January 10, 2024

Re: PCOMPBIOL-D-23-01609. "Informing policy via dynamic models: Cholera in Haiti"

Dear PLOS Computational Biology editorial board,

[TODO: Edit this letter] Thank you for arranging the review of our manuscript, and for the invitation to submit a revision. We are grateful for the thoughtful comments from the referees. Our revised manuscript has various edits for clarity, following the referee suggestions, as well as some small corrections in the text. The numerical results are unchanged.

To make space for some additional sentences, we cut a few sentences that seemed least essential. We also placed some of the new material in the supplement.

Below are point-by-point responses to the referee reports. The reviewer comments are shown in green, italic type. Our responses are in black, and material copied verbatim from the revision is in blue.

Sincerely,

Jesse Wheeler
PhD Student, Statistics Department
University of Michigan
jeswheel@umich.edu

Referee 1

The authors' manuscript aims to explain how public health decisions can best be informed by combining an understanding of the underlying pathogen transmission dynamics with epidemiological data. In particular, the authors make a number of recommendations of best practices that modellers can adopt when using mechanistic models to make statistical inferences from time series data. To make their recommendations more concrete, they perform a re-analysis of cholera incidence data from Haiti spanning the years 2014-2019. These data were previously analysed in a modelling study performed by a separate team of scientists, published in Lee et al. (2020), see Ref. 1. Using refinements of three models from Lee et al., the authors demonstrate a number of ways in which the analysis could have been improved. Furthermore, the authors present differences between the findings of their reanalysis and the original analysis of Lee et al., for instance in the predicted probability of eliminating cholera from Haiti (see Figure 8).

My most substantial concern is with the framing of the paper, and whether that might mislead potential readers. To me, the paper serves more of an educational purpose, presenting ideas on how to best use mathematical models to gain insight from time series data. While the case study into cholera in Haiti is methodologically sound, it does not seem to be the main focus of the manuscript. To such an extent that, in the discussion, the only comment to on the Haiti analysis is that the authors "used the same data and models, and even much of the same code, as Lee et al. [1], and yet ended up with drastically different conclusions."

This major concern can be resolved in two ways, both of which seem acceptable to me, either a) reframe the paper as a more general series of recommendations for how to better perform statistical inference on time series data using mechanistic models (including a longer introduction and discussion of existing methodological deficiencies) or b) perform a more in-depth analysis and commentary on cholera in Haiti (some of which has already been done but is buried in supplementary information files that are only briefly mentioned in the main text, e.g. S5).

I found the methodology used to be well-suited to addressing the aims of the paper.

Overall, I found the paper interesting and informative, but a little unfocused. I expect with some heavy rewriting in places (especially the introduction and discussion), the paper will be acceptable for publication in PLoS Computational Biology. I do not anticipate the authors having to perform additional computational analysis.

Abstract

The authors state they "develop data analysis strategies leading to improved statistical fit." It's not clear to me exactly what novel strategies have been developed, it seems the authors exclusively use pre-existing techniques from the literature. This is fine, if the focus of the paper is either a) pedagogical or b) on understanding a specific disease system (cholera in Haiti). If any methods are novel, this should be more clearly stated.

Introduction

To make clear that the issues being addressed by this paper go beyond one modelling study of Haiti, the authors should provide examples and references to support the (accurate in my opinion) statement about, "common modeling decisions that may not provide an adequate statistical explanation of the data".

Given the focus of the paper on how to better perform inference, explicitly commenting on what are the "existing guidelines for creating models to inform policy [4, 5]". It would then also be helpful in the discussion to comment on how the recommendations of this study extend these pre-existing guidelines.

Methods

Fig 2. Model parameters. I think this should be "Table 1"?

"and $z \in 1 : Z$ describes hypothetical vaccination programs" It would help to give an example of the different vaccine programs here. E.g. does this mean the number of doses the individual has received, or differences in the vaccine administered?

Eq 6 and line 1: I think the notation for the asymptomatic fraction might be clearer and more consistent with the other models if you used $f_z(t)$ instead of $\vartheta_z(t)$. In fact, this is the notation used in the table in Figure 2.

Line 198: Reference for where value of v_{rate} comes from.

Usually, vaccine efficacy = 1 implies the vaccine completely blocks transmission and 0 implies no effect. In model 2 the definition of vaccine efficacy (lines 200-205 and Eq 14) implies the opposite. I suggest re-parameterising the model, or renaming the variable. For what its worth, models 1 and 3 seem to follow the usual convention.

I found the commentary on comparing fitted mechanistic models to a (statistical) benchmark model to be informative and helpful. I was wondering if the authors had any comments on whether (or not) to directly compare the AIC values of the different fitted models directly. Am I right in thinking this should be possible for models 2 and 3 given (I think) they are fitted to the same data? Model 1 is fitted to a different data set, which I understand makes such a comparison impossible.

