# Response to reviewers

We would like to thank the two reviewers for their constructive comments and questions on our work. We addressed them in the manuscript and its supplementary material and we explain our answers in the document below.

## *Review by Wayne Landis*

*Page 5. Define MCMC for this paper, not everyone is familiar with Markov Chain Monte Carlo.*
**RESPONSE:** The reference to MCMC algorithms was removed from this paragraph, and we strengthened section 2.5 with more context on the tool we use for inference, with a definition of MCMC.

*Page 8. Figure 1. The graph is difficult to read even enlarged and in color. Perhaps there needs to be a larger contrast in the gradient of deaths in the key and in the body of the graphic. I find the very subtle difference in blue to indicate the lower range of mortality difficult to see.*
**RESPONSE:** This figure's color scale was updated to use a logarithmic scale, resulting in a stronger contrast for low values. The figure was also enlarged in order to make it easier to read.

*Page 11. Figure 3 very nice. The nature of the data indicates that there may be reporting error or delay that the model smooths out. The word lockdown looks like it is beyond the printable area of the page. Similar note for Figure 3, page 13. Nice figure.*
**RESPONSE:** The placement of the labels was fixed.

*Page 14. Title for section 3.1.2. I try to be careful to not use significant in a scientific paper so that it is not confused with statistical significance. The term "significance" also has a large cultural context. I would just say important or even simply "The reduction of viral transmissibility due to the lockdown."*
**RESPONSE:** The title was changed to "Reduction of viral transmissibility due to the lockdown"

*Page 16 and 18. I found it difficult to evaluate Figure 6b and the dots. These are simulated distributions so the dots are examples that were outside the boxes? The dots do not show up in the key and should be explained. Same comment with Figure 8b.*
**RESPONSE:** We added more explanations in the caption of Figure 6b:
"Each box shows (from top to bottom) the 3rd quartile, median and 1st quartile of the distribution. The vertical line on top of each box extends up to the largest value of the sample no further from the 3rd quartile than 1.5 times the inter-quartile difference; larger values are then represented as dots and can be interpreted as possible outliers. The vertical lines below each box are constructed in an analogous way for low values."

*Page 20 Figure 10. I like the way this figure summarizes a great deal of information. The regions appear to be in alphabetical order. In my observation there are some clear patterns. The regions Auvergne.rhone.alpes, Bourgogne.franche.comte, Bretagne, centre.val.de.loire, hauts.de.france, occitanie, payes.de.la.loire, and provence.alpes.cote.dazur seem to have a similar pattern. The region ile.de.france is distinct from the other regions. Are there specific reasons why these patterns occur because of the distribution of urban areas, demographics or other factors?*

**RESPONSE:** We added explanations to the presentation of the figure: "The distributions appear to be bimodal, which is expected given the underlying two categories of $\alpha_{lockdown}$ used in the mixture model. The sizes of the modes vary depending on the region, which reveals that the two $\alpha_{lockdown}$ values fit the regions differently. The lower $R_t$ fold change fits best the regions Île de France or Corse, while the higher $R_t$ fold change fits best Hauts de France and Occitanie. " Our investigation with a linear model found that population density was the only variable that could explain this difference in $R_t$ fold change between regions.

*Page 21-22, Counterfactuals and Figure 11. Counterfactuals are a key tool to examine how well the model describes causality. It appears to me that the base and mixture model are very close in prediction. The importance of an early lockdown on the total mortality is very clear.*

*Questions.*
*What kinds of datasets would it take to be more definitive that weekends for voting did not make a difference.*

**RESPONSE:** We would like to point out that another work studied whether elections had an influence on the total number of deaths. To this end, they used admissions and deaths together with regional participation to the election and also found an absence of evidence that the elections had had a detectable impact on viral spread (Zeitoun et al., 2020). In a way, our simulation study explains why: for a single day to have a visible impact on the total number of deaths, the number of contacts on that day would need to be multiplied by a factor of 2 to be detectable. Incorporating the number of hospital admissions, which are more noisy, would probably not help (and it did not seem to help much in Zeitoun et al.'s study). However, we would like to note that several reports suggest that the elections may have contributed to the contamination of individuals in various regions:
https://www.francetvinfo.fr/sante/maladie/coronavirus/elections-municipales-a-saint-ouen-un-tiers-des-assesseurs-ont-declare-des-symptomes-evocateurs-du-coronavirus_3928917.html
https://www.ouest-france.fr/sante/virus/coronavirus/coronavirus-des-assesseurs-et-presidents-de-bureaux-de-vote-des-municipales-positifs-au-covid-19-6789229
https://www.francetvinfo.fr/sante/maladie/coronavirus/coronavirus-les-regrets-des-elus-depuis-le-premier-tour-des-elections-municipales_3887843.html
https://actu.fr/auvergne-rhone-alpes/saint-fons_69199/coronavirus-assesseurs-infectes-une-candidate-elections-municipales-porte-plainte_32674042.html

