Excess death in relation to Covid-19

Aurora Hofman, Giovanni Rigolino, Paul Rognon

16 de May, 2020

Contents

1 Introducion **1**

2 Data **2**
2.1 Description 2
2.2 Exploratory data analysis 2

3 Model **4**
3.1 Excess mortality 4
3.2 Excess mortality on top of the reported COVID deaths 7

4 Results **7**
4.1 Poisson model 7
4.2 Negative binomial model 8

5 Discussion **17**

6 Future work **18**

1 Introducion

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 of the severity of COVID-19 and many has compared it to the yearly flu outbreaks. 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 out the likelihood of 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 within regions have casted more doubts on the veracity of the reported 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 analysts is a first approach but is 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 the 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 in Table 1.

Table 1: Number of deaths by week in the Netherlands

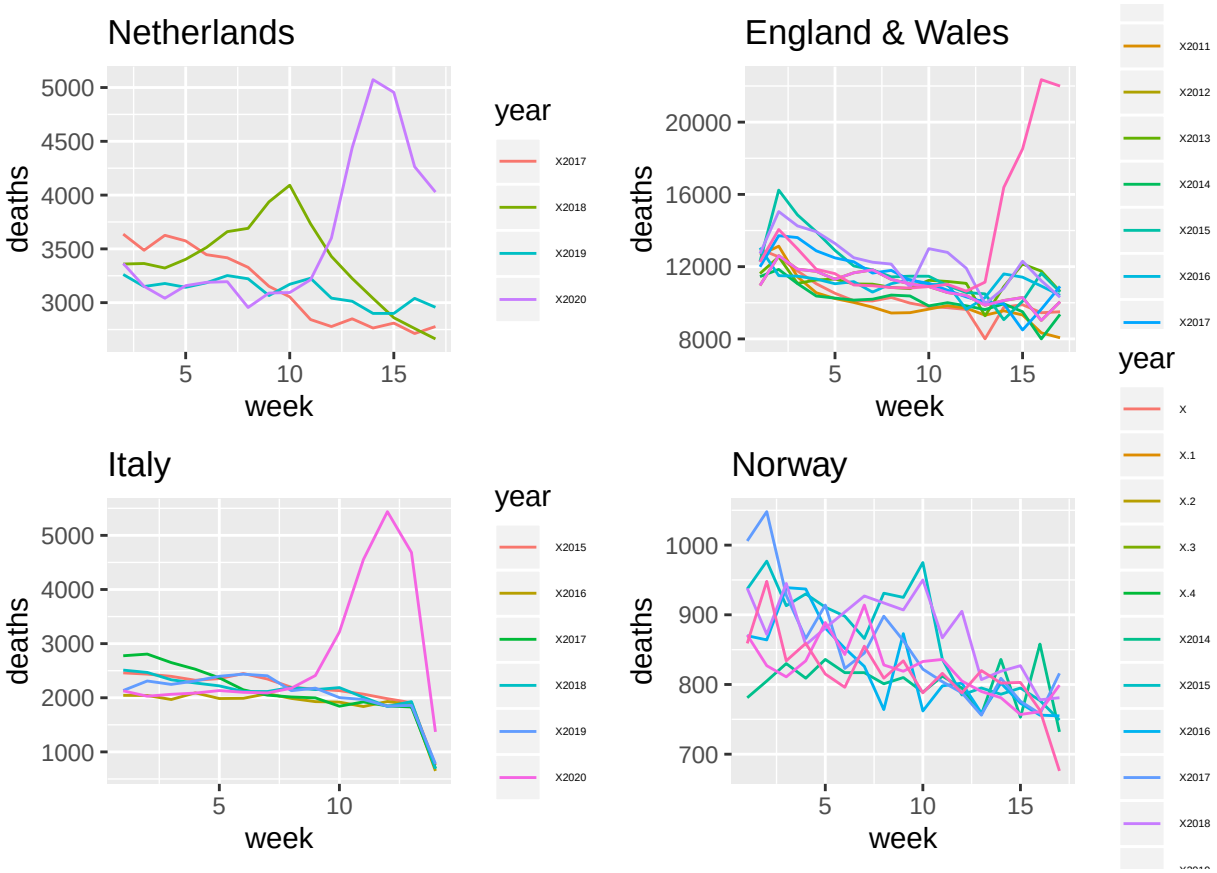
week	X2019	X2018	X2017	X2020	average
2	3262	3359	3637	3364	3637
3	3150	3364	3487	3152	3487
4	3178	3322	3626	3041	3626
5	3143	3403	3574	3158	3574
6	3185	3513	3446	3189	3446
7	3252	3660	3417	3195	3417