Lines 308-309: Isn't this a log-linear trend in the transmission rate?

While the data support a linear trend at the 95% confidence level, it's worth commenting on the magnitude of the trend, e.g. the total reduction in $\beta(t)$ over the period of study.

I might have misunderstood it slightly, but I don't entirely follow the reasoning of lines 340-346. It seems that two different claims are being made: A) the model can't identify which mechanism underlies a trend in $\beta(t)$ and B) the model can't definitively state there is a trend in $\beta(t)$ compared with some other time-varying parameter (e.g. reporting probability). For model 1, I agree with the authors that claims A and B are valid. What I take issue with is the final statement, which seems to be stronger and apply to both claims and mechanistic models more generally: "we argue that a decreasing transmission rate is a plausible way to explain this, but the incidence data themselves do not provide enough information to pin down the mechanism." I agree that the "incidence data themselves" (which I take to imply "without additional covariates data") make claim A valid regardless of the model. However, I don't think claim B holds regardless of the model. By changing model 1 to include time variation in the reporting rate I don't see a priori *why the "the incidence data themselves" might not provide enough information to distinguish between time variation in transmission compared with time variation in reporting without the need for additional data. The deficiency is in model 1 rather than the information content of the data. I suggest the authors rework the paragraph to make the reasoning clearer.*

Lines 53-56: I push back against this point. The models assume an exponential distribution for vaccine-derived immunity. Assuming a mean duration of 10 years, the proportion of individuals who remain immune 9 years after vaccination is . Even for much larger durations of immunity a non-negligible fraction of the population will lose immunity after 9 years, e.g. . I therefore don't expect values for the duration of immunity around 10 years to "effectively result in the same model dynamics".

Discussion

In contrast with the introduction, which almost exclusively focuses on cholera in Haiti, the discussion only mentions it once, briefly: "We used the same data and models, and even much of the same code, as Lee et al. [1], and yet ended up with drastically different conclusions." It would help the reader to have those discrepancies summarised. It would be helpful to summarise the recommendations the authors make attempting to fit mechanistic models to time series data (even as a list). Along similar lines, it would be helpful to present a more concise summary of what the "more accurate policy evaluations" found by using the approaches outlined in this study.

underlineTypos

Eq. 1: Is there a missing star on $y_{1:N}$ on the right-hand-side?

Line 146: delete "the" before describing

Referee 2

This well-written article is a strong piece of work that will be useful to the readers of this journal and, in particular, to researchers who perform statistical inference on compartmental models of infectious diseases. Furthermore, this paper is enhanced by its transparency and reproducibility, as the provided code is well-documented and organised in an R package. Additionally, the supplementary information is a valuable resource for researchers in this field.

In a nutshell, the authors argue that existing criteria to evaluate the validity of a disease model are insufficient. Therefore, they propose more stringent standards for evaluating models' ability to fit the available data in order to obtain more reliable forecasts. The authors use a Cholera case study to outline their suggestions. To highlight, a key contribution from this work is the recommendation of employing inductive (associative) models as a goodness-of-fit benchmark, as evidenced by this sentence: "It should be universal practice to present measures of goodness of fit for published models, and mechanistic models should be compared against benchmarks". Undoubtedly, this approach provides an objective measure to judge the ability of mechanistic models to fit the data. In the following sections, I express my opinion on how this paper may be improved.

Major Concerns

1. Literature

While the arguments provided throughout the article are well-articulated, there needs to be more supporting literature at the beginning of major sections. For instance, I don't need to be convinced that the structure of a model should be based on a realistic theory about the observed phenomenon; namely, models should be a white box. However, not everyone is on board with this premise, and supporting literature that argues in favour of this approach should be mentioned. In short, more citations should be added at the beginning of each major section.

2. Limitations

In various passages of this paper, it is hinted that modellers should refrain from deterministic models and instead opt for more realistic stochastic representations. While the critique of ODE models is valid, the shift to stochastic structures is not a free choice. For example, introducing extra-demographic variability adds one additional parameter (infinitesimal variance). In ODE models, one extra parameter can lead to unidentifiability, and there's no apparent reason why this would differ in a stochastic version. Moreover, transitioning to stochastic models involves abandoning well-established MCMC algorithms in favour of methods still in development. Therefore, modellers should not assume that more realistic models with additional parameters are necessarily better without proper caution. In my experience, diagnosing unidentifiability is easier in ODE models than in POMP structures. Hence, there is a trade-off between benefits and costs.

