Review of: "An Optimal Control for Ebola Virus Disease with a Convex Incidence Rate: Imputing from the Outbreak in Uganda"

Joe Hilton[1]

1 University of Liverpool

Potential competing interests: No potential competing interests to declare.

The authors have developed a mathematical model for the transmission of Ebola incorporating three control measures (curfew/lockdown, contact tracing, and treatment of cases), identified optimal combinations of control measures within their model, and performed cost-benefit analyses of these optimal combinations to identify policy packages best suited for the control of an Ebola outbreak. While the mathematical analysis has been performed to a good standard, I have several comments on their model formulation and I feel that with some adjustment this work could be more policy-relevant than it currently is.

My main concern is with the framing of the study, starting with the title. I would advise the authors against including the word “Imputing” in the title of the manuscript given that this has quite a specific meaning in statistics and the authors do not perform any imputation in this study. While the authors have developed their model with a broadly Ebola-like infection life history and a broadly-Ebola like parameterisation, there is no direct reference in the study to data from the 2022-23 Ugandan outbreak. I would therefore advise against making any implication that this is a modelling study specifically relating to that outbreak. If the authors do want to tie their study more explicitly to the 2022-23 outbreak they could try fitting their model to incidence data from that outbreak, which is available here: https://github.com/globaldothealth/ebola. A very close fit will be difficult to obtain and is not strictly necessary, but fitting a few key transmission parameters through MCMC or a similar method would help to make the model a bit more context-specific. This work could still be useful though even with broadly Ebola-like parameters which are not directly estimated from the 2022-23 epidemic.

In the description of the SEITRD model on page 3, the authors state three conditions which they impose on the convex incidence rate. Condition H2 requires that the incidence rate be zero at the “edges” of S-I-T-D space (i.e. whenever any of these quantities are equal to zero). I am not convinced that this is a sensible condition on an incidence function. While we would expect zero incidence when S=0 since there is no-one to infect, for S>0 incidence will only be zero if I+T+D=0, with non-zero incidence if any one of the infecting compartments is non-empty. In their model description on page 6 the authors set f(S, I, T,D) = S (ßI(1 +a0I) + ß0T(1 +a_1T) + ß1D(1 +a_2D)), which does not satisfy H2 but nevertheless seems like the correct formula.

In Table 1 the authors describe the parameter γ as the “Rate of recovered humans to become Susceptible”, with a value of 0.5336/day. This description is consistent with where the parameter appears in Equations (2.1) and (2.2).

However, the paper they cite for this choice of parameter value actually infers this number as a recovery rate, not a rate of waning immunity. When used as a rate of waning immunity, this value would suggest that immunity to Ebola lasts on average less than two days. This study suggests that the duration of immunity to Ebola actually persists on the order of months rather than days: https://www.nature.com/articles/s41586-020-03146-y. Given the short duration and small size of the 2022-23 outbreak of Ebola in Uganda, it seems very unlikely that waning immunity had any impact on the outbreak's dynamics because the epidemic was already over by the time any recovered individuals would have had their infection-derived immunity wane, and so waning immunity is probably not needed in this particular model.

More generally, I notice that the model parameters are all quoted from the literature rather than estimated directly from data from the 2022-23 outbreak. As I have stated above, this means that while the parameterised model should display broadly Ebola-like behaviour, it is not informed by the dynamics of the 2022-23 outbreak and will not be able to replicate those specific dynamics. At least some of the studies referenced in Table 1 relate to the Zaire strain of Ebolavirus rather than the Sudan strain, making them unlikely to reflect the behaviour of the Sudan strain which caused the 2022-23 outbreak. Similarly, parameters relating to treatment, contact tracing, and burial will all be specific to the outbreak under consideration since they reflect human activities which will vary over time and geographical location and so we should not expect values taken from the literature to be particularly relevant to the 2022-23 outbreak. With this in mind, the authors' model could still be a useful tool for performing general policy calculations relating to Ebola but is not a specific model of the 2022-23 Ugandan outbreak. In particular, the authors need to remove the sentence on page 19 which states that they "formulated and analyzed a deterministic nonlinear model for the transmission dynamics of Ebola virus disease premised on the 4-month 2022 Ebola outbreak in Uganda" since their model is not actually informed by the dynamics of that outbreak.

The model formulation in Equation (2.1) suggests that susceptible and recovered individuals who die due to non-Ebola mortality will still transition into the D compartment. Given the individuals in D can transmit Ebola, do the authors think this is a correct assumption? Surely it would be more reasonable to replace the \mu*Q term in Equation (2.1) with \mu*(E+I+T) or \mu*(I+T) so that only individuals who are currently infected with Ebola end up in the D compartment – otherwise individuals who are not infected with Ebola are able to transmit infection after dying.

On page 8 the authors perform an R0 calculation for their model. While this is certainly correct within the context of their model, it is important to emphasise that this is **not** an estimate of the basic reproductive ratio of the 2022-23 Ugandan outbreak since it is not based on data from that outbreak. In the Discussion and Conclusion section on page 19 the authors claim that this is an estimate of R0 for this specific outbreak, and this claim needs to be removed. The authors also claim that an R0 greater than 1 will wipe out the entire population. This is not correct; firstly the final size of an epidemic with R0>1 will be less than or equal to the total population size, not equal to it, and the final size is not the same as the total number of people to die in an epidemic unless the infection fatality ratio is 100%. Secondly, control measures are not the only thing that can reduce R0 – virulent epidemics can drive behavioural changes in the population which reduce contact/transmission rates even in the absence of control policies.