2.2 Exploratory data analysis

2.2.1 Visualization

Fig.**??** and Fig.1 show the weekly deaths from weeks
1 to 17 for each of the countries and each of the years. Most countries have graphs showing the same pattern

with a large peak between weeks 10 and 16 in 2020 which clearly stands out of from other years. Notably, Switzerland graph 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. On the contrary it shows a decrease, which might be due to a lag in the registration of deaths.



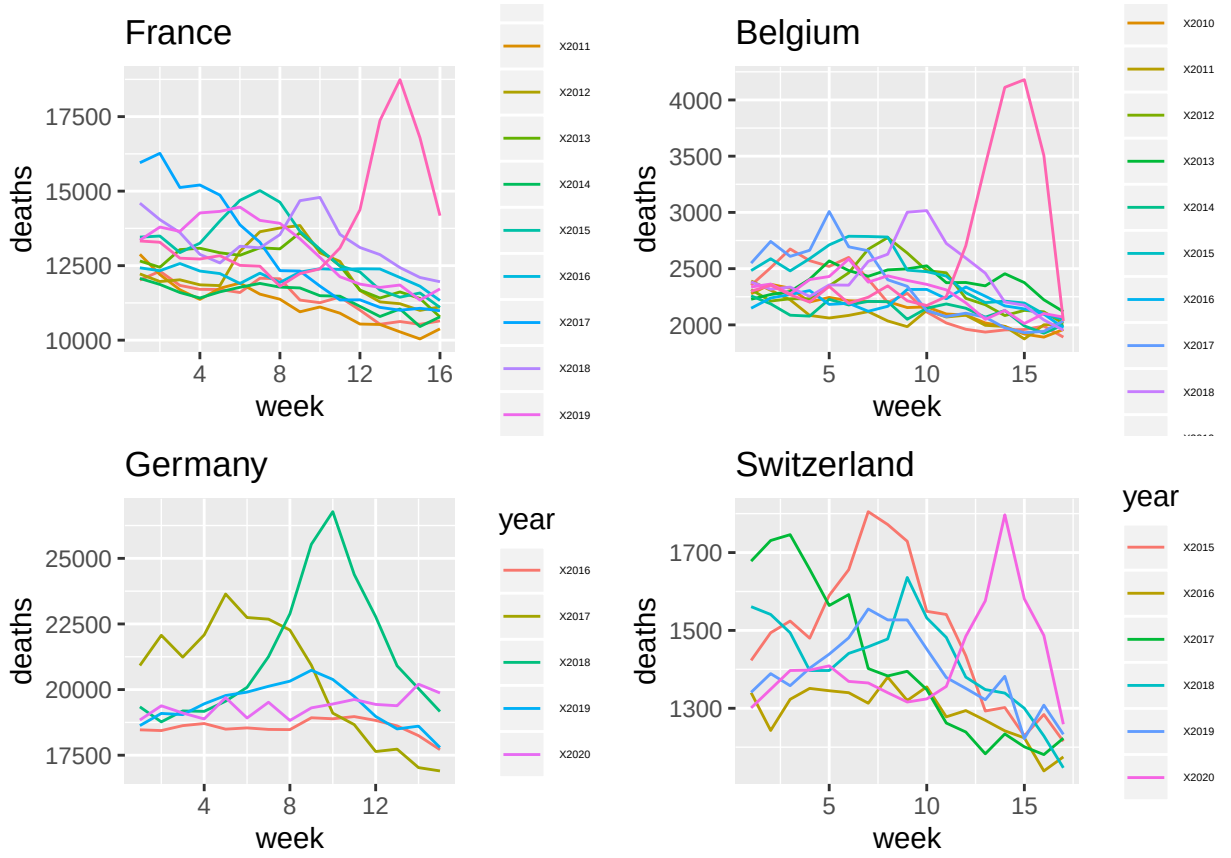


Figure 1: Weekly deaths by year and country

2.2.2 Overdispersion

Our data is characterized by much larger variances than means. As an example Table 1 shows the weekly means and variances 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) which 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 an extra parameter that allows to correct for overdispersion observed when using the Poisson model. In other words the Negative Binomial distrubution allows a much higher variance.

