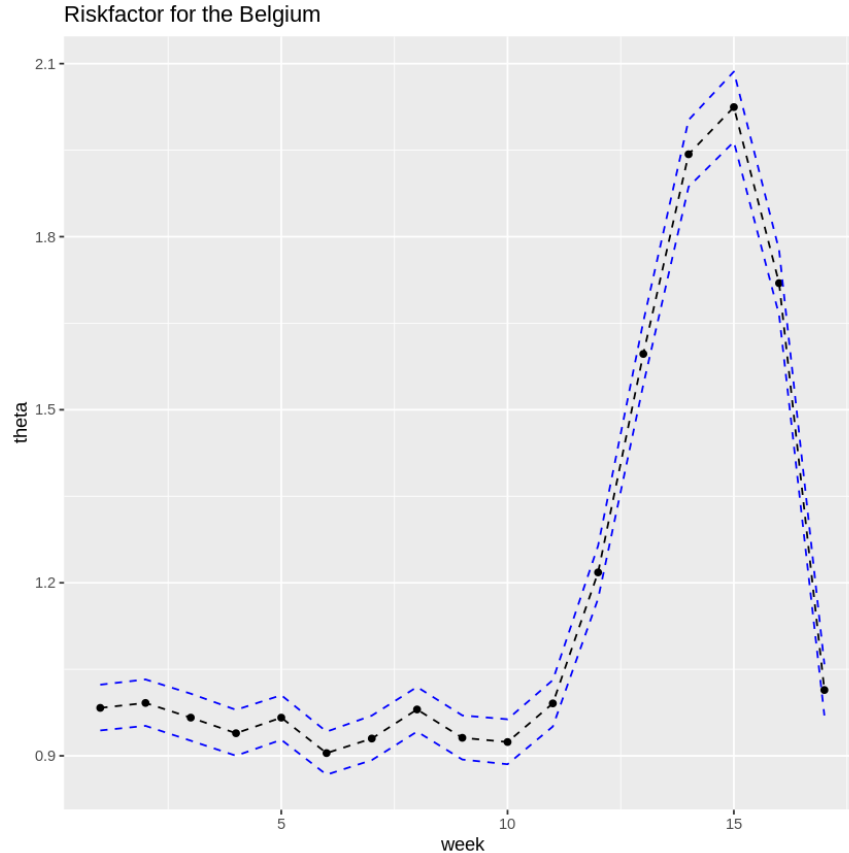
Figure 3: Poisson model estimates for Belgium

4.2 Negative binomial model

The negative binomial model proved to be much more suited to our data. This is not surprising givent the overdispersion we observed in the exploratory data analysis.

Reading the data

```
data_netherland<- read.csv("data/Death_week_2_17_netherland.csv", sep = ";")
data_england_wales <- read.csv("data/England_Wales_weekly.csv",sep = ";")
data_italy <- read.csv("data/italy_week_deaths.csv", sep = ",")
data_norway<- read.csv("data/Norway_total_death.csv", sep = ";")
data_germany <-read.csv("data/germany.csv", sep = ";")
data_switzerland <- read.csv("data/switzerland.csv", sep = ";")
data_france <- read.csv("data/data_france/weekly_deaths_france_week1_16_years2010_2020.csv", header = T
data_belgium <- read.csv("data/data_belgium/weekly_deaths_belgium_week1_17_years2009_2020.csv", header
```

Extracting the data for the Netherlands

```
mean_netherland <- as.vector(as.matrix(data_netherland)[,6])
netherland_2020<-as.vector(as.matrix(data_netherland)[,5])
```

Extracting data England and Wales

```
mean_england <- as.vector(as.matrix(data_england_wales)[,12])
england_2020 <- as.vector(as.matrix(data_england_wales)[,13])
```

8

Extracting data Italy

```r
data_italy <- data_italy[-14,]
mean_italy <- as.vector(as.matrix(data_italy)[,9])
italy_2020 <- as.vector(as.matrix(data_italy)[,8])
```

Extracting the data for Norway

```r
data_norway <- data_norway[-17, ]
mean_norway <- as.vector(as.matrix(data_norway)[,8])
norway_2020<-as.vector(as.matrix(data_norway)[,7])
```

