Reviewer’s report

Title: Real-time analysis of the diphtheria outbreak in forcibly displaced Myanmar nationals in Bangladesh

Version: 0 Date: 08 Oct 2018

Reviewer: Alex Cook

Reviewer's report:

This paper describes a model that was developed and fit to incoming data from an outbreak in a refugee camp in Bangladesh. The fact that it was developed 'live' means the model is not as polished as it would be if the authors had the comfort of the outbreak being over to refine it, and the authors present this as a lessons-learned article.

Having said that, there are several parts where the authors could probably provide a bit more criticism of the approach taken during their analysis, for the benefit of future researchers in a similar situation. I highlight these below together with a few other points for consideration.

(1) I don’t see why the model structure (short as it is) needs to be placed in the appendix. The journal is online only and it would give a better sense of what was done if this could be integrated into the methods section. I can’t recall whether BMC Med has strict word counts but if so, I would implore the editor to consider relaxing it in this instance, for the benefit of the reader.

(2) Did the MCMC chains converge in your analyses? The histograms of sampled parameters look somewhat ragged (Fig S2 and subsequent ones).

(3) It’s unfortunate that the prior distribution for the proportion of cases reported clashes with the evidence from the data (Figure S2: note that there are some parameter values that have effectively zero prior weight but high posterior weight, meaning the likelihood is exceptionally high for those values). I’ve similarly messed up my prior specifications during a live outbreak, so I’m very sympathetic, but this should be flagged up quite prominently in the discussion. Given the inconsistent prior and likelihood for this parameter, the suggestion of non-convergence of the chains mentioned above, and the chance that readers will look to the paper to set an example of how to do similar modelling in the future, I wonder if the authors would consider reporting, in addition, what the predictions would have looked like if they had run the model with the full benefit of hindsight? (For instance, being able to revise priors so they do not bias the posteriors, or running chains for longer.)
(4) More details should be provided on the MCMC algorithm: a few simple things like run length, burn in, number of chains, proposal distributions, and convergence diagnostics.

(5) The authors have used a mix of non-informative prior distributions, informative priors, and (degeneratively informative!) Dirac delta priors on different parameters. I think they should justify the decisions that led to this mix a bit better in the discussion. Also, the priors do not seem to be as described in the manuscript. They say they used a log-uniform(0,1) prior for the initial infected proportion, which should take values between 1 and e=2.72, but the histograms in Figure S2 etc do not agree with this.

(6) If the predictions had been done daily, instead of approximately weekly, when would it have been clear that the epidemic had peaked? Is there a case for daily forecasts?

Minor points

L60: the acronym WASH is used here but never again. Consider removing it.

L77: similarly, EPI is never used again. To remove. (it's not even in the list of abbreviations on line 338!)

L135: similarly, SEIR is never used again.

L143: would not the presence of competition mean that the mean generation time would be less than 4.5d? (Note also that a constant hazard does not give rise to a uniform distribution for infection time.)

L154: the authors claim that two parameters cannot be jointly estimated because the two parameters are inversely correlated. This should be more carefully written. The parameters' posterior distribution is correlated, or their prior distribution? Why does inverse correlation prevent estimation? You may have a lot of information on the *joint* distribution of the parameters, but insufficient information to estimate either *individually* if non-informative priors are selected---this should not, however, deleteriously affect the posterior distribution of derived quantities such as the forecasts. In short: please elaborate on this claim to justify it better.

L180: should 'prediction intervals' actually be called 'posterior predictive intervals', seeing as they are derived from the posterior predictive distribution.
L247: GOARN is never used again: remove acronym

L267: again, not sure I agree with the claim that it is not possible to estimate all parameters. You put priors on, and you estimate them. The issue is more that with non-informative priors, there is not enough information in the dataset to estimate a few parameters well individually.

L278: I'd be a bit cautious about making the claim that deterministic models cannot deal with time varying parameters. If $\beta(t)$ is the per capita$^2$ infection rate and $\beta(t) = b_1 \exp(b_2 t)$ then you have a deterministic model that arguably has time varying 'parameters'.

L332: would accurate predictions have been available earlier if the predictions were done daily? L405: the caption uses terminology too loosely. What is meant by 'posterior range'? If it is a 95% credible interval, please call it that. If it is the entire range of the posterior distribution, please provide details of the methodology to obtain this from a finite MCMC sample.

Figure 5: consider slightly offsetting the x coordinates for the age stratified estimates so they do not overlap. Consider making one a hollow circle to help colour blind people interpret the chart.

Figure S2: The main issue with the presentation of this plot is that numbers which should be compared (i.e. those from the two areas, top and bottom rows) have different x axis ranges, prohibiting such comparisons. However, in addition, this figure has unlabeled tick marks, weird labelling on the axes (such as 0, 2000 and 6000 on one), tiny fonts, and the prior distribution for one panel has been truncated off. This figure looks incredibly slapdash to be honest! Similar comments for Figures S3 to S5).

Are the methods appropriate and well described? If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls? If not, please specify which controls are required in your comments to the authors.

Not applicable

Are the conclusions drawn adequately supported by the data shown? If not, please explain in your comments to the authors.

Yes
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I have no competing interests

Statement on potential review bias
Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?

2. Are you currently or recently affiliated at the same institution as an author of this manuscript?
If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria’. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria. However, my institution has an MOU with LSHTM, to which some of the authors are affiliated. I do not think this would bias my review.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal