# Réponse-review COVID

June 23, 2021

## 1 Anonymous reviewer

### 1.1 Abstract

Rev: *First, your statement about the intensity of interventions, I believe, is right on track, and outbreaks of serious diseases that have been arrested before they became pandemics can result from careful intense intervention throughout the period of the disease. Also, the necessity of considering stochasticity at the beginning and at the end, when numbers are small, and considering super-spreading events is crucial in a disease model.*

Ans: Nothing to add

Rev: *A perhaps minor point, I'm not sure the term "memory effects"will be understood by general readers, and should be explained in a few words in the abstract where it first appears, I think. In fact, I was not sure of what it meant when I read the abstract, having it conjure thoughts of memory less distributions like the exponential, and other processes that are completely specified by only a state variable, such as a one-dimensional dynamical system.*

Ans: Thank you for this comment. We specified more clearly in the revised manuscript what we mean by the term "memory effects".

### 1.2 Introduction

Rev: *Another perhaps minor point, when you say in the introduction, "Most models,"that opens it up to unnecessary argument. In fact, I immediately thought that by far most of my work on epidemiological topics has been on large-scale individual-based models, all of which were intrinsically stochastic, with the stochasticity emerging for small numbers and behaving deterministically for large numbers. So if you would just say, "Many models"it might be more accurate, and would avoid pointless objections.*

Ans: This is definitely true, thank you for pointing out this exaggerated notation.

Rev: *A positive point, I think your discussion of the importance of the distribution of individual $R_0$ values is crucial and right on the point, especially for something like Covid where some large groups tend not even to believe in the existence of the disease, or in the value of any measures to contain it.*

Ans: Thank you for this kind remark.

### 1.3 The discrete Stochastic Model

Rev: *One question I have is, who is your audience? Are you aiming this at experienced mathematical modellers, or would you like to reach more general audiences, including students of epidemiology? If more general, then I would suggest a sentence explaining why the Poisson distribution is relevant here. It is the distribution of counts, so it applies here, but many aren't aware of the subtleties of different distributions, and a sentence or two could have a positive effect by helping to communicate that.*

Ans: We are indeed aiming for general public. We added a few phrases explaining the reasoning behind this model. In either way it is always good to clarify explicit the hypothesis.

Rev: *Related to audience, when I first saw $t$ introduced, I saw it with the notation $t \in \mathbb{N}$,and I thought two things, (a) that $t$ is a natural number, as indicated by the notation,and (b) hmmm, why use this mathematical notation? I think about what a great now-late mathematician told me when I was younger, "Do not put in symbols that which adequately can be explained in natural language."All of my mathematically trained students and colleagues know the meaning of those symbols, but I think few of my biologically trained students do. You do shortly thereafter say that this a number of days, but then shortly after that, in Equation 1, the first summation iterates $i$ from 0 to $t$. That leads to confusion,because zero is not a natural number. And I agree it is convenient and common to use $t = 0$ as the starting time. But then you must not say $t \in \mathbb{N}$, right? Won't that generate confusion, or make it look like you are careless with meanings?*

Ans: Indeed, we did not realise that there could be such a confusion with the notation $\mathbb{N}$. We removed the symbols in order to be more explicit.

Rev: *Regarding Equation 1, which is the starting point for the modelling discussions, let's see how well that can be understood by the general epidemiological modeller. First,$Y_t$ is given as the modelled incidence on day $t$ '- which is to say the number of new infections on day $t$.(Though incidence can also mean the proportion of the population newly infected on day $t$, or other time period, so it might be good to clarify that.) The calculation then sums over all days of the disease period, starting at time 0 and ending at the present day $t$ $\left(\sum_{i=0}^{t}\right)$. For each day within that, it sums over all new cases for that day, adding up the number of infections caused by those new cases for that day $\left(\sum_{k=1}^{Y_i} F_{k,i}\right)$. Wait, something seems wrong! You can't calculate based just on new cases that day. It must be calculated based on all individuals who are infectious on that day,mustn't it? Do you mean to say that $Y_i$ is prevalence rather than incidence?*

Ans: Actually, we do mean incidence, but it is calculated based on all individuals who are infectious that day tanks to the summation over all days of the disease period ($\left(\sum_{i=0}^{t}\right)$). The infectious status of individuals is implicitly determined by the $\omega_{t-i}$ parameter: even if we sum over all days, after 11 days, $\omega_{t-11}$ is null. We detailed the equation in the manuscript to make it clearer.

Rev: *Anyway,$F_{k;i}$is defined as the force of infection for individual $k$ first infected at time $i$, and that is multiplied by the number of individuals infected (multiplied in effect by summing over all such individuals). But force of infection is not to be multiplied by the number of infectious individuals, rather by the number of susceptible individuals who have been exposed, right? Are you referring to the infectivity ($\beta$) instead? That is the parameter that is to be multiplied by the number of infected individuals, properly prorated by the number of susceptible individuals as the infection expands through the population.*

Ans: Indeed, we meant the infectivity ($\beta$). We clarified it in the manuscript.

Rev: *I see that in Equation 2, where all individuals behave identically, you seem to switch to the term infectivity rather than force of infection. In summary of this part, and throughout your whole manuscript, I would recommend that you carefully write out what you mean by each parameter, rather than just applying terminology like incidence, force of infection, infectivity, and so forth, and also make sure that your terminology matches its common usage. It doesn't seem to here, and that will cause confusion and doubt in readers, as it does in me. If there is confusion in the literature on some terminology, then state that and explain how you are defining the terminology for your manuscript. Otherwise it will become difficult for your readers to discern what your equations mean, or they will give up and conclude that your methods are not verifiable or comprehensible.*

Ans: Thank you for your comment, we clarified the manuscript and carefully reviewed the use of those terms to make them more consistent thoughout the manuscript.

### 1.4 Computation

Rev: *A principle that can be followed to help insure correctness is not to trust any mathematics that has not been verified numerically, nor to trust any numerical procedure that doesn't have a mathematically representation in its simpler forms. Of course, this cannot always be observed, because there are some procedures that do not have simple mathematical underpinnings. Related to that principle, I see in your supplemental material that you provide differential equations further extending some of your work. Since supplemental materials are not particularly limited in size, it could also be useful to provide the actual source code, which would allow others directly to replicate and extend your work. That of course requires careful documentation of the source code, but such documentation also improves reliability of the results.*
*Earlier in the Covid pandemic, the world learned of a model used for consequential purposes that consisted of thousands of source line of undocumented C code, which was apparently not available at that time for review. Reliance on such models could back re on the idea of modelling itself, and you could help combat that by putting well-documented source code in your supplementary materials.*

Ans: We made the code available in a gitlab repository: https://gitlab.in2p3.fr/ete/origin-end-covid-19-epidemics

### 1.5 Summary

Rev: *I think you have important material to discuss concerning the stochastic nature of the onset and demise of disease outbreaks, including the current pandemic, and also about the effects of different rates of infections among different individuals or sub-populations. However, I suggest that you very carefully review your use of terminology, and explain and verify how it fits into the mathematical forms, with simplified examples as part of the explanations, to help move your preprint to the next stage.*

## 2 Bastien Boussau

### 2.1 General comments

Rev: *The manuscript was most of the time clearly written and easy to follow. However, some figures were difficult to interpret, and in some cases the description of the results seemed to include mistakes (see specific comments). In spite of these mistakes, the results appeared convincing. I could not find links to the data or the implementation of the models to reproduce the results. Finally, I believe the discussion could be extended a bit as I explain below.*

Ans: The code is available on this gitlab repository, and we corrected the manuscript to share the link to have access to the repository: https://gitlab.in2p3.fr/ete/origin-end-covid-19-epidemics.

