Poison_time_dep

Aurora Hofman, Giovanni Rigolino, Paul Rognon

10 de June, 2020

Contents

1 Introduction

As of May 15th 2020, the global death toll of the unfolding COVID-19 outbreak stands at 302 493 [1]. 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 referred the COVID-19 as a 'little flu' and refused to implement in his country the drastic lockdown measures that many other countries have enforced [2].

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 [3].

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 [4].

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 1 to 18 (end of April) in the following countries and following years:

- Norway: years 2014 to 2020 [1]
- Belgium: years 2009 to 2020 [2]
- France: years 2010 to 2020 [3]
- England and Wales: years 2010 to 2020 [4]

We picked those four countries for different reasons: Belgium for its peculiar death count methodology, France and England and Wales because they are among the most affected countries and Norway because it did not have a large outbreak and therefore offers a good benchmark. 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.

The reported COVID deaths for Norway, Belgium and France have been extracted from the Center for Systems Science and Engineering's repository at John Hopkins University Whiting School of Engineering [5]. The reported deaths for England and Wales from the Office of National Statistics [4].

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx SUMMARY AND TODO:

In Fig.1 try and fixs the legends (matching of colour and year) It seems like 2020 has been labled as 2019. Also for england & wales some years are missing.

Validation looks good for all countries, would be nice is the line for 2020 was in black or some colour it is possible to see. This is however not important!!

Implementation with koronadata: Done and looking good :) Seems like a lot of countries experienced a period of less deaths after "lockdowm / focus on covid" this oculd be due to people in general beeing a lot more carefull and hence not dying from "stupid things".

Report:

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Table 1: Number of deaths by week in England and Wales

Week	2017	2018	2019	2020
1	11991	12723	10955	12254
2	13715	15050	12609	14058
3	13610	14256	11860	12990
4	12877	13935	11740	11856
5	12485	13285	11297	11612
6	12269	12495	11660	10986
7	11644	12246	11824	10944
8	11794	12142	11295	10841
9	11248	10854	11044	10816
10	11077	12997	10898	10895
11	10697	12788	10567	11019
12	10325	11913	10402	10645
13	10027	9941	9867	11141
14	9939	10794	10126	16387
15	8493	12301	10291	18516
16	9644	11223	9025	22351
17	10908	10306	10059	21997
18	9064	10153	11207	17953

2.2 Visualization

2.2.1 Weekly deaths

Fig.1 show the weekly deaths from weeks 1 to 18 for each of the countries and each of the years. We observe several common patterns across countries. Firstly, they all show a downtrend with larger numbers of deaths in the first weeks of the year than in spring. Secondly, there is noise and peaks around this trend with some years being more deadly than others. There are very few dips in the data, the mortality can increase sharply but rearly decrease very rapidly from one week to the next. Earlier weeks are more often subjects to outbreaks of flu and other winter diseases than later weeks. Finally, most countries have a large peak between weeks 10 and 17 in 2020 which clearly stands out of from other years. On the contrary, Norway shows no peak.

2.2.2 COVID deaths and mean weekly deaths

In Fig. 2, we compare the 2020 weekly deaths to the COVID reported deaths summed to the mean weekly deaths. We observe for England and Wales a large difference between the two curves for weeks 13 to 18: there is excess mortality on top of the expected deaths plus reported COVID deaths. In Norway, 2020 deaths are below average up to week 13. After this the 2020 deaths are higher than the mean but still lower than COVID and mean deaths together, suggesting a lower mortality in Norway in 2020 for non COVID related deaths which COVID deaths do not compensate. Finally for Belgium and France, we observe larger 2020 deaths than mean deaths and COVID together up to week 15, suggesting again excess mortality on top of expected deaths plus reported COVID deaths. However from week 16, the mean deaths plus COVID curve is higher than 2020 deaths.

Overall in this quick exploration of our data, we find no issue or dubious data point.