Moreover, Monte Carlo methods, such as the Particle Filter and, by extension, Iterated Filtering, aim to approximate integrals (the posterior or filtering distributions). However, one cannot take for granted that these methods provide accurate descriptions of these targets without proper validation. In more 'traditional' MCMC methods, diagnostics like the potential scale reduction factor and effective sample size play a crucial role in the inference process. Unsatisfactory values of these diagnostics render inferences unreliable, often necessitating model reformulation. In contrast, in the literature on POMP models (including this paper), diagnostics are tangentially mentioned. Users (like me) sometimes face uncertainty about whether the lack of fit is due to model misspecification or problems with the Monte Carlo algorithm exploring the parameter space.
For example, in Figure 5, department Ouest exhibits substantial uncertainty from 2014 onwards, and this figure is on a log scale. Essentially, the inference suggests that 'anything can happen'. I would like to pinpoint the nature of this collapse in uncertainty. Identifiability issues might be at play, given the possibility of more estimated parameters than the incidence data can inform. Sometimes, we ask too much from the data. Observe the discrepancy in trends between the average behaviour and the uncertainty ribbons.
In summary, the authors should elaborate on the limitations of the proposed approach.

3. Conclusions

I find that the conclusions are somewhat disconnected from the introduction and abstract, which state that the paper presents a methodology to diagnose model misspecification, develop alternative models, and make computational improvements. It would improve readability to include a summary in this section. Specifically, link each contribution to a particular example. For instance, in the case of model 1, computational improvements increased the log-likelihood. In short, connect the findings more explicitly to the research question and stated goals.

Minor Comments:

1. In the author summary, this part is hard to follow: "and provides careful justification of valid conclusions from the fitted model. Objective measures are used to benchmark model fit; when these are combined with reproducibility, a framework emerges for continual improvement when revisiting the data and models." Please rephrase.

2. Lines 1-8. Please add more citations.

3. Line 36. Can you be explicit about what the forecasts predicted? Did the models predict a rise in cases?

4. Line 42. Add hyphen: "Model-based conclusions".

5. Lines 67-78. Add more citations.

6. Please add uncertainty intervals to the estimated parameters in Table 1. If necessary, consider splitting the table into two or including this additional information in the supplementary material. I suggest this update because there is an indication of unidentifiability when parameter estimates are fairly broad.

7. In line 159, it is stated that "$v^\star(t)$ is efficacy at time t since vaccination for adults" and then "single and double vaccine doses were modeled by changing the waning of protection; protection was modeled as equal between single and double dose until 52 weeks after vaccination, at which point the single dose becomes ineffective". I examined the reference from which this function is based but found only a numeric table. Please provide the equation of this function or a detailed description in the supplementary information. It would be beneficial for readers to understand how to model this complex feature.

8. Lines 209-211. In model 2, an incidence measurement is employed to configure a prevalence compartment. As the authors may know, initial values severely condition the dynamics of a model. Please explain this decision. Is it because that was the approach followed in the original formulation (Lee at al's paper)?

9. Lines 245-247. Same comment as before. What's the justification for assuming incidence measurements as the basis for prevalence states? What are the risks?

10. Line 262. "deterministic Model 2 is a degenerate case of a stochastic model". Please explain why or provide a reference.

11. Lines 260-274. Please add more citations.

12. Lines 343-345. This sentence is key for this paper, but it's hard to follow: "The robust statistical conclusion is that a model which allows for change fits better than one which does not—we argue that a decreasing transmission rate is a plausible way to explain this, but the incidence data themselves do not provide enough information to pin down the mechanism". Please rephrase.

13. Line 357. "Determining the necessary computational effort needed to maximize model likelihoods and acting accordingly" How do we determine the necessary computational effort?

14. Please add the predicted intervals to Fig 4. I would like to see the effect of the log-normal measurement model.

15. Lines 513-515. Please clarify the comparison between disaggregated models and a benchmark. Let's say I have spatial units 1 and 2, for which I have observations y1 and y2. Should I fit the disaggregated mechanistic model to y1 and y2 simultaneously (as usual) but keep a record of the individual log-likelihoods (log-lik y1 and log-lik y2). In parallel, fit the benchmark independently to y1 and y2, and then compare by log-likelihoods or information criteria by spatial unit.

16. Line 528. Please add hyphen: "Model-based inference".

17. Lines 576-577. "We notice that the calibrated model favors higher levels of cholera transmission than what was typically observed in the incidence data (S5 Text)". After this fragment, please summarise in one or two sentences what it will be found in S5 Text.

18. Lines 599-601. "The decision not to do this partially explains the unsuccessful forecasts of Lee et al. [1]: their Table S7 shows that the subset of their simulations which were consistent with observing zero cases in 2019 also accurately predicted the prolonged absence of detected cholera". The fragment before the colon says that Lee was unsuccessful. However, the fragment after the colon says that was in part successful. Please clarify.

19. Line 631. Since Model 1 accounts for infections at the national level, how are scenarios V1 and V2 handled?