Extracting the data of Germany

```r
mean_germany <- rowMeans(data_germany[, c(2,3,4,5)])
data_germany <- cbind(data_germany, mean_germany)
germany_2020 <- as.vector(as.matrix(data_germany)[,6])
```

Extracting the data of Switzerland

```r
mean_swiss <- rowMeans(data_switzerland[, c(3,4,5,6,7)])
data_switzerland <- cbind(data_switzerland, mean_swiss)
swiss_2020 <- as.vector(as.matrix(data_switzerland)[,2])
```

Extracting the data of France

```r
mean_france <- rowMeans(data_france[, c(2:11)])
data_france <- cbind(data_france, mean_france)
france_2020 <- as.vector(as.matrix(data_france)[,12])
```

Extracting the data of Belgium

```r
mean_belgium <- rowMeans(data_belgium[, c(2:12)])
data_belgium <- cbind(data_belgium, mean_belgium)
belgium_2020 <- as.vector(as.matrix(data_belgium)[,13])
```

Setting up the stan lists

```r
create_list <- function(N, E, O){
  data_list <-  list(
  N = N,
  E = E,
  O = O,
  phi_a = 1,
  phi_b = 6000,
  sigma_a = 0.01,
  sigma_b = 4,
  alpha_mu = 0,
  alpha_sigma = 4,
  beta_a = -1,
  beta_b = 1,
  sigma_time_mu = 0,
  sigma_time_sigma= 25
  )
  return(data_list)
}
```

```r
data_list_netherland <- create_list(length(mean_netherland), mean_netherland, netherland_2020)
data_list_england_wales <- create_list( length(mean_england),mean_england,england_2020 )
```

```
data_list_italy <- create_list(length(mean_italy), mean_italy, italy_2020)
data_list_norway <- create_list(length(mean_norway), mean_norway, norway_2020)
data_list_france <- create_list(length(mean_france), mean_france, france_2020)
data_list_belgium <- create_list(length(mean_belgium), mean_belgium, belgium_2020)
data_list_germany <- create_list(length(mean_germany), mean_germany, germany_2020)
data_list_switzerland <- create_list(length(mean_swiss), mean_swiss, swiss_2020)
```

Fitting the stan models

```
data{
  int <lower = 0> N;
  real E[N];
  int O[N];
  real phi_a;
  real phi_b;
  real sigma_a;
  real sigma_b;
  real alpha_mu;
  real alpha_sigma;
  real beta_a;
  real beta_b;
  real sigma_time_mu;
  real sigma_time_sigma;

}

parameters{
  real log_theta[N];
  real log_theta_time[N];
  real <lower  = 0> phi;
  real <lower  = 0> sigma;
  real alpha;
  real <lower=-1, upper=1> beta;
  real<lower=0> sigma_time;
}

model{

  O[1] ~ neg_binomial_2(E[1]* exp(log_theta[1]),phi);
    log_theta[1]~ normal(0,sigma);
    log_theta_time[1] ~ normal(0,sigma);

  for (i in 2:N){
    O[i] ~ neg_binomial_2(E[i]* exp(log_theta[i] + log_theta_time[i]),phi);
    log_theta[i]~ normal(0,sigma);
    log_theta_time[i] ~ normal(alpha+beta*log_theta_time[i-1], sigma_time);
  }

  phi~ uniform(phi_a,phi_b);
  sigma ~ uniform(sigma_a,sigma_b);

  alpha ~ normal(alpha_mu,alpha_sigma);
  beta ~ uniform(beta_a,beta_b);
  sigma_time ~ normal(sigma_time_mu,sigma_time_sigma);
```

```
}
```

```r
risk_netherland <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_netherland, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth

risk_england_wales <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_england_wales, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treede

risk_italy <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_italy, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth = par

risk_norway <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_norway, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth = pa

risk_germany <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_germany, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth = p

risk_switzerland <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_switzerland, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth

risk_france <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_france, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth = pa

risk_belgium <- stan(params$stanfile, iter = params$iter, chains = 4,
  data = data_list_belgium, seed = 1, control = list(adapt_delta = params$adapt_delta, max_treedepth = p
```