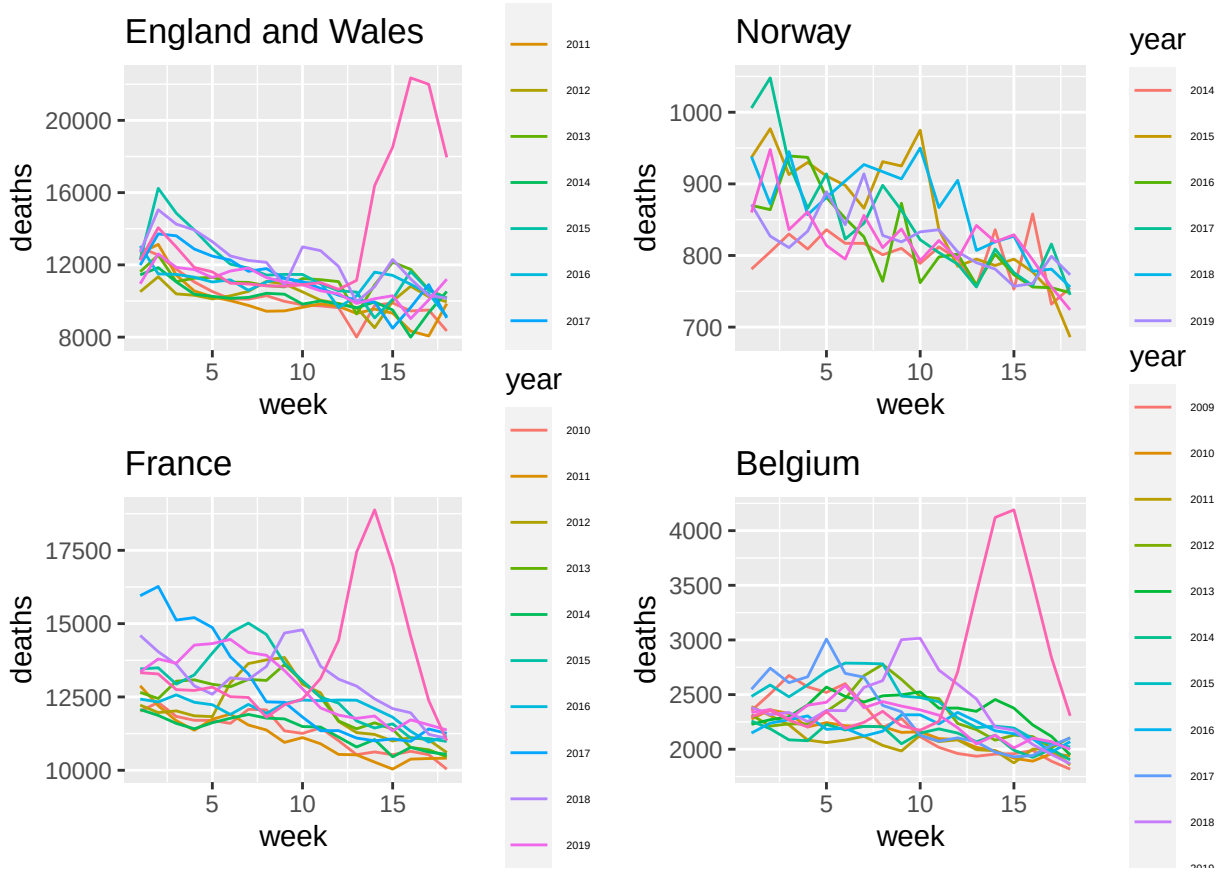

Figure 1: Weekly deaths across years

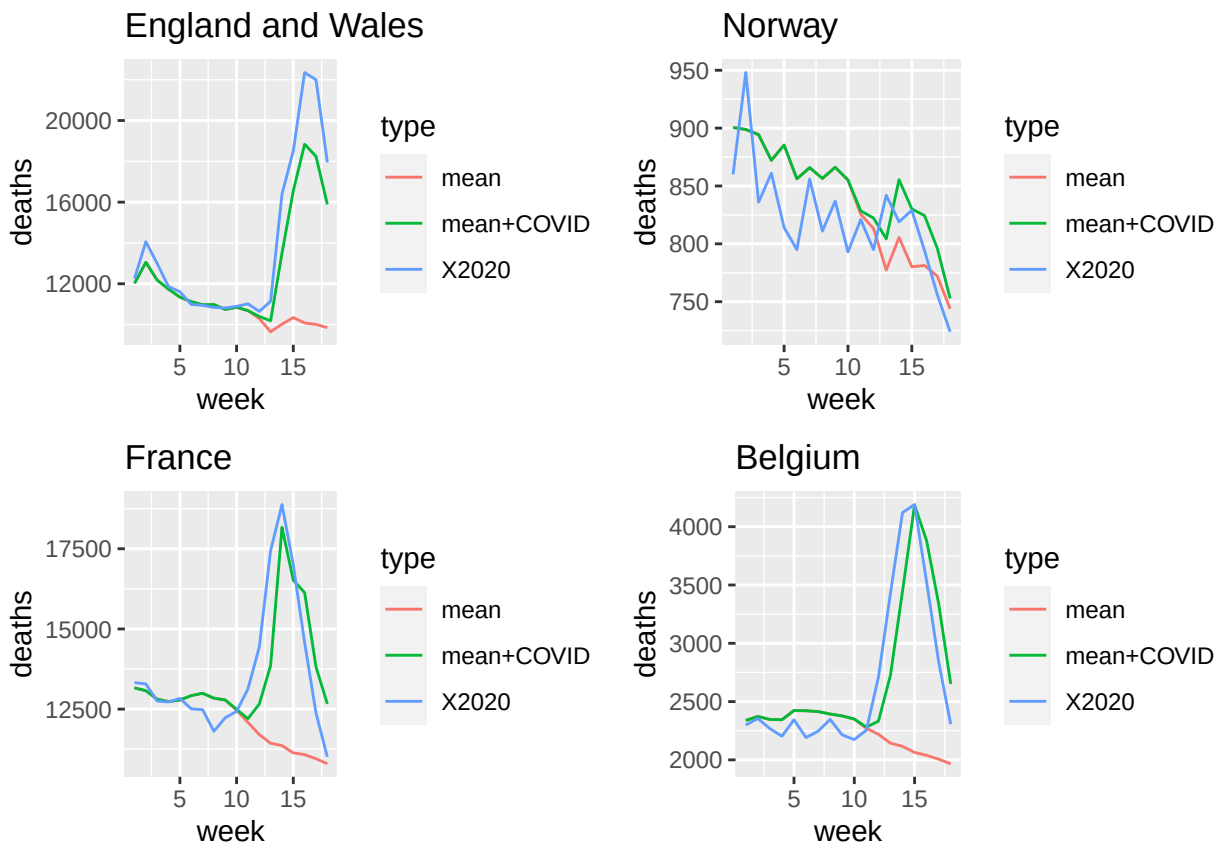

Figure 2: 2020 weekly deaths, mean weekly deaths and COVID reported deaths

3 Model

3.1 Excess mortality

3.1.1 Statistical model

For each country separately, we approach the modelling of excess mortality by estimating the relative risk γ understood as the ratio of weekly mortality during the COVID-19 outbreak and the weekly mortality in non-outbreak times.

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

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 across years.

The number of deaths O is a count variable. For such variable, a commonly used statistical model is the Poisson model

$$\forall i \in \{1, ..., 18\}, O_{COV,i} \sim \mathcal{PO}(\lambda),\ E(O_{COV,i}) = Var(O_{COV,i}) = \lambda$$

We have:

$$\forall i \in \{1, ..., 18\}, O_{COV,i} = O_{noCOV,i} \cdot \gamma_i \Rightarrow E(O_{COV,i}) = E(O_{noCOV,i})\gamma_i = E_i\gamma_i$$

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 model for each country separately:

$$\forall i \in \{1, ..., 18\}, O_{COV,i} \sim \mathcal{PO}(E_i\gamma_i)$$

In these models, a value of γ_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.

From the exploratory data analysis, we see that there is a clear time dependency in the weekly mortality. We therefore decompose our relative risk γ_i betwen a fixed component θ_i and a time structured effect $\theta_{i,t}$ in our models.

$$\gamma_i = \theta_i\theta_{t,i} \Rightarrow \log(\gamma_i) = \log(\theta_i) + \log(\theta_{t,i})$$

We use an auto-regressive structure of order 1 as time structure. Our models then become:

$$\forall i \in \{1, ..., 18\}, O_{COV,i} \sim \mathcal{PO}(E_i\gamma_i) = \mathcal{PO}(E_i\theta_i\theta_{t,i})$$

where:

$$\forall i \in \{1, ..., 18\}, \log(\theta_{t,i}) = \alpha + \beta \log(\theta_{t,i-1})$$

3.1.2 Bayesian model

We are modelling the number of deaths $O_{COV,i}$ across weeks with a Poisson distribution that depends on the relative risk of each week γ_i. We are assuming different risks for each week but at the same time the risks share characteristics such as being expected at a value around 1 and having a larger probability towards higher values (peaks) than lower values (no trough). A convenient way to model this is using a common distribution for each of the components of γ_i, that is to use a Bayesian hierarchical model. We then propose a Bayesian hierarchichal model based on the Poisson distribution for each country separately:

$$\forall i \in \{1, ..., 18\}, (O_{COV,i}|\theta_i, \theta_{t,i}) \sim \mathcal{PO}(E_i\theta_i\theta_{t,i})$$

with parameters priors:

$$(\log(\theta_1|\sigma), ..., \log(\theta_{18}|\sigma)) \sim \mathcal{N}(0, \sigma)$$

$$\forall i \in \{1, ..., 18\}, \log(\theta_{t,i}|\sigma_t) = \alpha + \beta \log(\theta_{t,i-1}|\sigma_t) + \epsilon, \;\; \epsilon \sim \mathcal{N}(0, \sigma_t)$$

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

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

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 γ understood as the ratio of weekly mortality during the COVID-19 outbreak and the weekly expected mortality in non-outbreak times. We now adjust, for each country separately, the expected mortality $\tilde{E}_i$ to account for COVID deaths.

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

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:

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

The updated statistical model is:

$$\forall i \in \{1, ..., 18\}, O_{COV,i} \sim \mathcal{PO}(\tilde{E}_i\gamma_i) = \mathcal{PO}(\tilde{E}\theta_i\theta_{t,i})$$

where:

$$\forall i \in \{1, ..., 18\}, \log(\theta_{t,i}) = \alpha + \beta \log(\theta_{t,i-1})$$

The priors and hyperpriors for the Bayesian model do not change.

3.2.1 Choice of priors

Our choice of priors and hyperpriors has been driven by prior knowledge or lack of it. First of all, we choose a normal distribution for $\log(\theta_i)$ with mean 0 and standard deviation the hyperparamter σ. As a result θ_i is lognormally distributed and centered at 1, the expected value for the relative risk. The lognormal distribution is particularly suited for several reasons. It has a positive support and we know θ_i must be positive. The lognormal distribution has also a positive skew, so the median of θ_i is 1 but the mean is larger and very large values are possible. It is important to have such rather heavy right tail because we know the risk can increase largely, for deadly events. On the contrary we do not need a heavy left tail because the mortality is floored by the natural mortality determined by the population demographics. For the time dependent componenet of relative risk, we make the same distributional assumption where the center of the distribution is not 1 anymore but the autoregression function estimate $\alpha + \beta\theta_{t,i-1}$, that represents the dependency on previous value.

3.2.1.1 Choice of hyperpriors

The choice of a uniform distribution with support $(-1, 1)$ for β is motivated by our assumption of autoregressive model of order 1 for $\log(\theta_{t,i})$. When the autoregresive process is stationary, β is the autocorrelation between two consecutive values and must be between -1 and 1. The choice of a non informative uniform distribution reflects our lack of further knowledge on the distribution of β. We choose for α the normal distribution, commonly used in linear regression settings. We assume the distribution of α to be centered at 0 because in a stationary autoregressive model of order 1, a null constant term α makes a null expectation for the modelled variable:

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

As our modelled variable $\theta t_{,i}$ is log transformed, it converts to an expectation of 1 when untransformed. We choose a small value for σ_α, 0.2, to prevent large deviations towards extremely high risk.

Finally, we choose non-informative uniform distributions for σ and σ_t because we do not hold prior information on them. We define small supports, $(0, 0.2)$, to again prevent large desviations towards extremely high risks, as exemplified below.

We check our prior by visualizing the prior predictive distribution of the relative risk and the number of deaths. We first simulated the prior predictive distribution of the total relative risk γ across all weeks and present the results in Fig. 3. We see the obtained distribution is concentrated around 1, the expected value for the risk, with the majority of the density betwee 0.5 and 1.5, a reasonable range of values for the risk. The quantile 5% stands at 0.51 and the quantile 95% at 1.936. The right tail is heavy which is appropriate to capture peaks.

We simulated the prior predictive distribution of the number of deaths for weeks 1 to 18 and plot the results in Fig. 4 for each country. We see the obtained distributions are roughly centered around the historical mean up to 2019. For every week, the range of values is wide, allowing for large but reasonable variations of the number of deaths. Accordingly to the idea of accomodating peaks, there is more room for variation away from the center towards higher values.

