

Dear Editor:

Thank you for the chance to transfer this submission. Below please find our responses from the previous round.

Reviewer 1

This paper considers estimation of generation interval distributions from data, paying particular attention to the questions of censoring. These are important questions, and the paper clearly lays out the theory of generation times and its implications.

My main concern about the paper is audience; for a typical Proceedings B / applied audience, there is no work with real epidemiological data (although there is real network data); but for a more theoretical audience, the new analytic results are not quite as general as would be expected.

Given that the questions considered are important and worthwhile, my suggestion is that the paper should be revised with a substantial additional piece of work added involving work with a real epidemic dataset (not necessarily novel data) to demonstrate the theoretical developments, with R code in a tidier / more user-friendly format than is currently on the associated GitHub site.

As per editor's suggestion, we have transferred our submission to Interface journal, which better suits our manuscript in its current form. Due to time constraints, we will not be able to make substantial changes to our manuscript.

Generation-interval-based approaches (Euler-Lotak equation) are still widely used in outbreak analyses to estimate the reproduction number. While some of the analytical results that we present here (e.g., egocentric kernel) may not be as general as would be expected, we expect biological intuitions (the effects of censoring and spatial reduction on generation intervals and reproduction number) that we get from our results to be fairly general.

Some minor comments:

Introduction. “Can be characterized by”. These are properties but don’t uniquely determine the epidemic.

We have followed the reviewer’s comment in changing the text:

“An epidemic can be described by the exponential growth rate, r , and the reproductive number, \mathcal{R} .”

Fig 1 -¿ two things change between each figure. I would describe the left hand as an individual realisation of the random process, and the right hand as a population-level behaviour emerging from the random process.

We have followed the reviewer’s comment in changing the figure caption:

“(Left) an individual-level kernel of an infected individual with latent period of 11.4 days followed by infectious period of 5 days; this represents an individual realization of a random process. (Right) a population-level kernel of infected individuals with latent and infectious periods exponentially distributed with means of 11.4 and 5 days, respectively; this represents a population average of a random process.”

Eq(15) -¿ typo on last line of the equations.

Fixed.

Reviewer 2

The manuscript addresses very important questions. Some comments:

The same notation, e.g. (3) is used for equation three and reference 3 (and also some effects mention in beginning of Sec 5). Use different notation.

Fixed.

p5, eq 3: perhaps $\mathcal{R} \rightarrow \mathcal{R}(t)$?

We made minor changes to this equation for clarity:

$$i(t) = S(t) \int K(s)i(t-s)ds = \mathcal{R}_0 S(t) \int g(s)i(t-s)ds.$$

p5, eq 5: I think $S(t)$ should be removed as it is contained in R ?

This was a typo. We changed \mathcal{R} to \mathcal{R}_0 :

$$i_{t-\tau}(t) = \mathcal{R}_0 i(t-\tau)g(\tau)S(t)$$

p5-6: perhaps mention that t, s is for calendar time and τ time relative to/from infection.

We added the following sentence in the beginning of section 2 to clarify this:

“Hereafter, we use t, s to represent calendar time and τ to represent time relative to/from infection.”

p6, top: when you say and write “proportional” there are 3 time parameters. Be explicit and write: with respect to τ

Fixed.

p8, top and later: I have a problem with $N(a)$. Does your model allow for different number of acquaintances? But then your \mathcal{R}_0 is no longer valid. Also, you are always assuming neighbours/acquaintances are chosen at random - please stress this more.

We added the following sentences in the beginning of section 3 to make this clear:

“To explore the effect of multiple contacts on realized generation intervals, we no longer assume that the population is homogeneous (and therefore do not rely on the classical definition of the basic reproductive number). Instead, we assume that a disease spreads on a network; infected individuals contact

their “acquaintances” at random, but “acquaintanceships” are predetermined by the network structure before the beginning of an epidemic.”

p8, bottom: I did not follow how you got \hat{g} . What is a, $N(a)$ and $f(a)$? Explain more.

We tried to clarify this.

“... number of acquaintances $N(a)$...”

“The population-level egocentric kernel is found by integrating the individual-level egocentric over individual variations:

$$\hat{K}(\tau) = \int \hat{k}(\tau; a) f(a) da, \quad (1)$$

where $f(a)$ represents a probability density over a (possibly multi-dimensional) aspect space. Essentially, the population-level egocentric kernel further accounts for the probability that a susceptible individual has not been infected within an infected-susceptible pair. Trapman *et al.* used this same kernel (also assuming a constant per-pair contact rate) to study the effect of network structure on the estimate of reproductive number. The population-level egocentric generation-interval distribution is:

$$\hat{g}(\tau) = \frac{\hat{K}(\tau)}{\int \hat{K}(s) ds}. \quad (2)$$

The population-level egocentric generation-interval distribution describes the distribution of times at which secondary infections are realized from an *average* infected-susceptible pair ...”

p9, bottom: what is a homogeneous network? Complete network, alla have same degree or ...? Explain better why your networks of size 5 matter in a large community (local aspects?)

We tried to clarify this.

“ Simulations on a small homogeneous network (i.e., complete network) confirm this additional effect (Fig. 3, right panel). We can think of simulations

on this network as an approximation of (local) infection process in a small household, consisting of 5 individuals; we expect realistic local network structures (and their effects on the realized generation intervals) to lay between a star network and a complete network...”

Sec 5: some of the 4 effects described also affect R whereas others don't. Be more clear to the reader. For instance, the ordinary Euler-Lotka equations is not fulfilled in network situation. I think Ball, Pellis and Trapman have investigated different reproduction numbers for network epidemics and related.

All 4 effects affect \mathcal{R} . Our simulations show that Euler-Lotka equation can work even in network situation, as long as we can find the proper generation-interval distribution (by using temporal correction early in an epidemic). We tried to make this point clear.

“When generation intervals are sampled through contact tracing, there will be four effects present in the sample: (1) right-censoring effect, (2) egocentric depletion effect, (3) local depletion effect and (4) global depletion effect. We can correct explicitly for the egocentric effect and, in the case of exponential growth, the right-censoring effect; we showed that these effects shorten the mean observed generation intervals, which in turn will reduce the estimate of the reproduction number. While the other two effects are difficult to measure, we can make qualitative predictions about their effects on the realized generation intervals and reproductive numbers: both local and global depletion effects also reduce number of infections that occur and shorten generation intervals.”

“When temporal correction is performed early in an outbreak, during the exponential growth phase, the effective distribution should incorporate egocentric and local spatial effects but not the global effects; we expect this distribution to correctly link r and \mathcal{R} through the Euler-Lotka equation.”

In your analysis you assume that you observe the sampled contact traced generation times without error. It is known (e.g. your ref 18) that onset of symptoms rather than infection time are observed, and that often there are several potential infectors. You don't have to include this in your analysis, but at least it should be acknowledged.

We added the following sentences:

“For brevity, we assume that *all* generation intervals (no under-reporting) are observed without error. In practice, it is difficult to measure generation intervals precisely because (1) infection events are often unobserved and (2) there may be multiple potential infectors; these factors can introduce biases to the estimates of the reproduction number. We do not pursue this direction in this study.”

p14, sec 7.2: the S_i should vary over time. When is it measured in the algorithm?

We added the following sentence:

“... (and therefore decreasing S_i by 1 for all individuals i that are acquaintances of the newly infected individual) ...”

p17, middle. Suddenly you talk about under-reporting. This is another very important area which would affect your analysis in many ways. Mention that you neglect this elsewhere.

We added the following sentence:

“Here, we assume that *all* generation intervals (no under-reporting) are observed without error”