

Dr. Benjamin Bolker
Professor
Departments of Mathematics & Statistics and Biology

1280 King Street West
Hamilton, ON L8S 4K1 Canada
phone (905) 525-9140 x23320
fax (905) 522-0935
e-mail bolker@mcmaster.ca
December 13, 2016

To the editor,

ADD GENERAL POINTS HERE

Editor

In particular, we would encourage you to take into consideration the remarks of reviewer 1 on the presentation of your work. Furthermore, reviewer 3 raises important points regarding the limitation of the ODE equation approach. While you already discuss some of these points, it would be great if you put your study into the context of more realistic modeling approaches of HIV and other sexually transmitted diseases. It would also be insightful if you discussed to what extent your "heterogeneous" model covers age and sex structure.

Reviewer #1

The outstanding issue is that the paper remains quite challenging to read. It is just not written as clearly as it could be. In many paragraphs, the authors launch into discussion of complex points with long, multi-clause sentences, instead of explaining their motivation and results in simple, concise phrases. Too often the authors do not make it clear WHY they did something. Reading the original papers by Fraser's group on this topic (PNAS 2007 and PLoS Comp Biol 2011) provide good examples of clear writing.

Abstract: *The first few sentences of the are confusing, especially "Changes in the fitness landscape generally select for higher virulence early in an epidemic". Do you really mean to suggest the fitness landscape - which is the mapping between parasite genotype and phenotype - is dynamically changing? And do you think that the main driver for changes in virulence is simply the stage of the epidemic? If so, then it would be better to complete this sentence with a brief description of that idea and the expected trends in virulence over time. The other important factor here is that evolution of virulence is mostly likely to occur after a pathogen is introduced into a new host. It may have been optimally adapted to the original host, but differences between hosts mean it may no longer be optimal. This seems very relevant in the HIV case, as it is a zoonosis from SIV.*

Author summary: *In the first sentence, I think it would be clearer to say "infectivity" instead of "prevalence". Obviously disease prevalence changes over the course of an epidemic .. the epidemic starts with one individual being infected, and then rapidly increases. This has nothing to do with evolution - it would occur for any strain with $R_0 > 1$ as it spreads to others, even if no evolution could occur!*

Introduction: *While the intro has been improved since the last version, it is still relatively confusing. It would really be improved by explicitly and concisely explaining the goals of this study, and by addressing the other concerns below.*

Second sentence - would be much clearer to say something like "... pathogens with higher reproduction numbers - that is, the number of secondary infections caused by a single infected host over the course of its infectious period - will tend to increase in prevalence relative to strains with lower reproductive ratios." Your definition of "reproduction" was not very clear, and it's not a good idea to call R_0 a rate, as it does not have units of time.

Confusing sentence " ... in studies of discordant couples (...), HIV virulence as measured by the rate of progression

to AIDS was both heritable and covaried with the set-point viral load (...) and the probability of transmission". - You are combining results of multiple different studies here. I don't think there is a study that compared rate of progression to AIDS with probability of transmission. I think studies have a) compared SPVL to rate of progression, b) showed that SPVL was heritable, c) compared SPVL to probability of transmission. I think b) and c) required studying serodiscordant couples but I don't think most studies of a) did. The paragraphs starting at line 40 read like methods, not introduction

The term "epidemiological structures" is used throughout to mean the dynamics of the contact process (partnership dynamics), but this is not a common phrase and is not at all descriptive of what you want it to mean.

Methods: If you refer to parameters here, they must be defined here, not just in the SI (i.e. for c_u , c_w , etc)

What is the rationale behind starting at SPVL of 3? All of your results about the general time course of the mean SPVL depend on this. If it is simply to compare to Shirreff, then should state this.

Figure 1 is not well explained in the text or the caption. What is the point of these comparisons? Like where do these three values of r come from? Which one did Shirreff use? Same for other quantities.