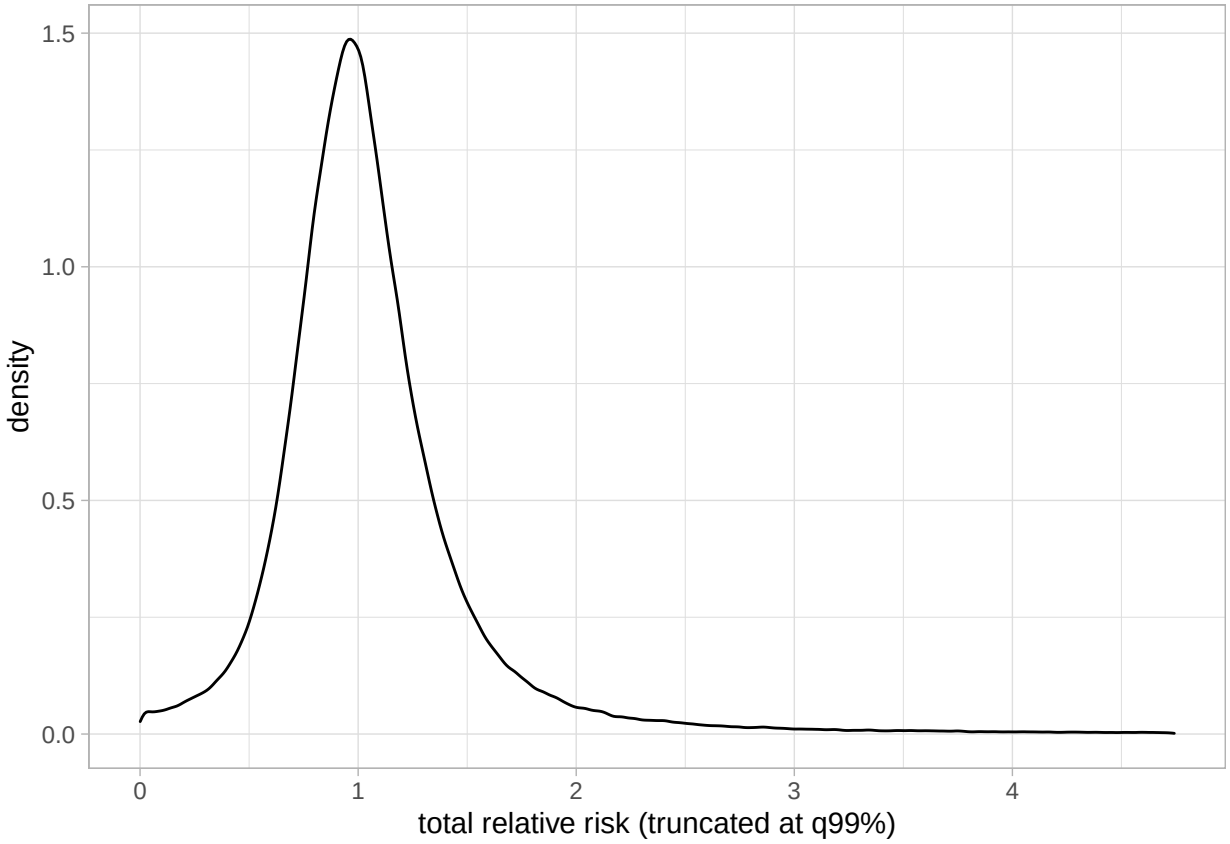

Figure 3: Density of prior predictive distribution of risk

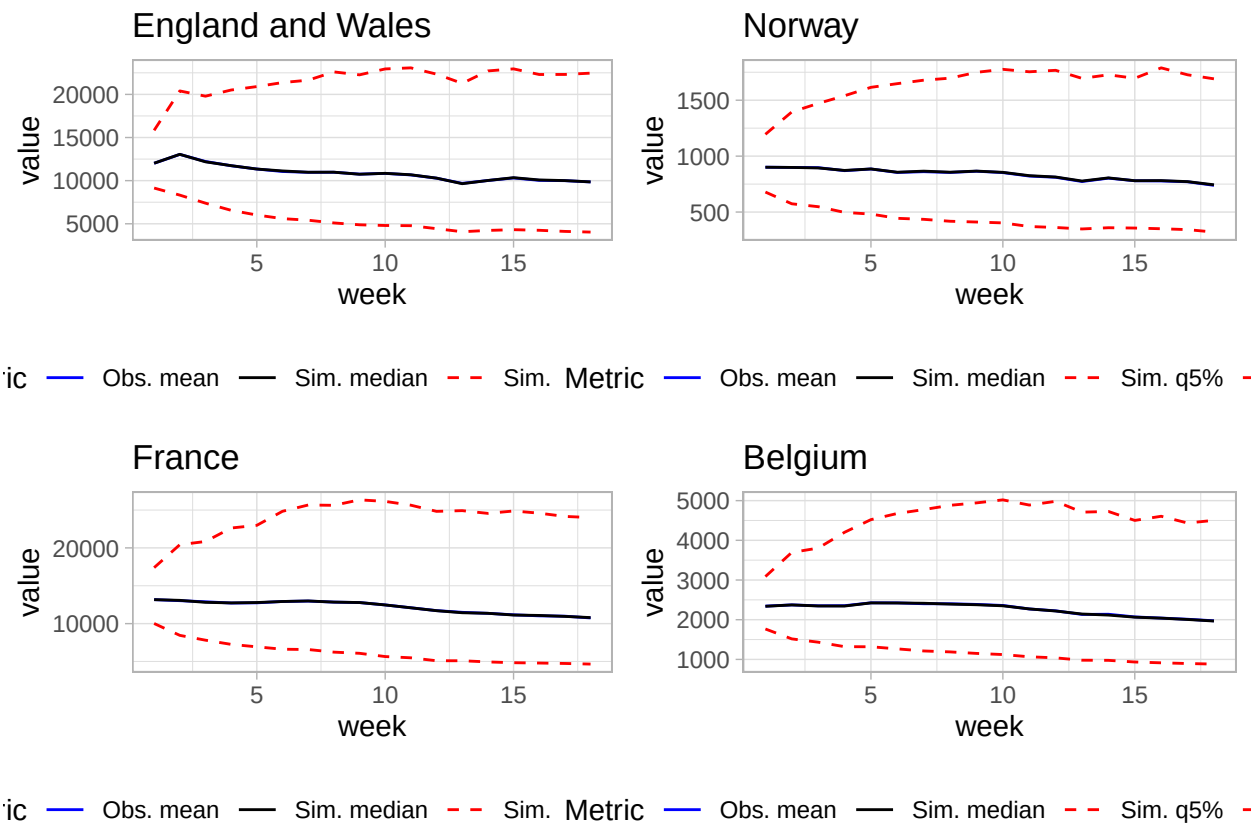

Figure 4: Prior predictive distribution of number of deaths

4 Results

4.1 Convergence

4.1.1 Excess deaths

We check the convergence of each of our four models (one per country) by checking the Rhat values for all parameters, plotting the chains traces and autorcorrations in chains. Table 2 gathers the maximum rhat value in each model. All the values are below 1.01 indicating satisfactory convergence. In Fig.5 and Fig. 6 we plot the chain traces and autocorrelation plots for four parameters in the England and Wales model as an example. We can see the chains mix well. There is a bit of autocorrelation in the plots of log_theta[16] but it is small and there is no autocorrelation for the other paramaters.