There is a problem of divergence in our model meaning means and medians could be unreliable.

```r
#####################
# Helper functions
#####################

#Function for output graphs
plot_risk <- function(data, country_name){
  g<- ggplot(data = data, aes(x = week, y = theta)) + geom_point() + geom_line(lty = 2) +
    geom_line(aes(week, CI_lower), col = "blue", lty = 2) +
    geom_line(aes(week, CI_upper), col = "blue", lty = 2) +
    geom_hline(yintercept = 1, col = "red")+
    ggtitle("Relative risk for" , country_name)
  ggsave(paste0("output_",country_name,".png"))
  return(g)
}

# Function for creating the dataframe with thetas and the bounds for the credible interval.

create_CI_theta_vec <- function(risk_data, N){
  fit<-summary(risk_data)
  results <- as.data.frame(fit$summary)

  CI_upper <- exp(results$`97.5%`[1:N] +results$`97.5%`[(N+1):(2*N)])
  CI_lower <- exp(results$`2.5%`[1:N] +results$`2.5%`[(N+1):(2*N)])

  theta <- exp(results$mean[1:N] +results$mean[(N+1):(2*N)])
  week <- seq(1:N)
```

```
  data <- data.frame(week, CI_upper, CI_lower, theta)
  return(data)

}
```

4.2.1 The Netherlands

```
data_netherland <- create_CI_theta_vec(risk_netherland, length(mean_netherland))
plot_risk(data_netherland, "Netherland")
```

```
knitr::include_graphics("output_from_Report_1/output_Netherland.png")
```
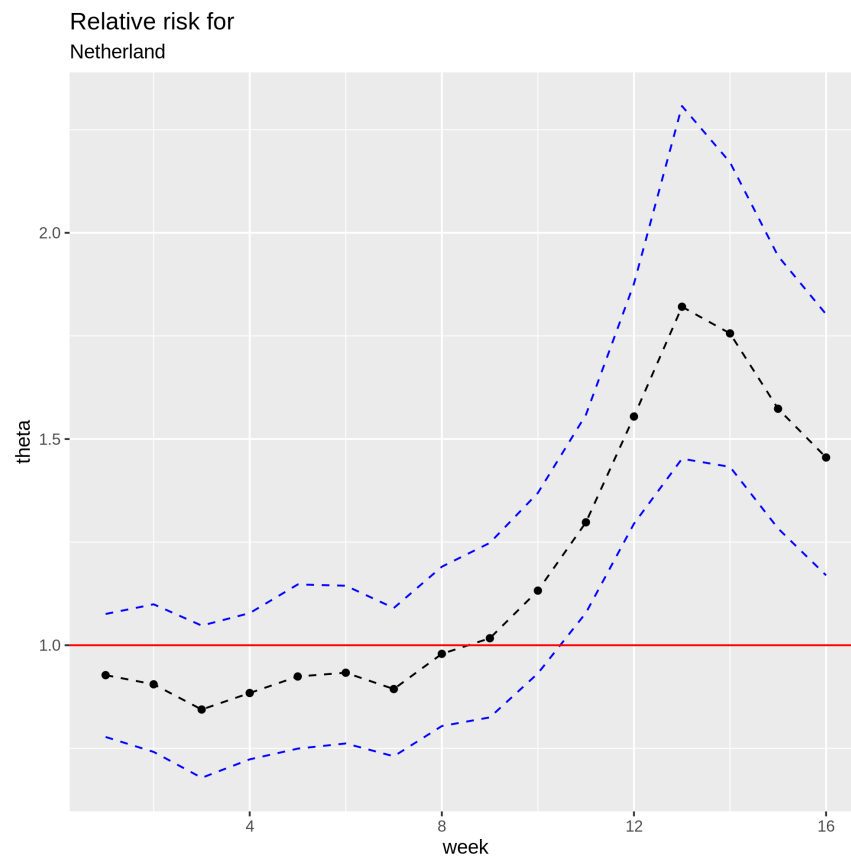

Figure 4: Negative binomial model estimates for the Netherlands

4.2.2 England and Wales

```
data_england_wales <- create_CI_theta_vec(risk_england_wales, length(mean_england))
plot_risk(data_england_wales, "England and Wales")
```

```
knitr::include_graphics("output_from_Report_1/output_England and Wales.png")
```