Rev: *The reliance on several models allows for testing the influence of different factors, including super-spreaders, age structure, and memory in the time from hospitalization to death. However, these models all rely on different implementations, and differ in several respects, making their comparisons difficult. It might have been cleaner to use one framework to implement all models and compare them by changing one parameter at a time; for instance, some Bayesian models that have been proposed in the literature on SARS-CoV-2 might be amenable to such an investigation. Nonetheless, the fact that the different models agree in a lot of their predictions suggests that the results would probably have been the same, and the reliance on several implementations also protects against implementation-specific bugs.*

Ans: Thank you for your remarks. Actually, our goal here is to use the model we developed, ie the discrete stochastic model, to answer the questions. However, the other models are classically used in the literature and we wanted to compare their impacts, knowing the differences in the assumptions made behind each of them.

Rev: *Among the results that stand out is the fact that several months of lock-down are necessary to reach extinction of the epidemic. This is not unexpected, but the relevance of it to public health is little discussed in the manuscript. In two places the authors mention "an audience not familiar with stochasticity"; if this means e.g. public health officials or the general public, then more discussion should be included. In particular, I believe that the relationship between the authors' result and the feasibility of the "zero-Covid" strategy should be discussed, as a cursory reading of the manuscript may be interpreted as an argument against the strategy. Along similar lines, it seems a bit much to ask of a lock-down that it brings an epidemic to its extinction, especially when the epidemic is tackled a bit late. Would a different objective, i.e. that of reaching daily incidence levels that are compatible with a zero-covid-like strategy (control points, local lock-downs) also require several months of lock-down? Would the modeling approach proposed by the authors suffice to answer such a question, if the data are available?*

Ans: Thank you for this remark, we completed this approach with Fig4. where we display the first passage to the threshold of 20 new cases per day. We observed that Taiwan and South Korea in particular could keep the epidemic under control without strict control measures such as a lockdown for months while incidence was kept under this threshold. We further introduced and discussed those results in the main text.

### 2.2 Specific comments

Rev: *p3: "Finally, we analyse a classical deterministic Markovian model, which is commonly used to analyse COVID-19 epidemics [? ]." : missing reference*

Ans: We corrected it.

Rev: *p4: "(see Figure S6)" : this is the first reference to a figure; it would probably make sense that this is Fig. S1, not S6.*

Ans: We corrected it.

Rev: *p4: "a value much higher than the outbreak threshold above which a stochastic fade out is unlikely [10]": the number of daily deaths is not directly comparable to the outbreak threshold values provided in the reference cited. It would be convenient for the reader to detail the computations that ensure that the value chosen is much higher than the outbreak threshold.*

Ans: We detailed it in the paragraph.

Rev: *Table S1: "Shape parameter (Gamma distribution)" : in this table, could the reader be reminded that the Gamma distribution is used to model heterogeneity in infectivity and/or infection duration?*

Ans: We corrected it in the manuscript.

Rev: *Supp mat p3: "where $\eta_n$ measures the public health intervention impacts on the disease spread at day $n$,": for consistency with the stochastic model, perhaps it would be clearer to use $t$ for the day?*

Ans: We reorganized the equations to make them clearer according to reviewer 1's comments.

Rev: *Fig. S1: the legend to this figure should at least explain the meaning of the compartments, and possibly the parameters.* Ans: We added the meaning of the compartments, the parameters are explained in the main text.

Rev: *Supp mat p3-4: "We compared this model to the discrete time non-markovian model, and a SEAIRH4D model in which memory in the delay from hospitalization to death is implemented" : I find this description too short to really understand what was done, and the meaning of the acronym SEAIRH4D should be provided.* Ans: We added a paragraph in the methods detailing the SEAIRH4D model.

Rev: *Supp mat p3: "The set of ODE shown in the previous paragraph is solved using 'odeint' function from Numpy on Python 3.8.3.": Is the code for the deterministic models available? If so it could be stated here.*

Ans: Yes, it is available in the gitlab repository which we mention in the end of the methods section.

Rev: *Supp mat p3: "We estimated the following parameters for the SEAIRHD model using a maximum likelihood procedure" : could the authors provide the likelihood formula and specify what algorithm was used to maximize the likelihood?*

Ans: We detailed the likelihood computation and the algorithm. The whole code is available on the gitlab repository.

Rev: *Figure S4: "Generation time standard deviation impact on the starting date inference.": there is an inconsistency between the y axis that states "Serial interval standard deviation" and the legend.*

Ans: We corrected the y axis.