Table 2: Max Rhat in excess deaths models

Country	Max rhat
England and Wales	1.006
Norway	1.009
Belgium	1.004
France	1.001

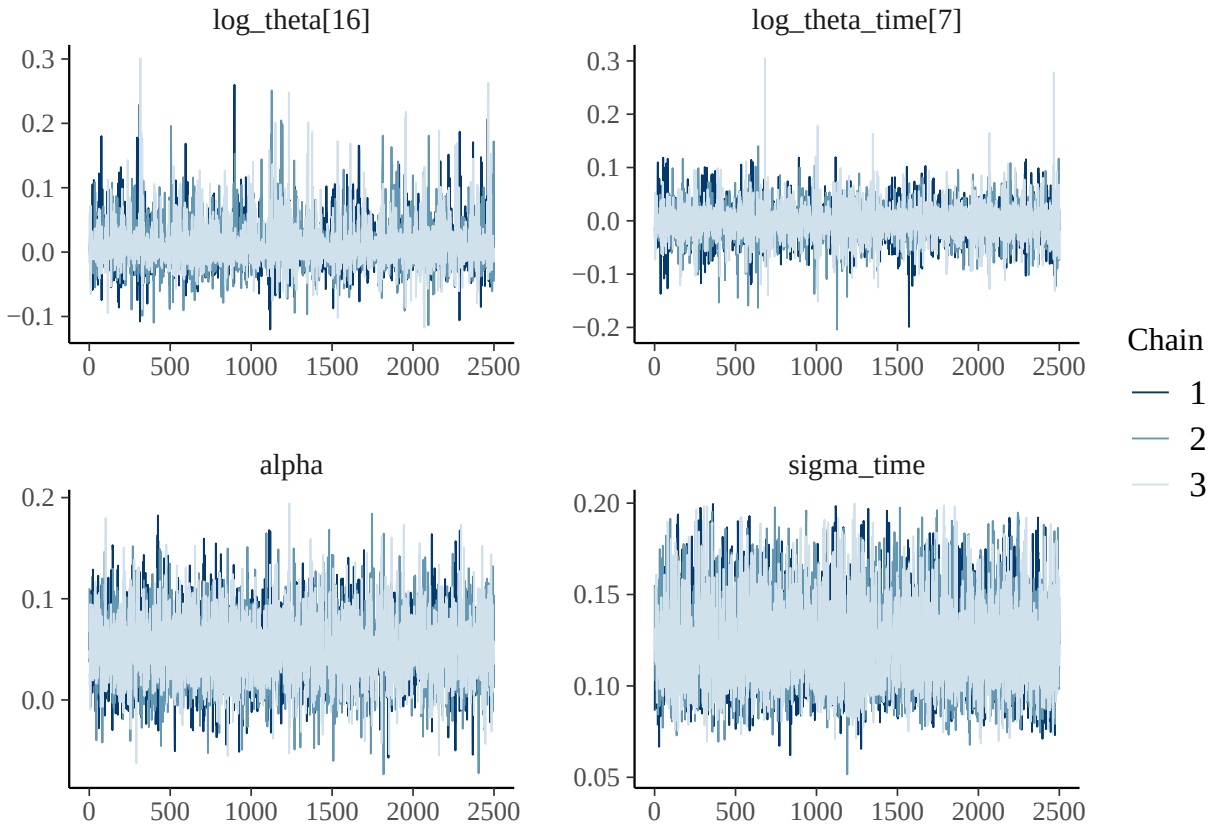

Figure 5: Chains traces for a sample of parameters from England and Wales excess deaths model

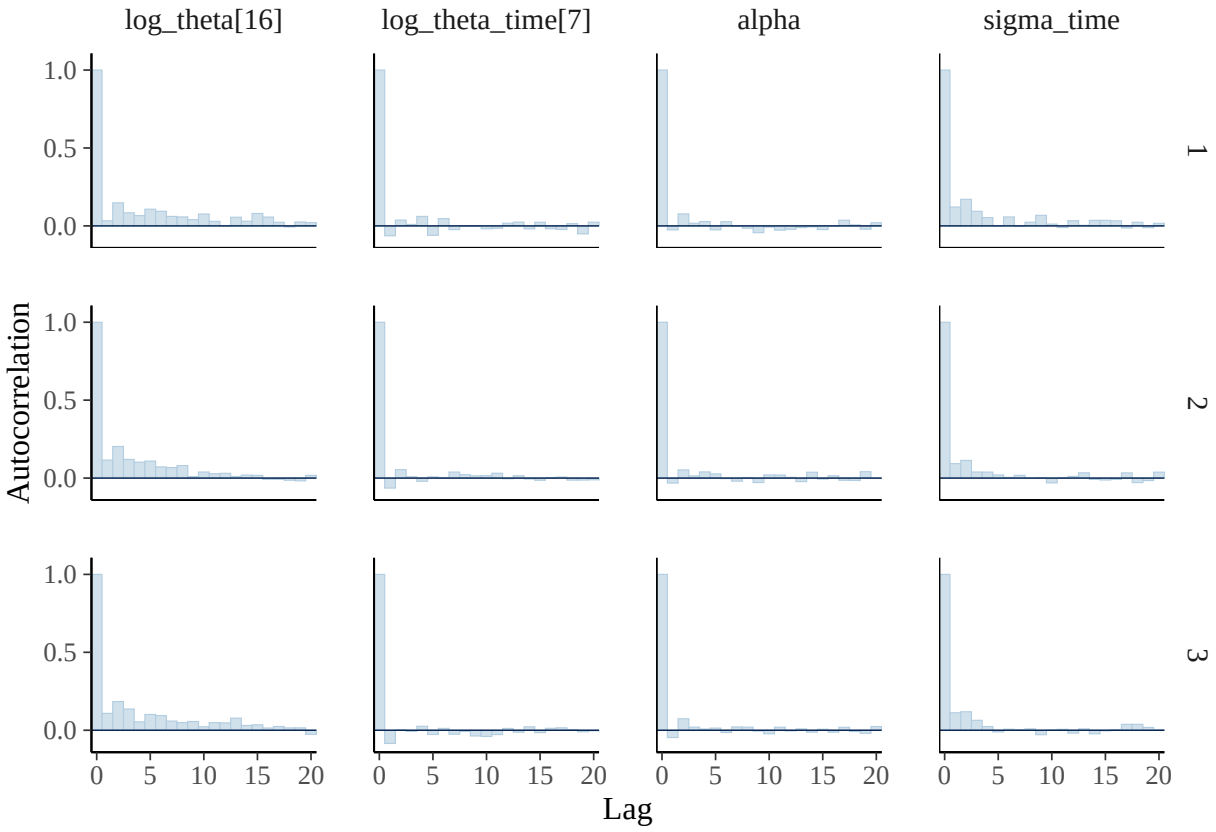

Figure 6: Autocorrlation plots for a sample of parameters from England and Wales excess deaths model

4.1.2 Excess deaths on top of the reported COVID deaths

We check again the convergence of each of our four models (one per country) by checking the Rhat values for all parameters, plotting the chains traces and autocorrations in chains. Table 3 gathers the maximum rhat value in each model. All the values are a bit larger below strictly 1.02 indicating satisfactory convergence. In Fig.7 and 8 we plot the chain traces and autocorrelation plots for four parameters in the Belgium model as an example. We can see the chains mix well. There is a bit of autocorrelation in the plots of log_theta[1] but it is small and there is no autocorrelation for the other paramaters.

Table 3: Max Rhat in models excess deaths on top of the reported COVID deaths

Country	Max rhat
England and Wales	1.004
Norway	1.004
Belgium	1.004
France	1.013

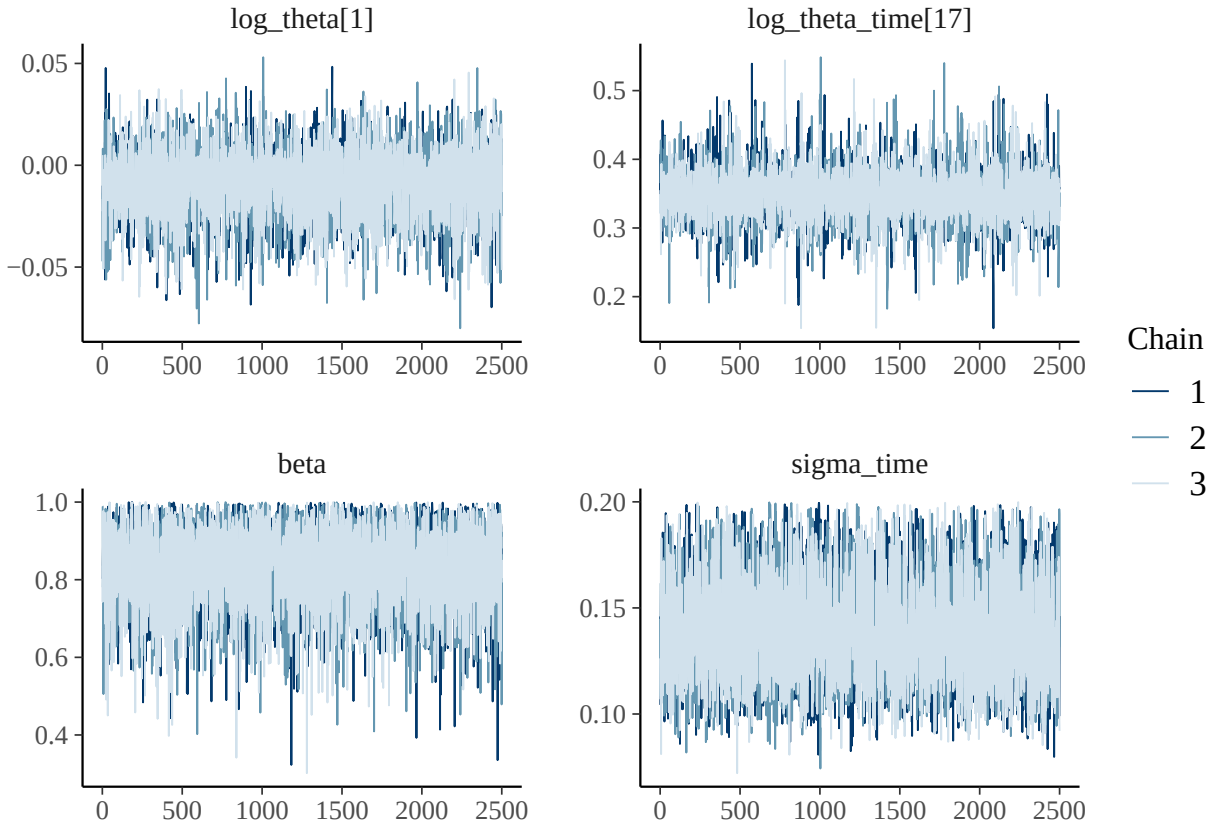

Figure 7: Chains traces for a sample of parameters from Belgium excess deaths on top of the reported COVID deaths model

4.2 Validation

To validate our model we work on the posterior predictive distribution. We have little information to compare that distribution to actual data as we have only one data per week in 2020. For each country and for each week, we compare the 2020 value to median and quantiles 5% and 95% of our posterior predictive distribution.