Relative risk for
England and Wales
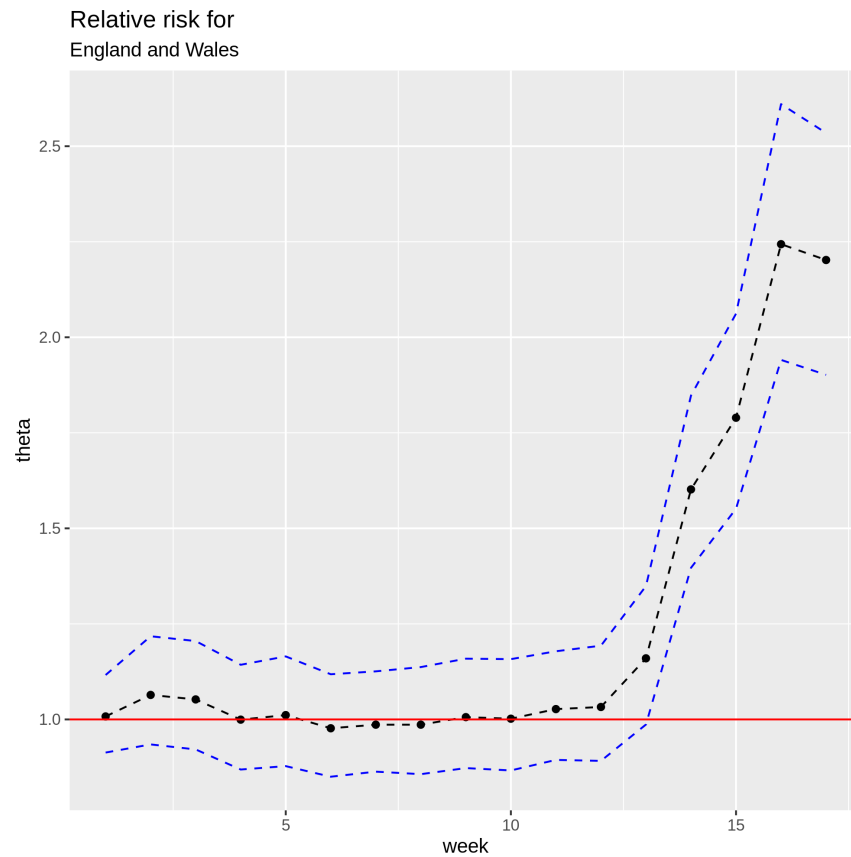
Figure 5: Negative binomial model estimates for England and Wales

4.2.3 Italy

```
data_italy <- create_CI_theta_vec(risk_italy, length(mean_italy))
plot_risk(data_italy, "Italy")

knitr::include_graphics("output_from_Report_1/output_Italy.png")
```
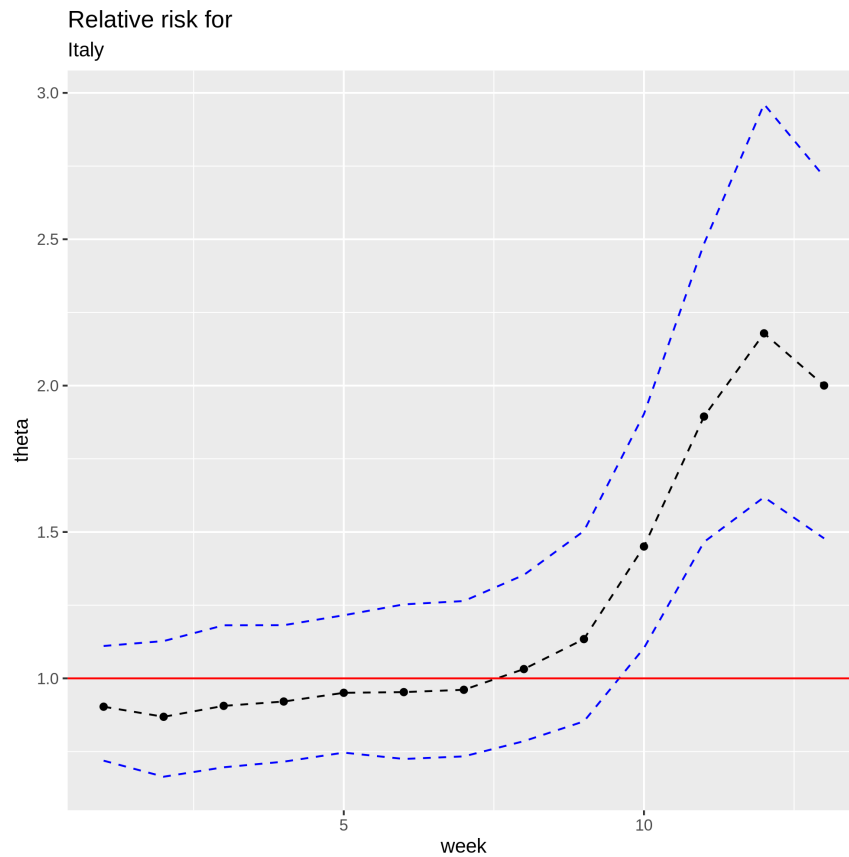
Figure 6: Negative binomial model estimates for Italy