Rev: *Figure S5: I assume a serial interval of 2.3 was used? It would be useful to point it out.*

Ans: We applied the corrections in the manuscript.

Rev: *Supp mat p7: "We can see that only the importation of new infected individuals during the first days has an impact on the epidemic.": I do not understand how this conclusion is reached: is it by comparison of Figs. S4 and S5? I would need more details on the reasoning and possibly another figure to understand this.*

Ans: We detailed the paragraph to make it clearer.

Rev: *p5: "with an estimated efficacy of 1 -$\eta_{FR}$= 76% [21]." : it would be good to define $\eta_{FR}$ here rather than a few lines later.*

Ans: We corrected it in the manuscript.

Rev: *p6: "finite lock-down extensions on the the probability": too many "the"s*

Ans: OK

Rev: *p6: So $\tau$ is defined per simulation, and $p_0(t)$ is averaged over all simulations?* Ans: Yes, we made it clearer in the text.

Rev: *p6: "SEAIRHD" : This model does not include the possibility that asymptomatic individuals become recovered without ever becoming symptomatic, which is a big feature of Covid. Could the authors comment on the expected importance of the lack of such a feature?*

Ans: This is a good point. However, this should nod affect the overall dynamic if we assume that asymptomatic individuals have the same infection kinetics as symptomatic individuals. This should simply over-estimate the prevalence of symptomatic individuals, which could indeed be an issue in the end of the epidemic if asymptomatic individuals are undetected: the epidemic could be declared as cleared before it is actually the case.

Rev: *p6: "Scripts for the SEAIRHD model can be found in the supplementary materials.": I have not found them.*

Ans: Indeed this is a mistake we forgot to add the link, it is now shown in the end of the materials section.

Rev: *p7: "the same as in our model" : the same as in our DS model*

Ans: OK

Rev: *p7: "The likelihood of the deterministic SEAIRHD model was computed assuming a Poisson distribution of the daily mortality incidence data." : I think it would be good to explain how parameter inference was achieved with the non-Markovian deterministic model.*

Ans: We referred the reader to the corresponding manuscript, where the parameter inference is further detailed.

Rev: *p7: "the time mortality incidence reaches" : I think it would help to remind the reader that this date is March 23.*

Ans: OK

Rev: *p7: "67 days (equivalent to a first case on January 16 in France), with a 95% confidence interval (95% CI) between 62 and 79 days" : the numbers given in this section do not seem to match Fig. 1 "DS without heterogeneity". Was there an inversion in the names of the violin/boxplots between with and without heterogeneity?*

Ans: This is indeed an inversion in the names of the y axis. We corrected it in the manuscript.)

Rev: *p8: "However, consistently with earlier studies [21? ]" : missing reference*

Ans: We corrected it.

Rev: *p8: "the median delay for daily incidence to reach 100 deaths is decreased by 5 days when the serial interval standard deviation is decreased by one third (Fig. S4).": isn't it the opposite?*

Ans: Indeed, thank you for the remark, we corrected it in the manuscript.

Rev: *p8: "However, when assuming a more realistic scenario where all those cases are not imported on the same day, we find a much more limited impact on the delay" : I find it hard to be convinced, looking at the figures and trying to compare the two panels of Fig. S5. Could the authors provide trends or numbers, or maybe an additional supplementary figure, that would precisely convey this information?*

Ans: Indeed, the differences are not that high. We added some numbers in the main text results to be clearer about the extent of those effects.

Rev: *p9 "Time to eradication": in this section a few comments about the results of the SEAIRHD model would be useful.*

Rev: *p10: "The results are shown in Figures 3 for the case without host heterogeneity and Fig. S8 with super-spreading events." : it is not clear to me why the authors chose to show the results of the superspreading model in supplementary material and the results of the model without superspreading in main? I would have expected the reverse.*

Ans: OK

Rev: *p12: "as stressed by earlier studies [21? ]." : missing reference*

Ans: We corrected it.

Rev: *p13: "higher k parameter value that the one used here (0.30 versus 0.16 here)" : than instead of that*

Ans: OK