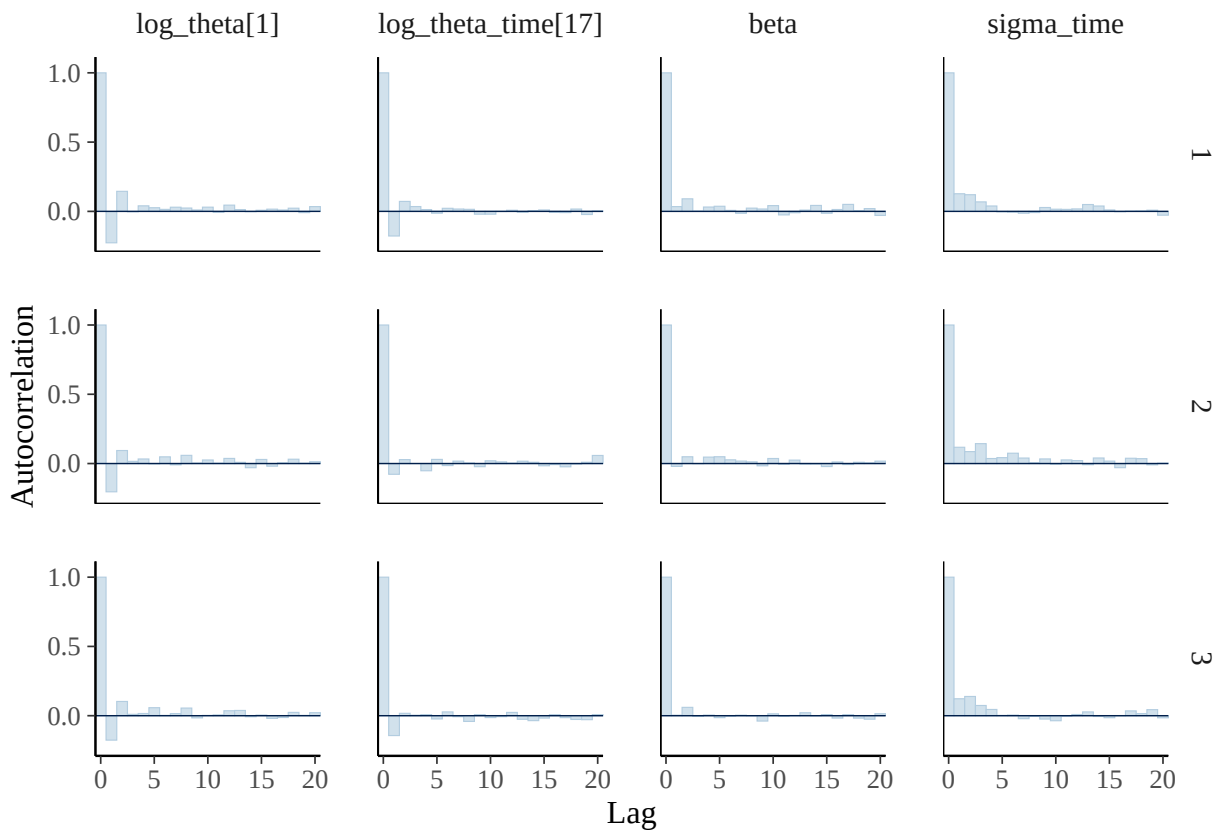

Figure 8: Autocorrlation plots for a sample of parameters from Belgium excess deaths on top of the reported COVID deaths model

In this validation procedure, a good model would be a model for which the 2020 data falls between the two quantiles for all weeks and close to the median. Aslo none of the simulations should deviate very far from the others.

4.3 Excess deaths

In Fig.9, **??**, 10, 11 we plot the 2020 deaths along with 10 simulations of posterior predictive distribution and the 5% and 95% quantiles computed on the posterior predictive for each week. We see that the 2020 deaths data falls well between the quantiles and in the mid of the simulated trajectories. In Table 4 and 5, we compare the 2020 deaths to the median of the simulations of the posterior predictive for each week. We see that for all the countries, the two values are very close.

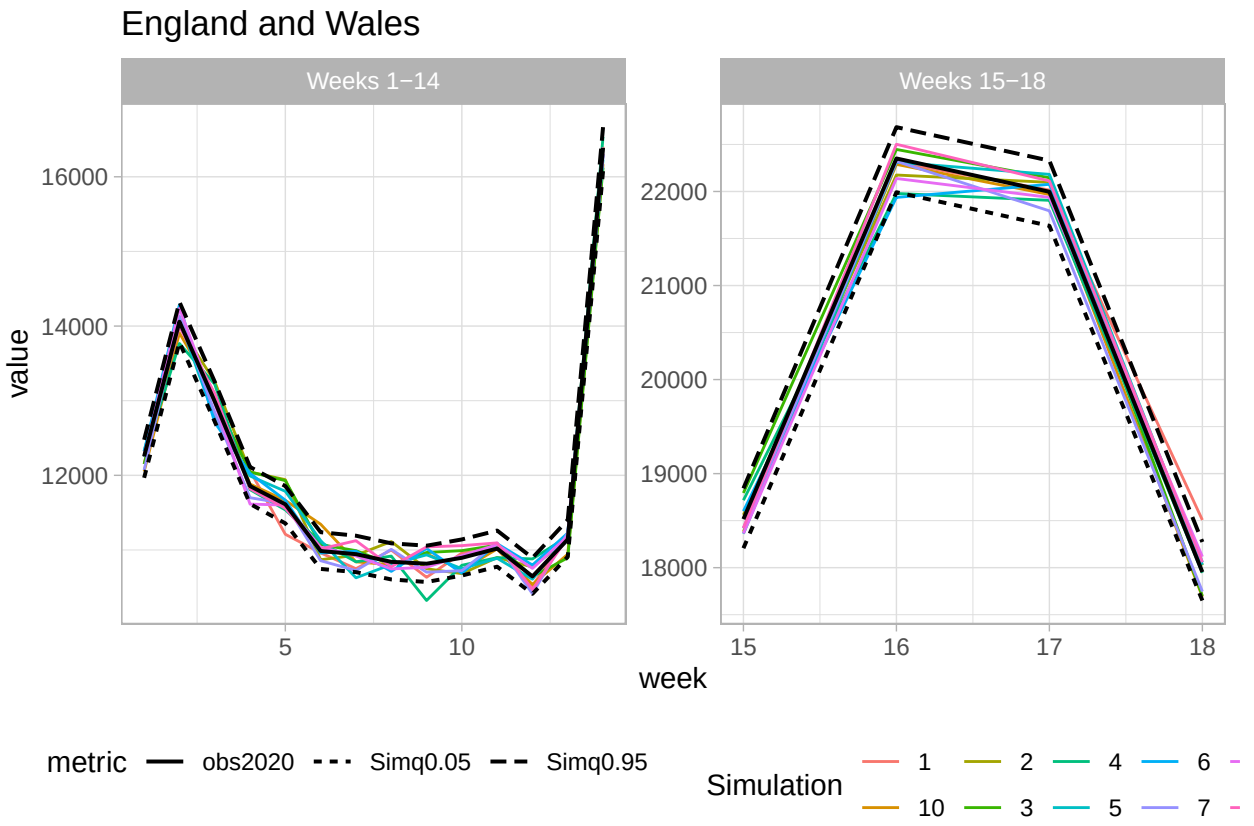

Figure 9: Validation of excess deaths model

4.3.1 Excess deaths on top of the reported COVID deaths

In Fig.13, **??**, 14, 16 we plot again the 2020 deaths along with 10 simulations of posterior predictive distribution and the 5% and 95% quantiles computed on the posterior predictive for each week. We see that the 2020 deaths data falls well between the quantiles and in the mid of the simulated trajectories. In Table 6 and 7, we compare the 2020 deaths to the median of the simulations of the posterior predictive for each week. We see that for all the countries, the two values are very close.