4.2.4 Norway

```
data_norway <- create_CI_theta_vec(risk_norway, length(mean_norway))
plot_risk(data_norway, "Norway")
```

```
knitr::include_graphics("output_from_Report_1/output_Norway.png")
```
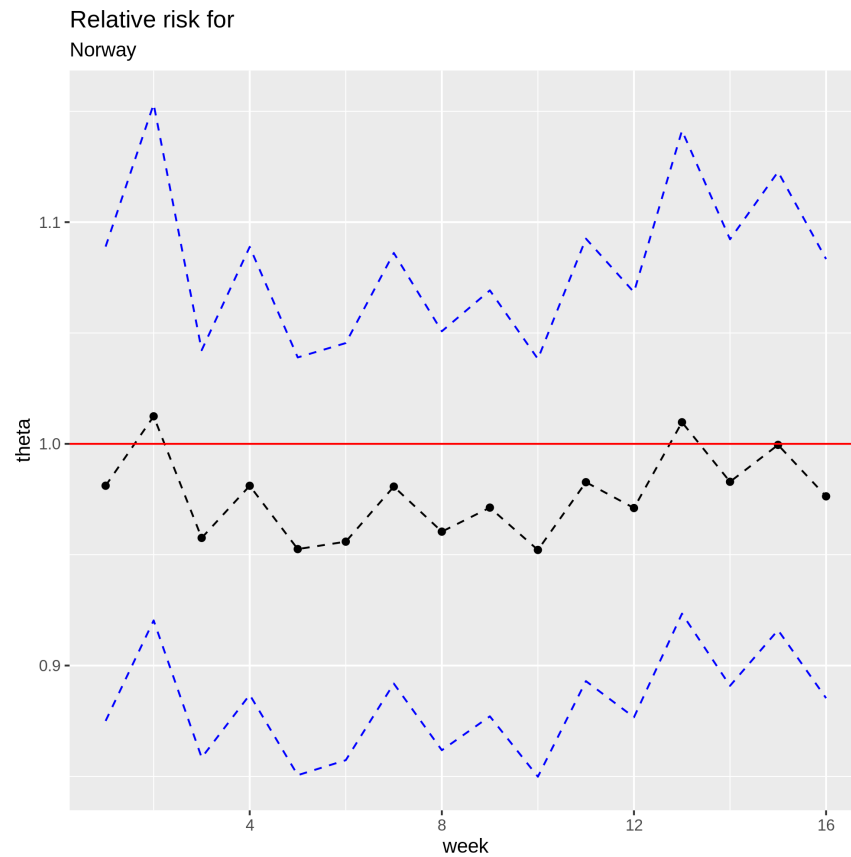
14