We have:

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

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:

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

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

In these 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 therefor introduce a time structured random effect θ_t in our models. We use an autoregressive structure of order 1 as time structure. Our models then become:

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

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

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):

$$\forall i, (O_{COV,i}|\theta_i) \sim \mathcal{P}\wr\rangle\int(E_i\theta_i\theta_{t,i})$$

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:

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

Similarly the negative binomial model (2) is:

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

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:

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

Choise of hyperpriories

The choises for the hyperpriories and there distributions are slightly based on trial and error.

For the dispurtion parameter it is important that it containes the value of the historical avarage since a dispurtion of this size would make the variance linear in the average instead of quadratic which corresponds

to variances observed in historical data. However we did not want to rule out smaller or larger values. Hence the choice fell on a uniform distribution with bounds $a_\phi = 1$ and $b_\phi = 6000$.

For σ we choose a uniform distribution since we do not have information about the variance of this risk factor. One wants the riskfactor to have the possibility of taking values with quite some variance hence the bounds were chosen to be $[0.01, 4]$

The α parameter is the costant with regards to the timedependency and is expexted to be around zero. This leads us to a normal distribution with $\mu_\alpha = 0$ and $\sigma_\alpha = 4$ to allow for other values.

The β parameter is the weight og the timedependency parameter which should fall in the interval $[-1, 1]$ hence a uniform distribution with these bounds are choosen.

Finally the parameter σ_t was chosen to be a halfnormal distribution with parameters $\mu_{\sigma_t} = 0$ and $\sigma_{\sigma_t} = 25$ to allow for a large variance also here. Although one expects the variance to be closer to zero. The halfnormal ensures possitive values.

3.2 Excess mortality on top of the reported COVID deaths

To model excess mortality on top of reported COVID deaths, we will use the same approach as for the 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 $\tilde{E}_i$ 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 was deemed insuitable after having tried numerous variations of the hierarchical model such as changing the priors and levels of the multilevel model. Indeed, as can be observed in Fig. 2 for Belgium, the model adequately captured the change in relative risk in peak weeks but failed to adequately reflect variance with extremely narrow credible 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")
```

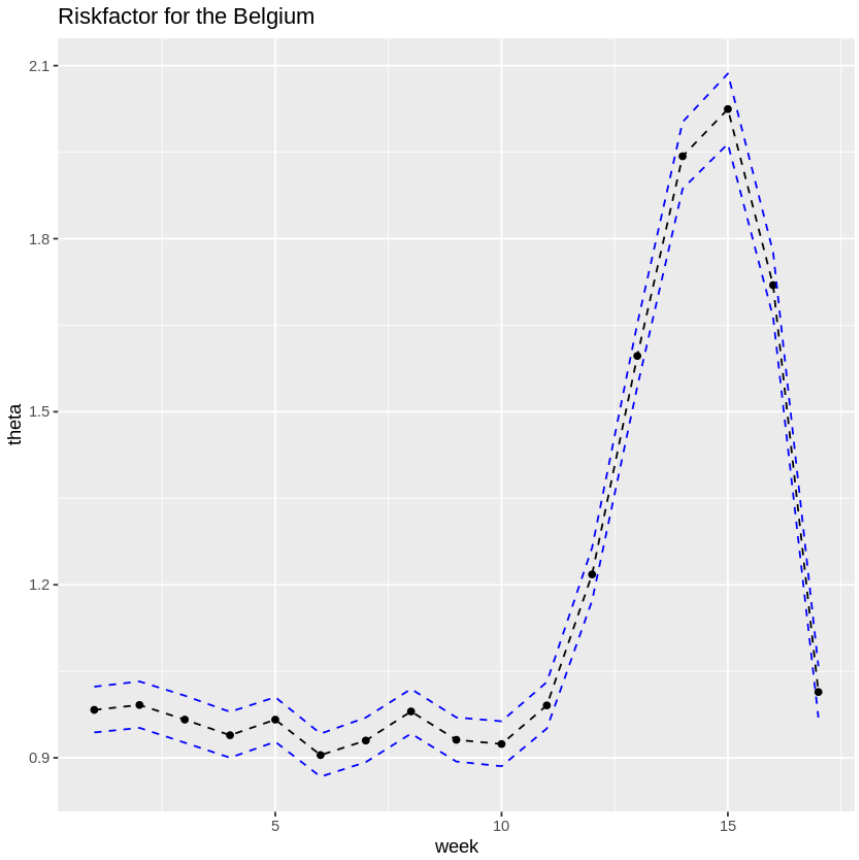


Figure 2: Poisson model estimates for Belgium

4.2 Negative binomial model

The negative binomial model proved to be much more suited to our data. This is not surprising givent the high variance we observed in the exploratory data analysis. For all countries the historical average and the deaths for 2020 were extracted form the data.

The following function is used to sett up the stan lists which is done for all countries.