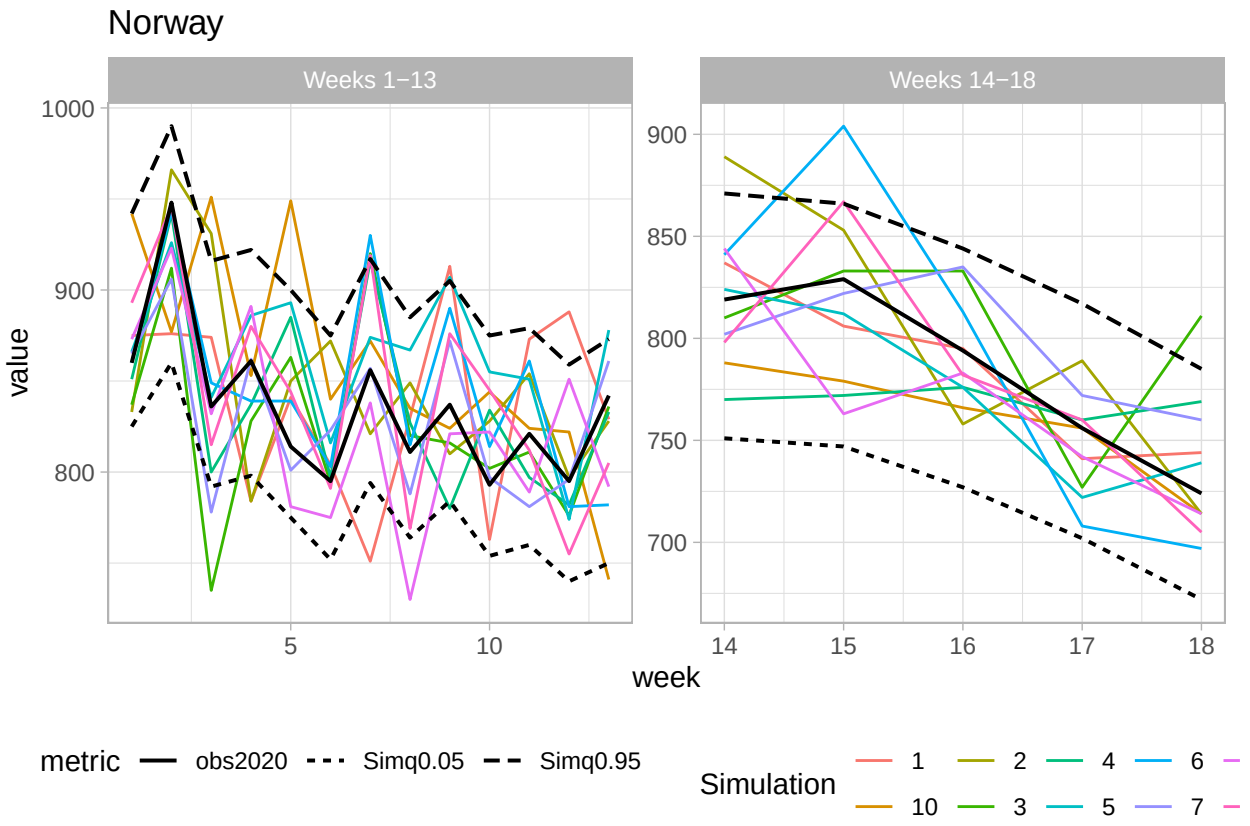

Figure 10: Validation of excess deaths model continued

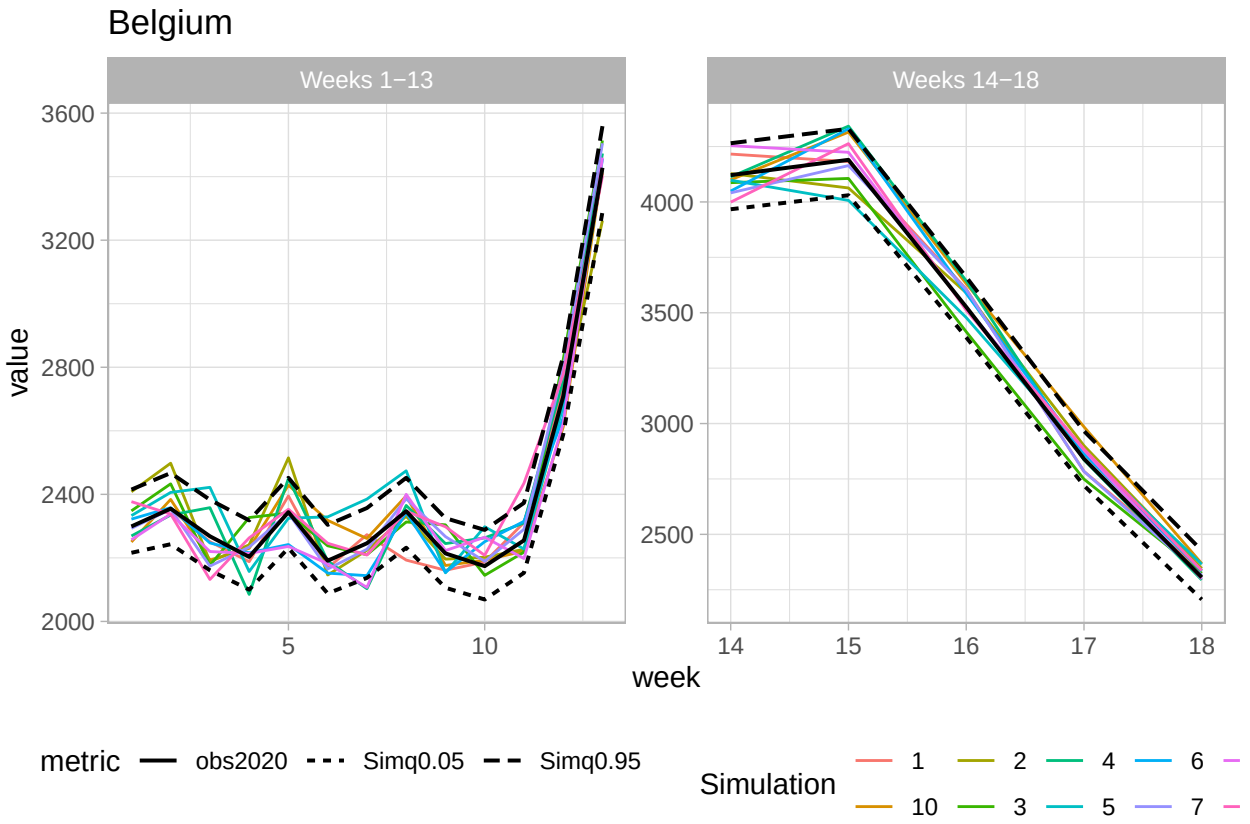

Figure 11: Validation of excess deaths model continued

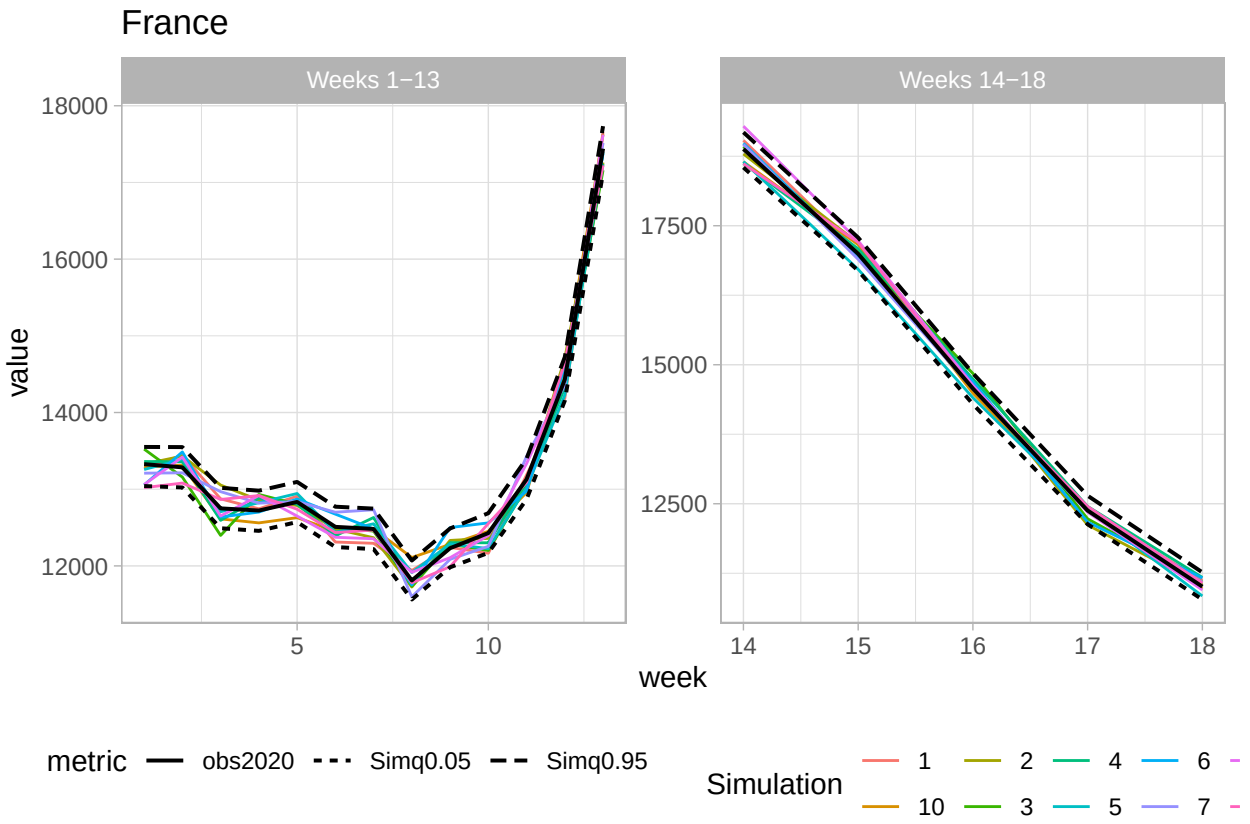

Figure 12: Validation of excess deaths model continued