*The model is promising for other countries as well that may have regional differences. Flaxman did more countries but not to the smaller, regional scale.*

*An important result is the sensitivity of the date of the lockdown on the mortality rate. Again an important result. I assume that the team is now modeling opening up the lockdown and then reapplying it to demonstrate how long it takes the lockdown to reduce the rate of infection at that time. In my role when serving on scientific panels it seems that many scientists and decision makers are not familiar with the long lag times between implementation of a control method and seeing a change in infection and death.*

**RESPONSE:** The question of which measures to use to prevent the second wave is indeed very important. We would like to direct the reviewer to a few preprints that provide valuable insights into possible strategies:
https://www.medrxiv.org/content/10.1101/2020.06.23.20138099v2
https://www.medrxiv.org/content/10.1101/2020.05.22.20110593v1

# *Review by reviewer 2*

*1. On Section 2.1.1: As one can see from Flaxman et al, the reason why the infection-to-death is modelled as a sum of two gamma distributions is because infection-to-death is seen as infection-to-onset plus onset-to-death. However I think that this point should be stated in a clearer form.*
**RESPONSE:** The paragraph was reformulated the following way:
"In practice π is the convolution of two Gamma distributions whose parameters are obtained from the literature. That is, the infection-to-death time is modeled as the sum of two independent random times : the incubation period, and the time between onset of symptoms and death. Both time components are Gamma distributed."

*2. Section 2.1.2: in the equations (1), (3), (5) μ m should be R t 0 ,m , with t 0 being the day at which lockdown is imposed. It would read better if the authors state this clearly.*
**RESPONSE:** Notation was changed from $\mu_m$ to $R_{0,m}$

*3. Still on Section 2.1.2 (and also section 3.4): I think that, while the basic model is well introduced and one can easily refer to Flaxman's work, mixture models are not well presented. I suggest the authors to explain them in a clearer way, maybe just expanding Section 2 of the Supplementary material. I would also suggest to add some references.*
**RESPONSE:** We have improved the description of the mixture model in section 2 of the main text, by clearly separating the intended model and the way it is encoded using Stan's language. We also added a few more steps and notations to make it more explicit.

*4. Section 3.1.1, sentence "This likely results from under-reporting on week-end days, and is not handled explicitly in the model": why should hospital deaths be under-reported during week-ends? Please explain.*

**RESPONSE:** The counts provided by French public health agencies are based on the date each event was reported, and not the date it occurred, and there is always a latency between the events occuring during the treatment process (e.g hospitalization, admission in ICU, decease) and their reporting. This latency is particularly longer during the week-ends, possibly because of reduced workforce, leading to increased numbers reported on the following Monday. Source :
https://www.liberation.fr/checknews/2020/04/07/covid-19-pourquoi-les-chiffres-des-deces-et-des-hospitalisations-sont-toujours-plus-eleves-le-lundi_1784460

*5. Caption of Figure 2: it might be clearer if the authors refer to the equation for D t,m at the end of page 4 when describing that dashed lines indicate the predicted number of deaths.*
**RESPONSE:** This was fixed accordingly in the text.

*6. Section 3.1.1, period "For instance, the model estimates that in total there had been 6231 deaths in region "Ile de France" when all the data up to May 11 is used, 6502 deaths when the data stops one week before May 11, 6829 deaths when the data stops two weeks before May 11, and 5894 deaths when the data stops four weeks before May 11": initially it seems that the number of deaths increases as more data are excluded, meaning*
*deaths(all data) < deaths(last week excluded) < deaths(last 2 weeks excluded)*
*However when 4 weeks are excluded (and we are still under lockdown) this trend is broken and the number of deaths is smaller than the number of deaths using all the data. Can the authors provide a justification for this? Does this mean that in the last 4 weeks something did not work very well?*
**RESPONSE:** This is an interesting observation. However we think it needs to be evaluated in the light of two considerations. Firstly, the trend observed by the reviewer is based on 3 data points; the fact that the fourth data point, when 4 weeks are excluded, does not fit the trend may not be so unexpected. Secondly, and more importantly, when four weeks are excluded, the prefix stops right at the end of a dip in the mortality data (Monday April 13, see Fig. 2, top right panel), presumably due to under-reporting during week-ends. This dip seems particularly pronounced, we believe it is because it corresponds to the Easter week-end and Monday, a national holiday in France. It appears that the model, which does not explicitly handle under-reporting on week-ends, interprets this dip as a *bona fide* decrease in the number of deaths, which affects its prediction for the rest of the period. However, with a lower amount of data, the credibility intervals are larger, and still include the observed value.