```
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)
}
```

Below the stan model used to fit the data is shown, the model is fitted to the data from all the countries.

```
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);
}
```

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

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")
```

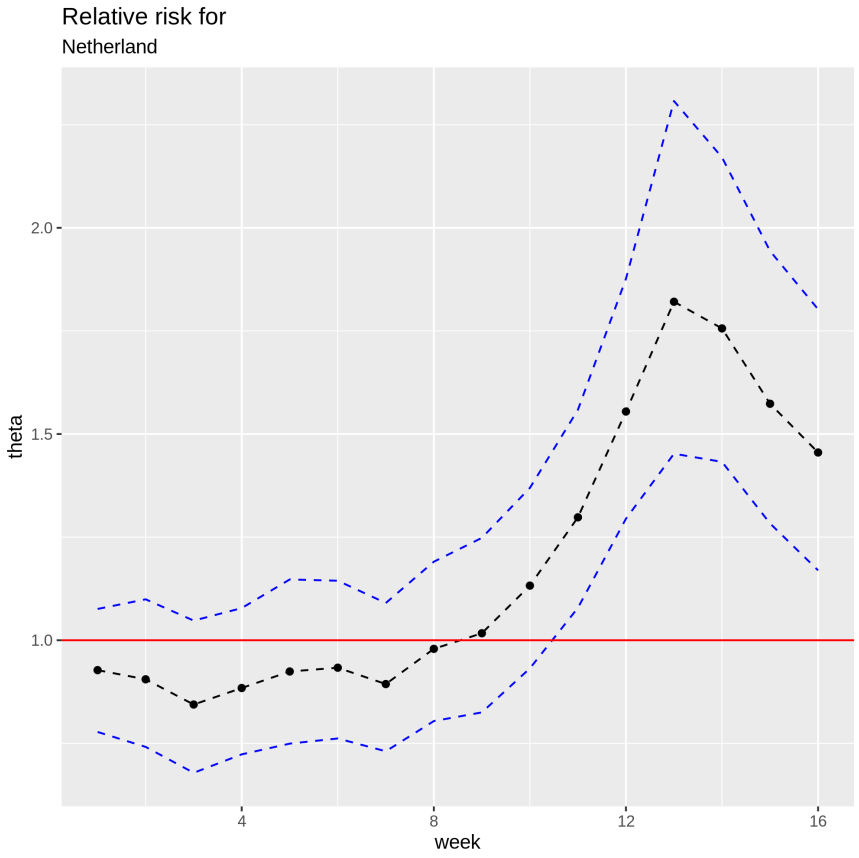


Figure 3: Negative binomial model estimates for the Netherlands

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")
```

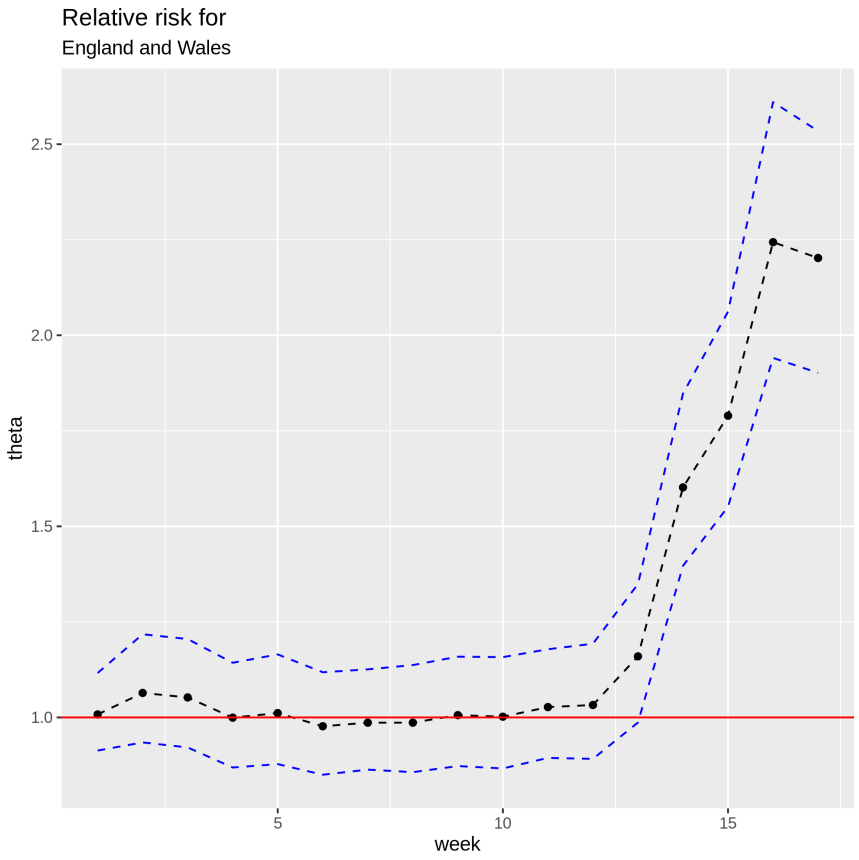


Figure 4: Negative binomial model estimates for England and Wales

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")
```