Figure 7: Negative binomial model estimates for Norway

4.2.5 Germany

```
data_germany <- create_CI_theta_vec(risk_germany, length(mean_germany))
plot_risk(data_germany, "Germany")
```

```
knitr::include_graphics("output_from_Report_1/output_Germany.png")
```
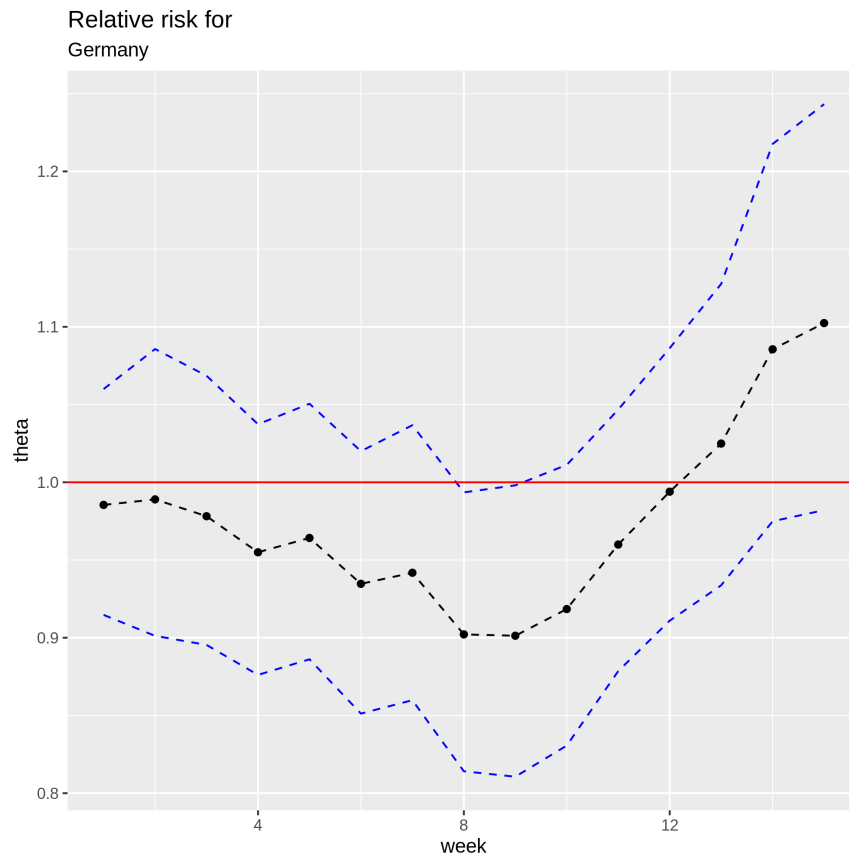
Figure 8: Negative binomial model estimates for Germany

4.2.6 Switzerland

```
data_switzerland <- create_CI_theta_vec(risk_switzerland, length(mean_swiss))
plot_risk(data_switzerland, "Switzerland")

knitr::include_graphics("output_from_Report_1/output_Switzerland.png")
```
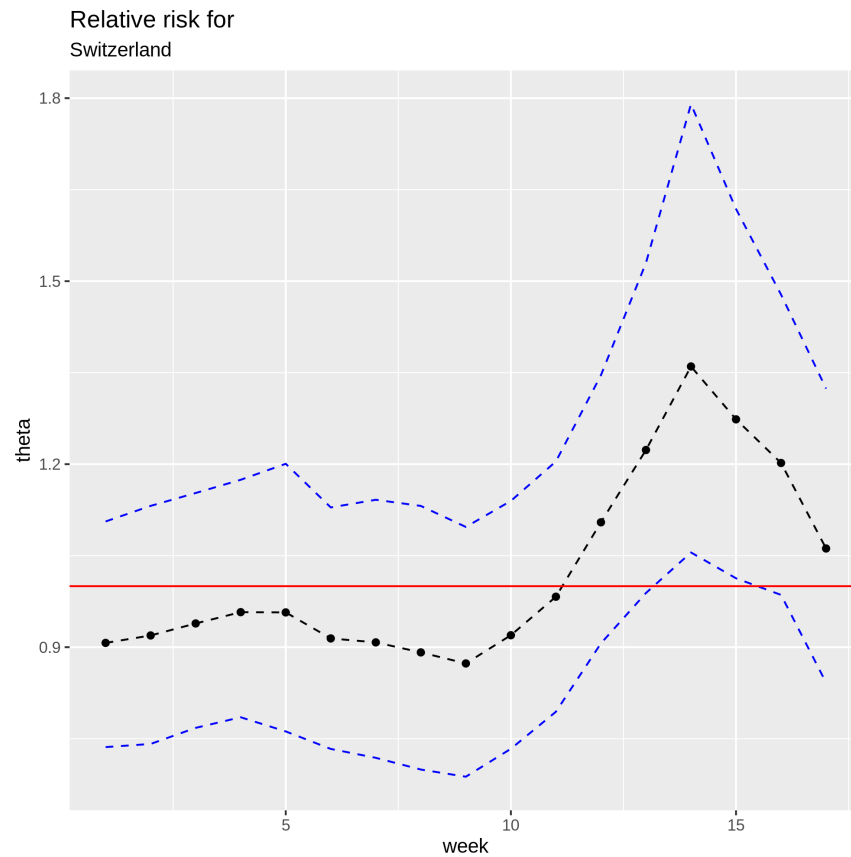
Figure 9:   Negative binomial model estimates for Switzerland