You can't assume that your reader knows exactly what was done in Shirreff's paper. Instead of always saying what you did relative to Shirreff, the comparison between your methods and theirs should be written in a way that is clear, even if the reader has never read Shirreff.

The description of the different contact dynamics is confusing. Again, don't assume readers have read Shirreff or Champredon - clearly explain what the models are. What you name the variables (e.g. SI, II) is not important, since you don't list the equations in this section anyways. The paragraph starting at 167 is more clear, but makes no sense to have this AFTER the more detailed description of the differences between the models - should be first. Start simple, add details.

Why did you assume a constant population size - that is, replace any dead individuals immediately with new susceptible ones? In this case, the limitation on susceptible that occurs near epidemic peak is greatly dampened, and is definitely going to change the infection dynamics and probably also the virulence trajectory. Since the goal of the paper is to add more realism by including more realistic partnership dynamics, then this seems like any easy place to avoid removing realism.

Results: I would highly recommend that the figures and descriptions of the results refer mainly to SPVL, and augment with results on time to progression. It is just much simpler and more intuitive to think about SPVL. SPVL is the only quantity that is actually observed in most cases. Referring to time to progression is very confusing because LOWER time to progress = HIGHER virulence so just takes more effort to interpret.

Throughout the paper you refer to the conclusion that the simplest and most complex models give the same results, and that models of intermediate complexity give different results. But none of the figures are really presenting this point, as someone looking at the figure probably does not remember the details of each model based only on its code-word, and the series in the figures do not seem to be organized in terms of model complexity.

First paragraph: Don't understand "As r decreases from 0.084 to 0.42" . Where did these r values come from? Is r changing over time? Or over different iterations? Why not just fit to the observed r value? And this change you've described is an INCREASE, not a decrease.

Why are you changing the initial infectious density?

Line 306: Why would the speed of virulence evolution affect the maximum attained virulence? Doesn't make sense. Should just influence the time to peak.

It is very hard as a reader to maintain any interest in the results of different models when there was so little description of them, and, there is just too much detail for one to store in one's head throughout the paper. I would recommend having a diagram, or a table, that clearly compares each model, for what features it does and doesn't have, so the reader can keep referring back to this to try to make sense of the findings.

I don't understand why Figure 5 shows the unscaled parameters, when in reality you had to scale the parameters. Does not make sense to show summary statistics for scenarios that you know are unrealistic (because they would lead

to epidemics with r values that do not match the observed value)

Figure captions are not publication level. Some general advice on figure captions, taken from another source: Figures should be able to be interpreted relatively independently of the text. Most readers will go straight to the figures before reading the text in detail, and so the captions should not assume the reader has internalized every detail of the text preceding the figure citation. Figure captions should walk the reader through what experiment (simulation) was done, what analysis (statistics) was done on the data, and what quantities are shown in each axis. 100-200 words is a normal length. The first sentence (might be called the title), is often bolded, and can be either a summary of what data (or model, etc) the figure is showing, or a statement of the main results. Try to be consistent. In general, figures are supposed to describe results, not conclusions. Next, describe each panel of the figure, referring to it by label (eg A)). For each panel, or at the end of the caption (if this info is the same for all panels), describe the values of any important quantities that were held constant in during the experiment that generated the figure, state any important assumptions, and state sampling size or time if relevant. Avoid using parameter symbols or abbreviations, but if they are used, provide definitions in the caption. These conventions are often ignored in technical mathematical biology papers (i.e. those in JTB), but are crucial for making theoretical papers palatable for interdisciplinary audiences.

Discussion: *Would be helpful to start off by summarizing the main goals of the paper and then the main findings of the paper. Then, can get into limitations*

It would be very helpful to be able to give a intuitive or at least qualitative explanation for why these mixing patterns have such an influence on virulence.

Reviewer #2: *The authors have appropriately dealt with my previous comments. Furthermore, I think that the addition of a model with heterogeneity in sexual activity is useful.*

Sincerely,



Benjamin Bolker