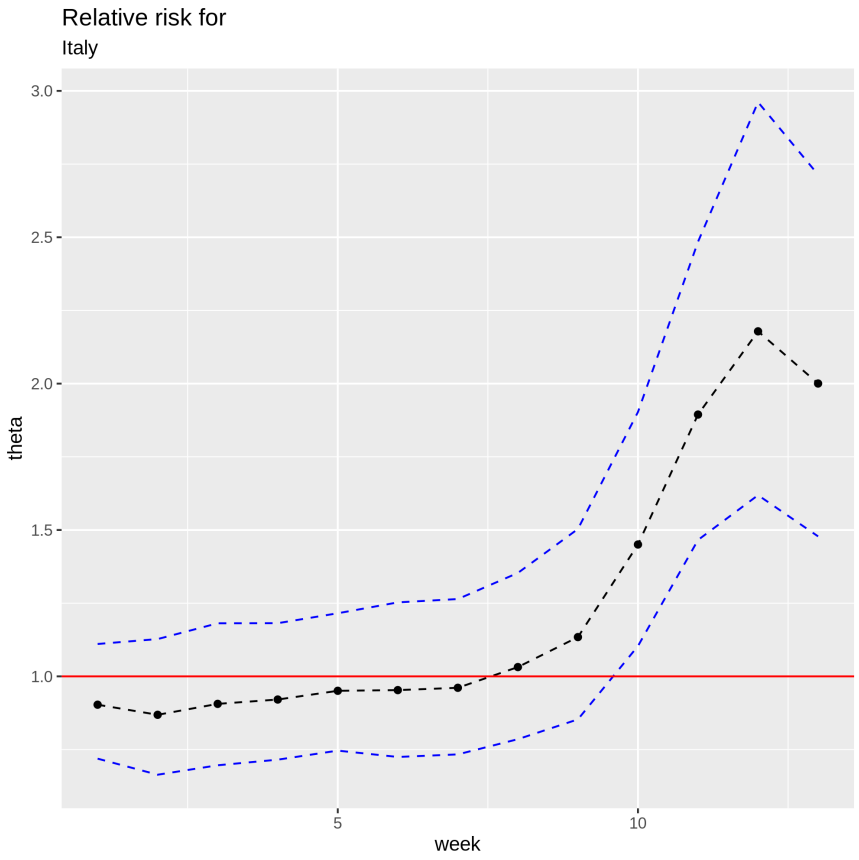


Figure 5: Negative binomial model estimates for Italy

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")
```

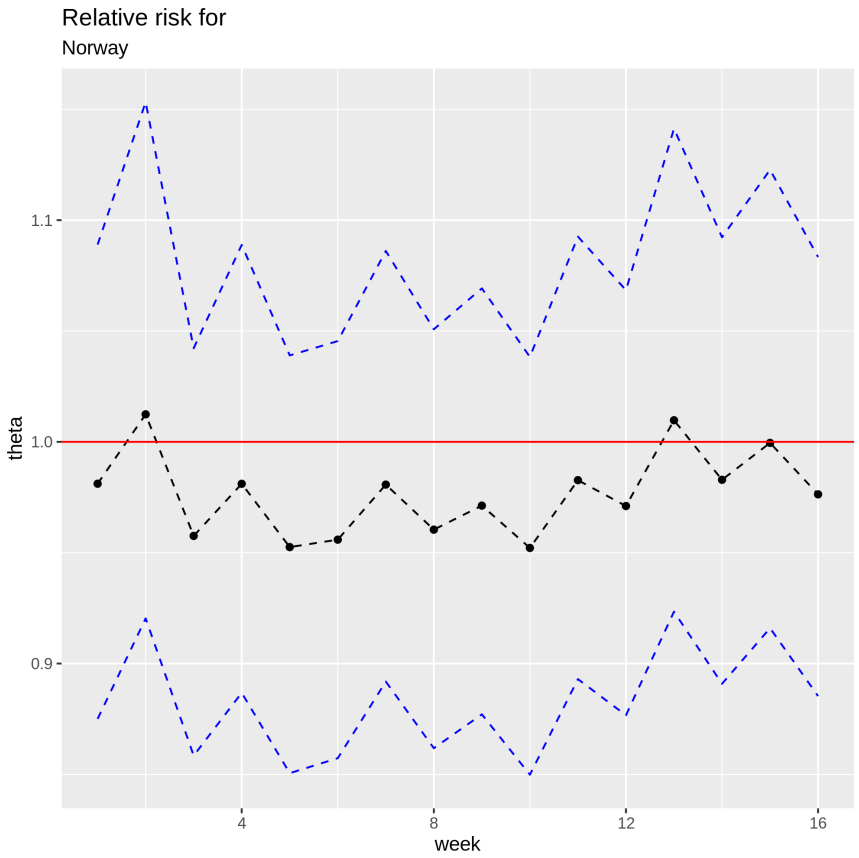


Figure 6: Negative binomial model estimates for Norway

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")
```