4.2.7   France

```
data_france <- create_CI_theta_vec(risk_france, length(mean_france))
plot_risk(data_france, "France")

knitr::include_graphics("output_from_Report_1/output_France.png")
```
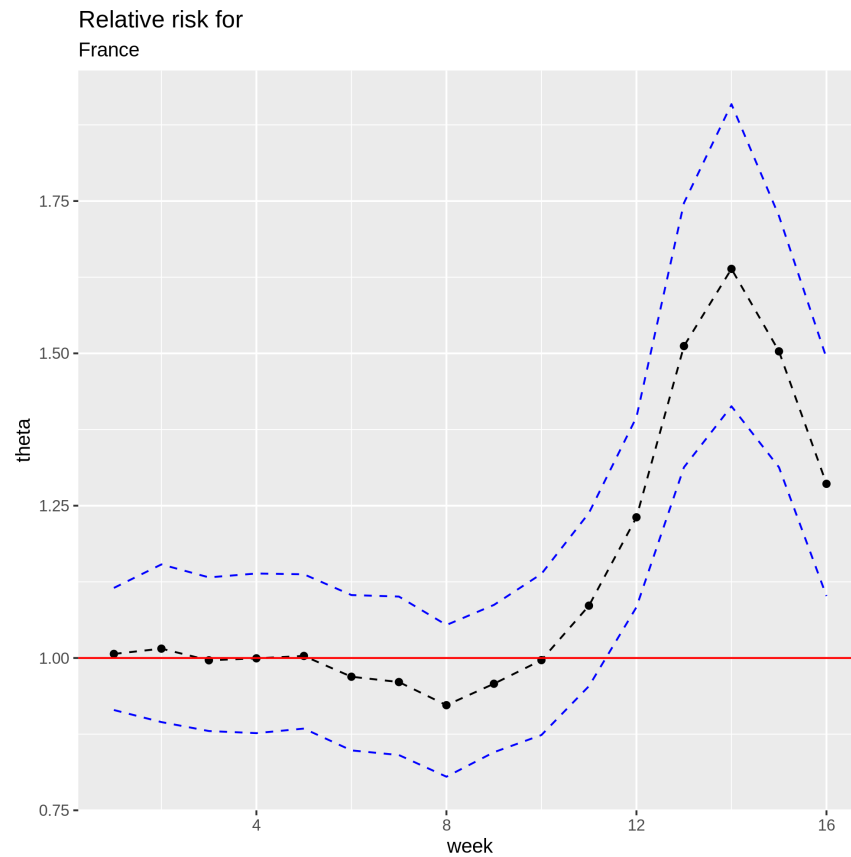
Figure 10:   Negative binomial model estimates for France