Given that Ebola has never reached endemic equilibrium and its high level of virulence means this is unlikely to ever happen, I am not sure Section 2.1.3 is particularly relevant to the rest of the study. It may improve the overall flow of the manuscript to move the endemic equilibrium calculation to an Appendix, or remove it entirely.

In Equation (3.1) the contact tracing parameter c_2 is attached to an E->I transition. Contact tracing does not act to make latent cases symptomatic and so I am not sure that this parameter should appear here. In this model formulation it might make more sense to have contact tracing cause an I->T transition, so that when an infectious individual is detected through contact tracing they enter treatment. This would make the impact of contact tracing the same as the impact of treatment in the model, although it could make sense to model the effect of increased treatment as increasing the proportion of T->R transitions relative to the proportion of T->D transitions rather than increasing the rate of I->T transitions – in this interpretation the c_3 parameter acts to make treatment more effective rather than just making treatment happen faster (which should be what contact tracing does).

It would be useful for the authors to state the method they use for solving their optimal control problem in more detail. Currently a reference is provided at the start of Section 4 on page 12, but a few sentences describing the method would help to make this manuscript more self-contained. I would also like to point out that the Data Availability statement at the end of the manuscript claims that “Data supporting this model are found in this manuscript”, but the authors have not made their code available and so the numerical analysis is not actually replicable. The study will be more replicable and thus substantially more useful to scientists and policy makers if the authors can make their code publicly available using Github or a similar service.

The authors do not state the optimal control parameters used in their numerical simulations; a table of optimal values would be very helpful since it would make it clear what kind of strategy combinations the optimal control analysis identifies as being desirable.

All numerical simulations are conducted with 20 initial cases – given only 142 cases were recorded in the 2022-23 outbreak this may be too large a “seed”. Starting with a smaller number of cases might give a closer analogue to the slow early growth of this epidemic.

The authors give no indication of what the values or units of k_1, k_2, and k_3 are in their cost-effectiveness analysis. This makes it very difficult to give any interpretation of what their estimated costs mean. It also not clear why strategy III (contact tracing and treatment) was chosen as the baseline scenario in the cost-effectiveness analysis given that in the 2022-23 outbreak local lockdowns were declared in Mubende and Kassanda districts, so that the real-life policy strategy corresponded to a situation with all optimal control parameters greater than zero. The authors also need to add a clear explanation of what is being plotted in Figure 23. It is not at all clear what the cost function being referred to here is, or why it takes on the shape it does.

The Discussion and Conclusion section as a whole is a bit vague and under-referenced. For instance, it is stated that “Contact tracing is a very laborious but rewarding exercise and involves even tracing contacts of contacts which can be time consuming and tedious” but no reference is given to justify the positive or negative claims about contact tracing given

here. Similarly, the authors claim that “All resources should be mobilized and availed to bring an end to an outbreak as soon as possible”, but the purpose of an optimal control analysis is to work out exactly which subset of resources should be applied to find the best possible outcome. This section needs to be rewritten to place the study in context with reference to other related studies and it would benefit from a more direct comparison between the optimal control strategies identified by the model and the actual policy package implemented during the 2022-23 epidemic. This is another reason for explicitly stating the optimal controls identified, since without stating the values of c_i (i=1,2,3) obtained it is impossible to know how this compares to the actual policies chosen.

Minor comments and typos

- Throughout the manuscript the authors are inconsistent in whether they abbreviate “Ebola virus disease” to “EVD” or “EBV”.
- Page 2: the sentence “Since it was first discovered in Africa was in 1976[2].” appears to be incomplete or left over from a previous draft.
- Page 2: “… in the most prone…” should be “… in the most *outbreak-prone*…”
- Page 2: Sudan ebolavirus is typically abbreviated to SUDV rather than SUVD – given the authors do not refer to this acronym anywhere else in the manuscript though, I am not sure it needs to be included at all.
- Page 2: Incomplete sentence “By 18th November 2022, 114 cases EBV has reportedly been”
- The caption of Figure 1 lacks an explanation of what the colour scheme and different arrow styles mean. The diagram in this figure would be much more useful if accompanied by an explanation, and possibly a legend.
- Page 3: In “However, an outbreak that comes back-to-back with COVID-19 pauses challenges on its control” I believe “pauses” should be “places”.
- Page 3: “…their impact have not been well described” should be “… their impact *has* not been well-described”.
- Page 3: “Hypothesise driven” should be “Hypothesis-driven”.
- In the last sentence of the first paragraph of page 3 the authors claim that mathematical models are more important to disease control efforts than actual interventions. This is an extremely contentious claim and should be removed from the paper.
- In the description of the SEITRD model the authors describe the susceptible and recovered populations as “showing” specific disease states; “representing” or “corresponding to” would be a clearer choice of words here.
- Stray asterisk in Equation (2.2) on page 6.
- Typo in caption to Figure 8: “poupulation” should be “population”.
- The variables being plotted are not specified in the captions to Figures 21 and 22.
- Stray comma after “2022” on the last line of page 19.