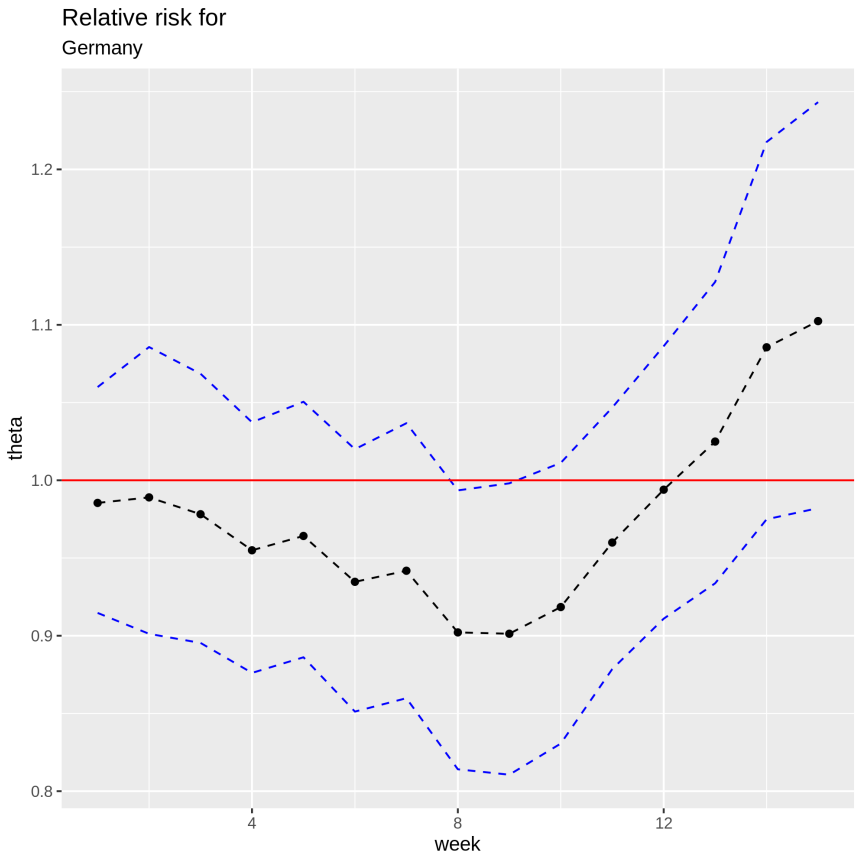


Figure 7: Negative binomial model estimates for Germany

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")
```

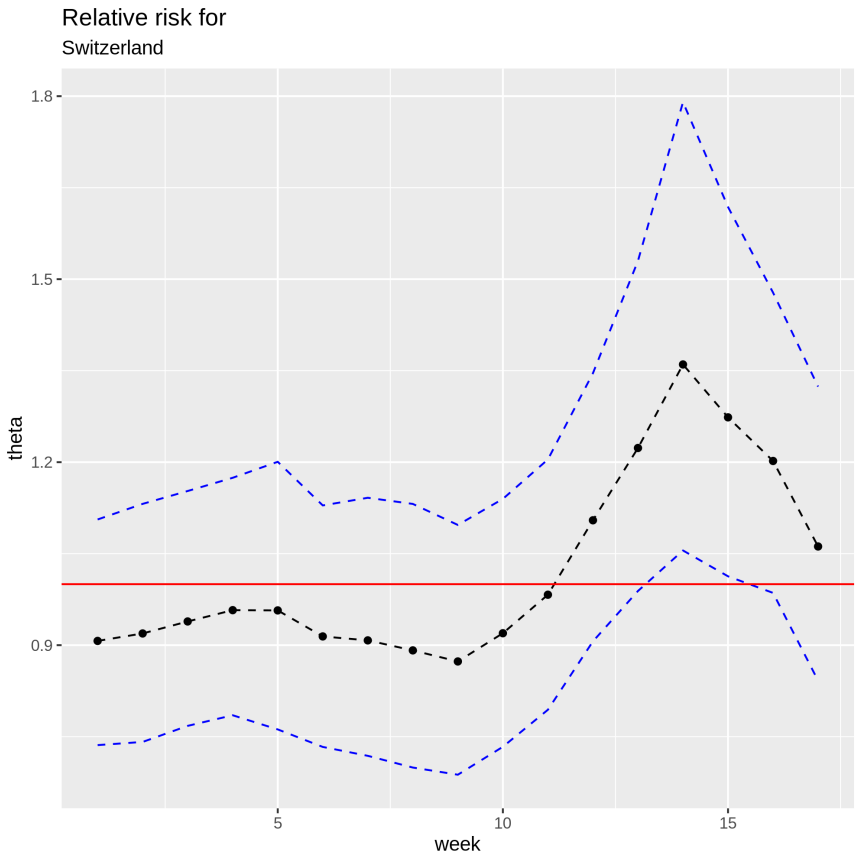


Figure 8: Negative binomial model estimates for Switzerland

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")
```