*7. Still on Section 3.1.1, data points used in the estimation process are called prefix, but, for a better readability, I would suggest to initially introduce this term.*
**RESPONSE:** A definition for the term "prefix" was added in the first paragraph of this section.

*8. Still on Section 3.1.1, at the beginning of page 14 I do not find the text "(9750 and 7300 deaths out of 9834 and 7824, respectively) " very clear to me.*
**RESPONSE:** This text was rephrased as "[...] making errors ranging from 0.86% (9750 estimated deaths for 9834 observed in data) to 6.70% (7300 estimated for 7824 observed) per day over the month of April"

*9. Section 3.2.1, sentence "They reveal that a Rt fold change of 0.75 seems necessary for it to have a detectable impact on the number of deaths ": try to expand and explain this concept in a clearer way.*
**RESPONSE:** More explanations are now provided in this sentence:
"They reveal that a $R_t$ fold change of around 0.75 seems necessary for it to have a detectable impact on the number of deaths, because the distributions obtained with an $R_t$ fold change of 0.9 overlap largely with the distributions obtained without a fold change."

*10. Section 3.2.2, sentence "Fig. 7 shows that the resulting posterior of Rt looks very similar to the posterior obtained without accounting for behavioural changes on week-ends": indeed the two graphs look very similar, but how do numbers change?*

**RESPONSE:** A figure comparing reproduction number estimates using the base model against the week-ends model was added as Supplementary Material, section 3.

*11. Section 3.3.2: a very important result is presented, but I think it should be supported by some ideas. For instance, masks were mandatory? Was safety distance applied? How was the voter turnout?*
**RESPONSE:** Some context was added :
"The first round of voting in the municipal elections took place on Sunday March 15, just two days before the nation-wide lockdown was enacted. The voter turnout amounted to 41.6%. Following measures were enforced : safety distancing, and a maximum of three voters were allowed at once in polling stations ; hydroalcoholic gel was available in every polling station, and masks were mandatory ; voters were encouraged to bring their own pen and ballot paper which was sometimes sent by mail."

*12. Section 3.4 parameters α 1 , α 2 are elsewhere indicated as α 1 , α 2*

**RESPONSE:** This was fixed in the text.

*13. Some questions about the data:*
*- Is it possible that different regional health systems have different reliability in reporting covid-related deaths?*
*- In the discussion there is the sentence "These mortality data are incomplete, as they only include the numbers of deaths in hospitals of patients positive for the virus": I was wondering how the testing was conducted, for instance, were all deaths in hospital tested against the virus?*
**RESPONSE:** The organization of the health system in all regions of France is the same, with regional agencies collecting information coming from the hospitals. As a result, differences in reliability between regions should come from individual and not organizational origins. Although we have no reference to back this up, we believe not all deaths in hospitals were tested against the virus: some diagnostic criteria had to be met for a test to be ordered, so not all deaths were tested against the virus.

*14. I think it might be useful to show a map of France including all the regions, indicating the population density (for example as a heat map) and the estimated Rt . In addition, Ile-de-France is never mentioned to be Paris region, which I think it is important to remark.*
**RESPONSE:** A mention to Île-de-France being the region of Paris was added in the introduction. A map of France including region contours and largest cities was added as supplementary material. Heat maps for population, population density, and reproductive number estimates per region were also included as supplementary material.

*15. Discussion section, sentence "We find reproduction numbers in our results are virtually unchanged by this scaling of the IFR": I would suggest to show this in the supplemental material.*
**RESPONSE:** A figure comparing reproduction number estimates using our base IFR against a downscaled IFR was added as Supplementary Material, section 3.

*16. Check hyperlinks, since these are not always working, especially when referring to figures in the supplemental material. As well check the References section.*
**RESPONSE:** Making hyperlink references work between two distincts PDF documents does not seem possible. References to supplementary material content were modified to not attempt to create hyperlinks.