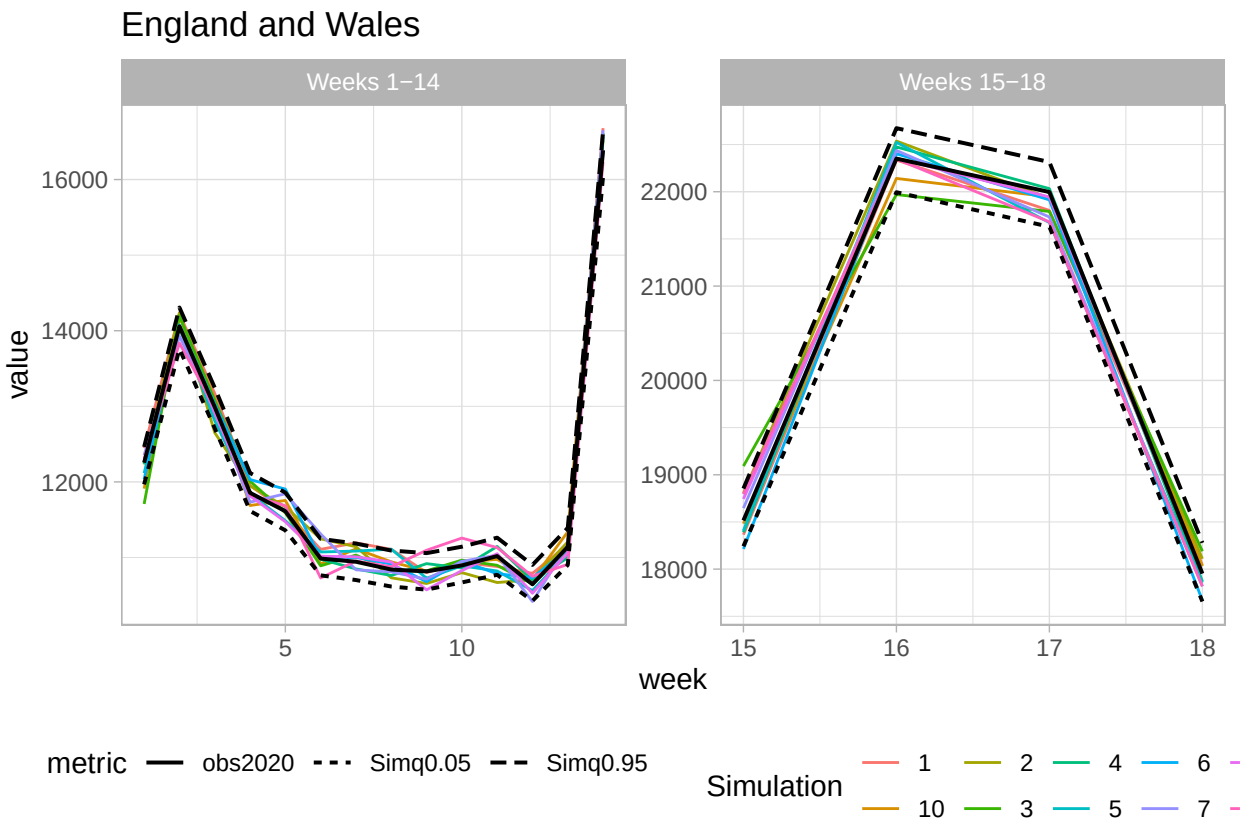

Figure 13: Validation of excess deaths on top of the reported COVID deaths model

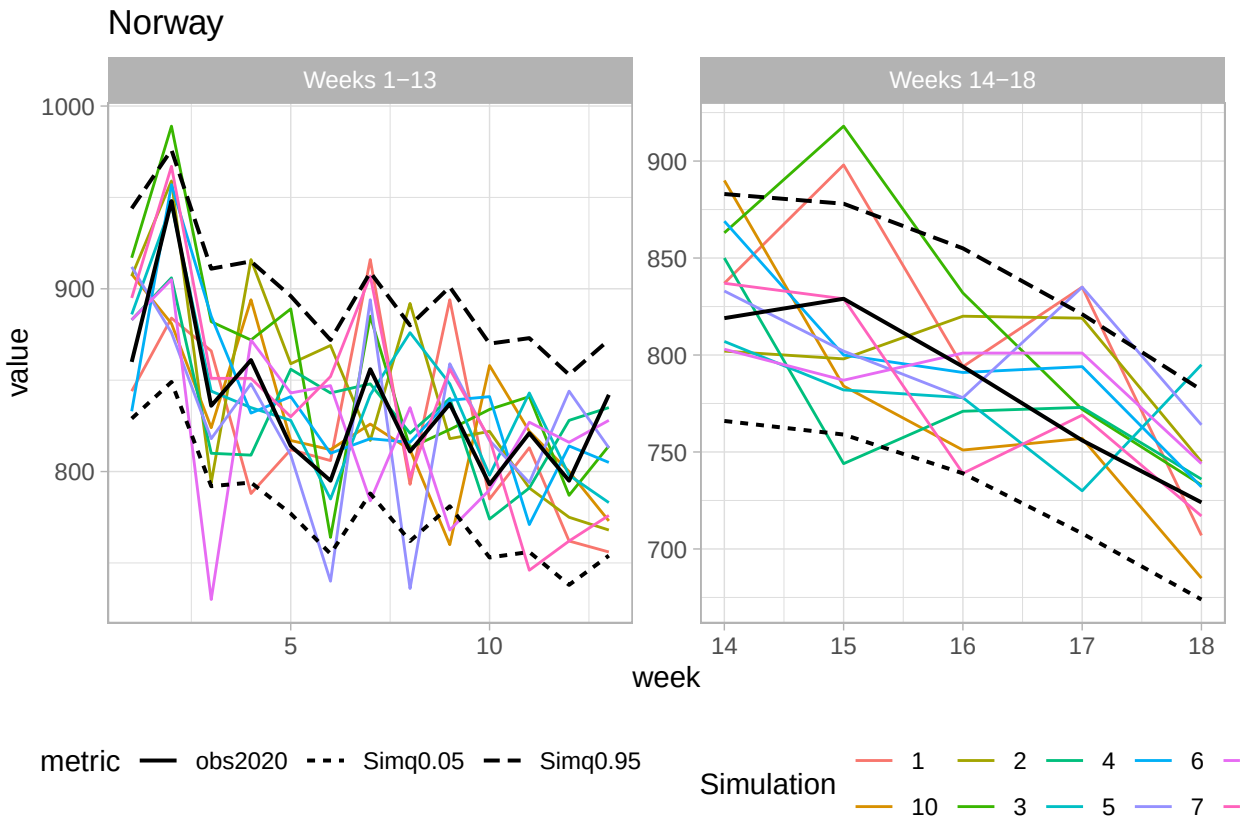

Figure 14: Validation of excess deaths on top of the reported COVID deaths model continued