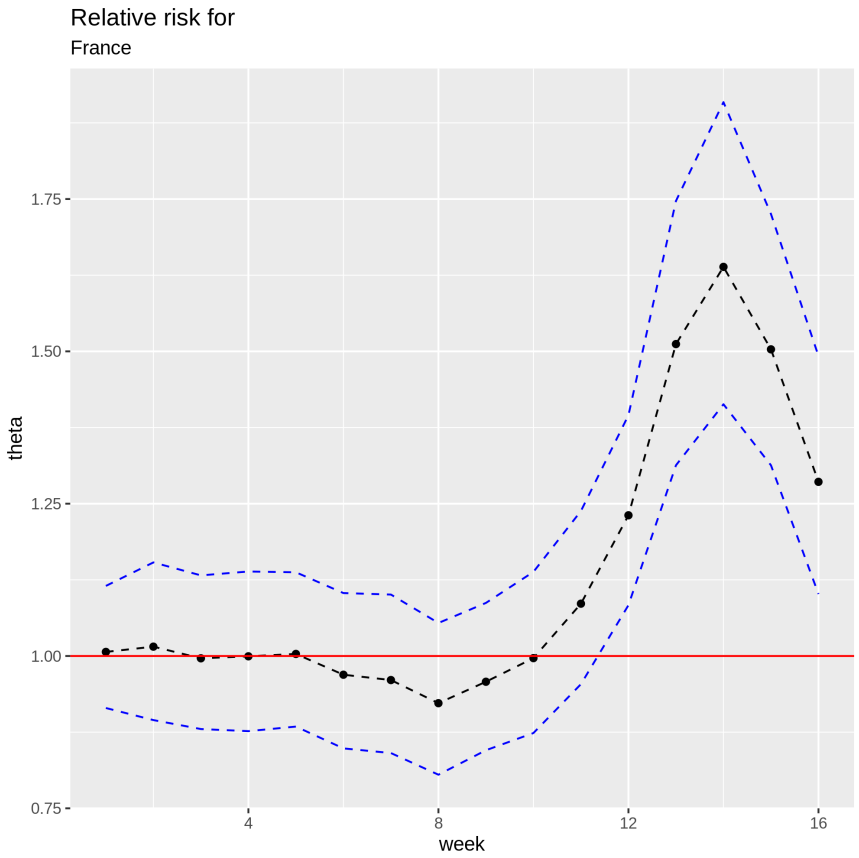


Figure 9: Negative binomial model estimates for France

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")
```

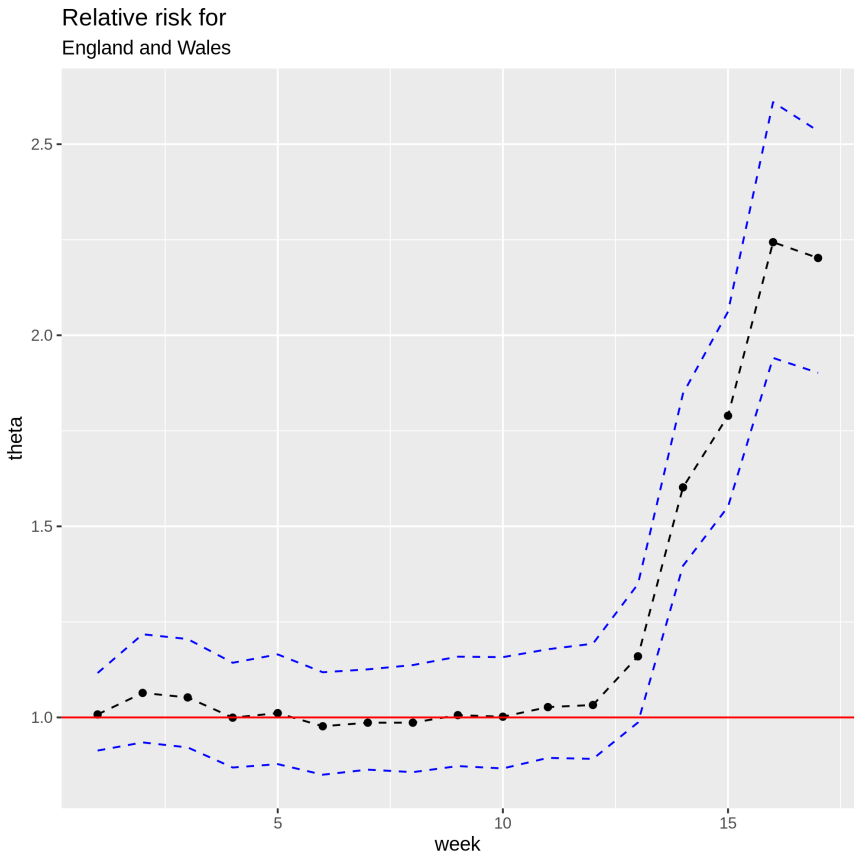


Figure 10: Negative binomial model estimates for Belgium

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

5 Discussion

Let us summarise briefly the main results coming out of this first analysis on COVID-19 data for each country. The analisis reflect the evolution of the pandemy throughout Western Europe where as expected for most countries one can observe a peak which is currantly on its way down.
In all the countries analyzed, expect for Norway, one can see a change in the behaviour of the risk factor around the time the virus impacted the different countries.

With respect to the question formulated in the introduciton:

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

One can for some countries conclude that for sevels weeks the risk has been significantly higher than in non outbrake times, however there are big national differences.

The countries Italy, England & Wales, France, Belgium and the Netherland have had periodes with an estimated relativ riskfactor above 1.6 and several weeks with a significant increase in relative risk. The countries Norway, Germany and Switserland has been much less affected and have had non or very few weeks with a significant increase in relative risk.

Hence when exploring the seccond question porposed in the introduction:

Is there significant excess mortality on top of the reported COVID deaths?

we will foxus on the first group of countries.

It is important to note that the model used for this analysis has some issues with regards to convergance. However it captures the time before the outbrake of COVID-19 relativly good for all the countries, meaning that one, the normal situation, falles withing the credible interval. Also the riskfactor obtained by the model is quite accurate when multiplied by the historical average to estimate the number of deaths in 2020. Hence one can say the model is usefull although not correct.

6 Future work

In the final report we will address the following points in more detail:

- convergence issues,
- analyse further the time structure significance,
- implementation and obtentain 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.