4.2.8   Belgium

```
data_belgium <- create_CI_theta_vec(risk_belgium, length(mean_belgium))
plot_risk(data_belgium, "Belgium")
```

```
knitr::include_graphics("output_from_Report_1/output_England and Wales.png")
```
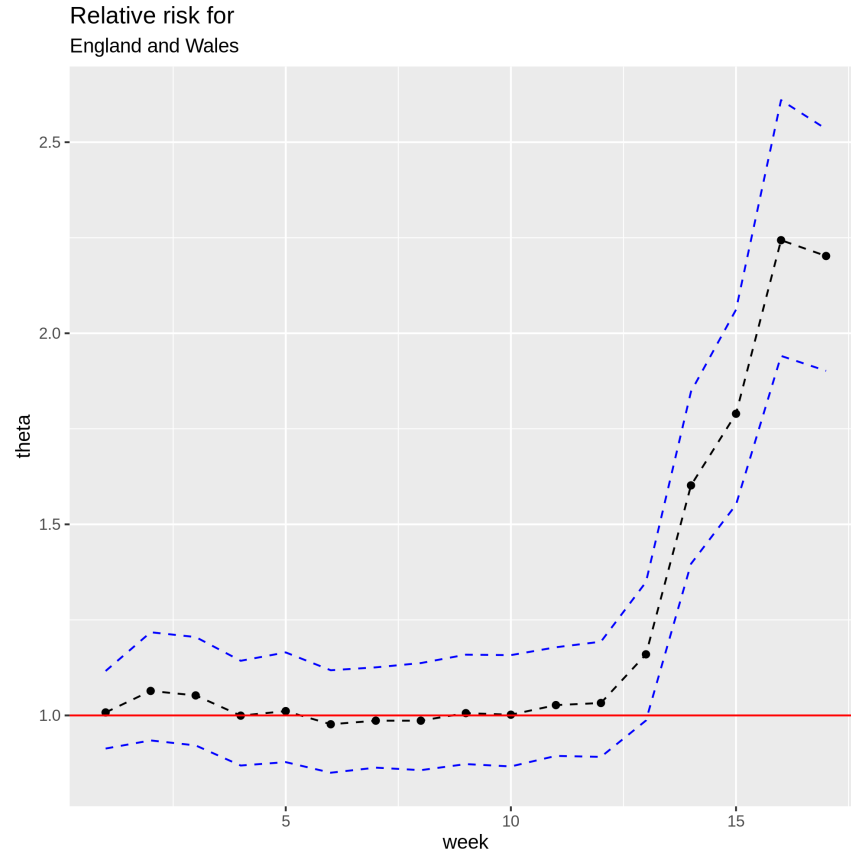
Figure 11: Negative binomial model estimates for Belgium

```
# Save output
save.image("excess_deaths_output.Rdata")
```

5   Discussion

Let's summarise briefly the main results coming out of this first analysis on COVID-19 data for each country. As expected, they reflect the evolution of the pandemy throughout Western Europe. In all the countries analyzed, expect for Norway, we clearly see a change in the behaviour of the curve of the risk factor, proving the incredible impact of the virus on our countries. We can practically divide the analyzed nations in two main categories: one in which falls the ones that reached the pick of deaths and that overcame it and the ones in which they're still fighting against the virus.
In the former group we can assign practically every country we studied, exception for Italy and England which we will comment separately. The trend is clearly ascending in the weeks in which pandemy exploded in Europe, but it decreases in the following period.
Even if the behaviour of the curve is the same, it's important to analyze the differences between these countries that come up from our analysis.
First thing that arises easily is that the values of the risk factor are disequal. For example in Germany, which claims a low number of deaths, the risk factor stays in the neighborhood of the value 1 which is taken as "normal threshold", while in France they reach pick of 1.7 and in Norway, in which there are around 200 certified deaths for COVID, the factor is less than 1.
It's worth to notice the beahviour of the curve of Belgium. Even if the total number of deaths registered for COVID-19 is low, the pick reached by the curve is really high.

Total different situation in Italy and England, probably the most affected countries in Europe, with the highest number of people who lost their lives because of the virus. Both the countries have a pick greater than 2. Italy data need to be analysed better since we have not a recent reliable source for the months of April and May. The risk, in this case, can be underestimated.

6 Future work

In the final report we will address the following points:

- convergence issues,

- analyse further the time structure significance,

- implementation and obtention of results for the excess deaths on top of reported deaths,

- exploring the possiblity of an additional layer in our hierarchical negative binomail model by gathering the different countries into the same model,

- validation of the model.