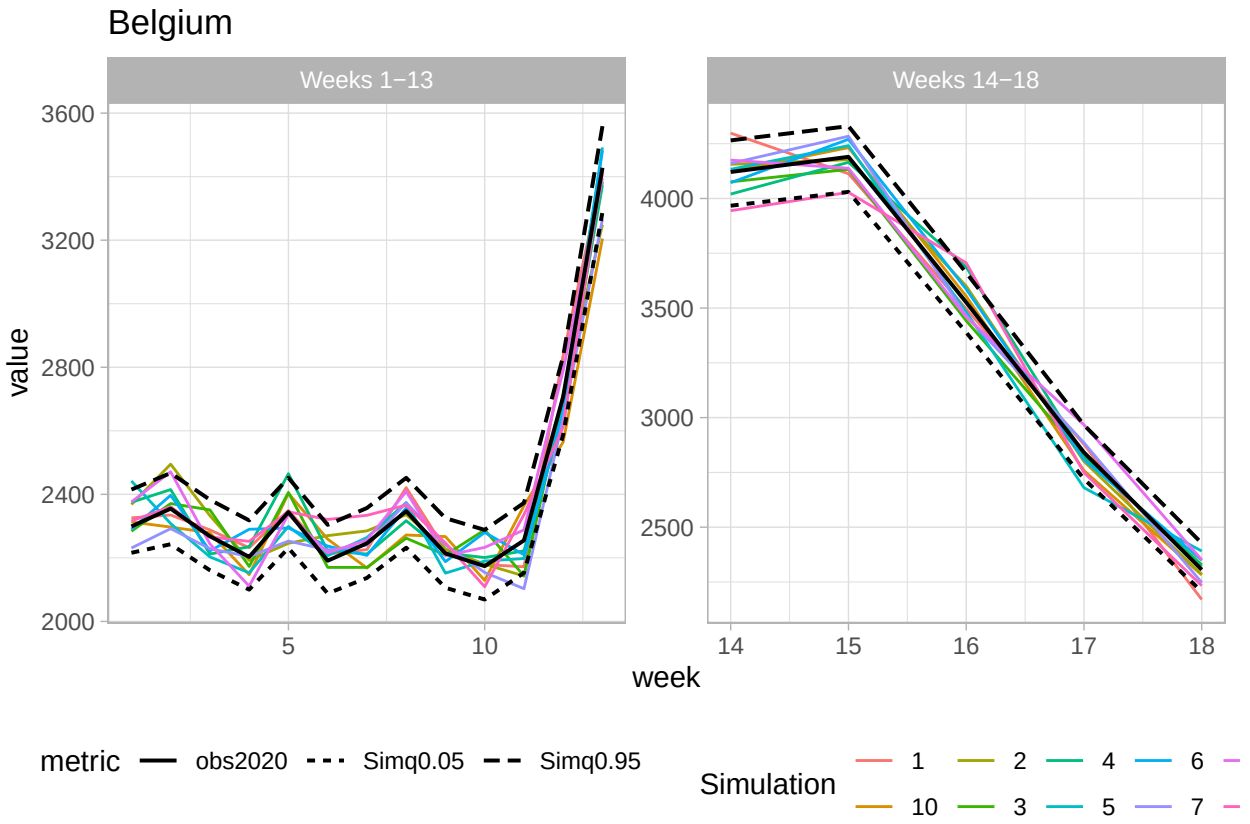

Figure 15: Validation of excess deaths on top of the reported COVID deaths model continued

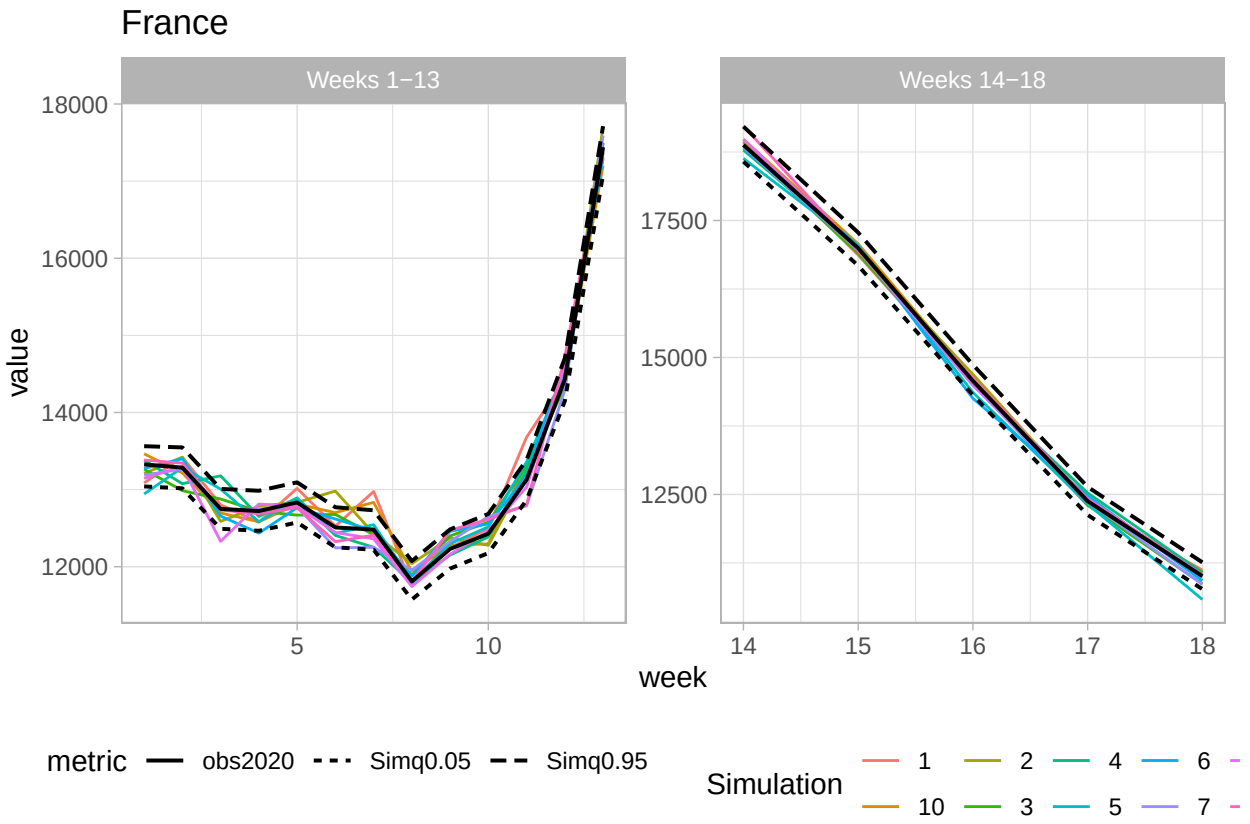

Figure 16: Validation of excess deaths on top of the reported COVID deaths modelcontinued

Table 4: Values observed in 2020 and simulations of excess deaths model

	England and Wales		Belgium	
week	Obs. 2020	Sim. Median	Obs. 2020	Sim. Median
1	12254	12220	2299	2315
2	14058	14058	2355	2353
3	12990	12986	2268	2267
4	11856	11856	2203	2206
5	11612	11611	2344	2341
6	10986	10989	2191	2195
7	10944	10941	2246	2246
8	10841	10846	2347	2341
9	10816	10815	2214	2215
10	10895	10896	2174	2178
11	11019	11013	2255	2262
12	10645	10648	2708	2711
13	11141	11152	3428	3423
14	16387	16372	4121	4115
15	18516	18521	4190	4180
16	22351	22332	3526	3525
17	21997	21979	2841	2842
18	17953	17963	2306	2316

Table 5: Values observed in 2020 and simulations of excess deaths model

	France		Norway	
week	Obs. 2020	Sim. Median	Obs. 2020	Sim. Median
1	13327	13295	860	884
2	13285	13286	948	923
3	12750	12752	836	854
4	12722	12719	861	859
5	12834	12832	814	837
6	12509	12508	795	814
7	12482	12477	856	853
8	11810	11815	811	822
9	12231	12227	837	843
10	12428	12431	793	812
11	13129	13129	821	818
12	14430	14436	795	798
13	17440	17425	842	810
14	18879	18862	819	810
15	16994	16991	829	805
16	14572	14572	794	785
17	12378	12383	756	758
18	11009	11018	724	728

Table 6: Values observed in 2020 and simulations median of excess deaths on top of the reported COVID deaths model

	England and Wales		Belgium	
week	Obs. 2020	Sim. Median	Obs. 2020	Sim. Median
1	12254	12211	2299	2315
2	14058	14036	2355	2353
3	12990	12974	2268	2267
4	11856	11872	2203	2206
5	11612	11605	2344	2341
6	10986	10997	2191	2195
7	10944	10943	2246	2246
8	10841	10849	2347	2341
9	10816	10817	2214	2215
10	10895	10903	2174	2178
11	11019	11013	2255	2262
12	10645	10663	2708	2711
13	11141	11144	3428	3423
14	16387	16339	4121	4115
15	18516	18544	4190	4180
16	22351	22337	3526	3525
17	21997	21966	2841	2842
18	17953	17965	2306	2316

4.4 Inferential results

5 Conclusion

6 Data sources

[1] Statistics Norway, https://www.ssb.no/statbank/table/07995/

[2] Statbel, https://statbel.fgov.be/fr/nouvelles/mortalite-jusquau-3-mai

[3] Institut National de Statistiques et Études Économiques, https://www.insee.fr/fr/statistiques/4487854, https://www.insee.fr/fr/information/4190491

[4] Office of National Statistics, https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales

[5] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv

Table 7: Values observed in 2020 and simulations median of excess deaths on top of the reported COVID deaths model

	France		Norway	
week	Obs. 2020	Sim. Median	Obs. 2020	Sim. Median
1	13327	13302	860	887
2	13285	13278	948	911
3	12750	12751	836	851
4	12722	12719	861	854
5	12834	12826	814	836
6	12509	12509	795	814
7	12482	12477	856	847
8	11810	11819	811	820
9	12231	12227	837	840
10	12428	12436	793	812
11	13129	13127	821	814
12	14430	14428	795	795
13	17440	17404	842	811
14	18879	18891	819	823
15	16994	16975	829	818
16	14572	14585	794	796
17	12378	12383	756	765
18	11009	11017	724	727