Supplementary Material: Analysis of Individual-level Epidemic Data: Study of 2018-2020 Ebola Outbreak in Democratic Republic of the Congo

Harley Vossler[1], Pierre Akilimali[2], Yuhan Pan[1], Wasiur R. KhudaBukhsh[3], Eben Kenah[1], and Grzegorz A. Rempała[1,*]

[1] *College of Public Health, The Ohio State University*
[2] *College of Public Health, University of Kinshasa*
[3] *School of Mathematical Sciences, University of Nottingham*

November 26, 2021

A Additional Figures

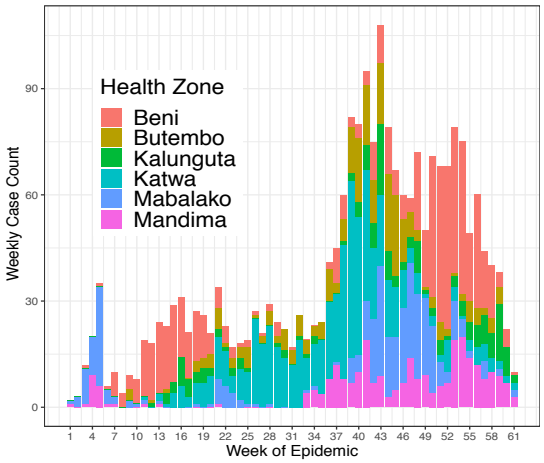

Figure 5: Proportion of cases, aggregated by week, for each of the six most outbreak-affected health zones: Beni, Butembo, Kalunguta, Katwa, Mabalako, Mandima (top to bottom in legend).

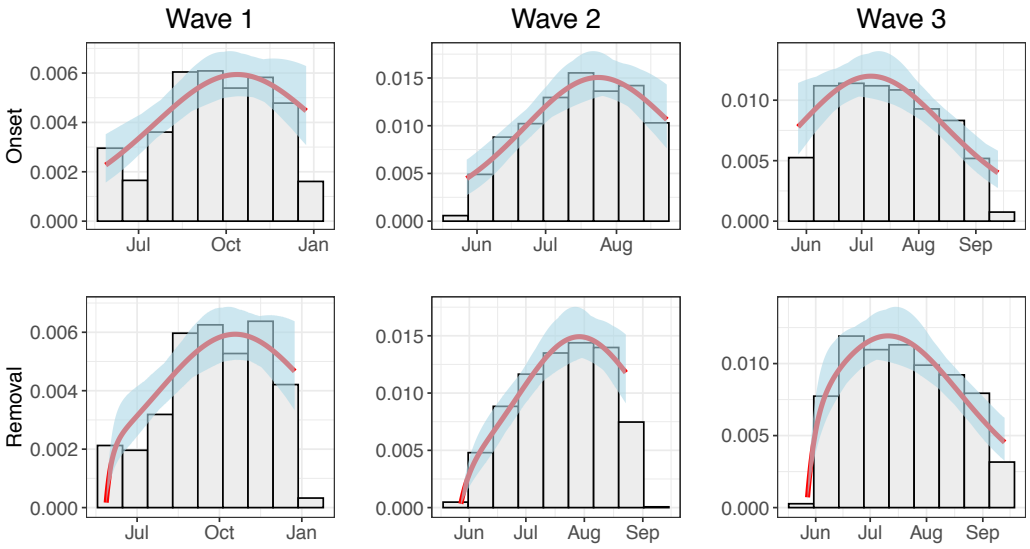

Figure 6: **Daily Incidence and Removal Rates** The model fit versus actual daily data on cases incidence (onset) and removal. As described in Section 2.3, the analysis was conducted for each wave separately, leading to different sets of parameter values. To facilitate the comparison with Figure 3, the results are plotted as relative counts (density functions).

B Derivations for DSA Model

The idea of dynamic survival analysis (DSA) is to derive individual-level independent Markov processes such that their aggregation leads to the desired ecological models [10]. In case of the SIR model, equation (1) describes for each individual member of the population their probabilities of being in the susceptible (S), infected (I), or recovered (R) state. This can is done by interpreting the rescaled proportions s_t, ι_t, and r_t as the respective state probabilities whose evolution is described by the stochastic *master equation* [24] given by equation (1) with a random initial condition. This DSA interpretation of the SIR model leads to the derivation of the individual-level likelihood function as follows.

Likelihood. We assume that observations of new infections and recoveries are available up to some time horizon T such that $T \in [0, \infty]$. Denote

$$\tau_T = 1 - s_T. \tag{5}$$

Note that τ_T is not decreasing in T and that $\tau_\infty = \tau < 1$, the final epidemic size. Since from the first and the third equation in (1) one obtains, for $t \in [0, T]$,

$$s_t = e^{-\beta \int_0^t \iota_u du} = e^{-\mathcal{R}_0 r_t},$$

The quantity $(s_t + \tau_T - 1)/\tau_T$ may be readily interpreted as the statistical survival function [29] on $[0, T]$. It follows that $-\dot{s}_t/\tau_T$ may be interpreted as a density function of infection times on $[0, T]$ for any $T \leq \infty$. Since the DSA Markov assumption implies that the infectious period is exponentially distributed with

rate parameter γ, we obtain the following individual-level likelihood function for an initially susceptible individual i observed until time T with infection and recovery times t_i and T_i, respectively:

$$\mathcal{L}_i(\theta|t_i, T_i, T) = (-\tau_T)^{-1} \dot{s}_{t_i} \gamma^{w_i} e^{-\gamma(T_i \wedge T - t_i)}.$$

Here, w_i is the event indicator satisfying $w_i = 0$ if $T_i \wedge T = T$ and $w_i = 1$ otherwise. The likelihood for the set of n individuals in the population (3) is simply the product of the individual likelihoods, reflecting the assumption that the infection events are approximately independent in a large population.

Effective population size and outbreak size. Since only infections and recoveries are recorded are recorded during a typical epidemic, it is often difficult to determine the size (N) of the susceptible population at risk of infection. This is known in the literature as the problem of estimating N the *effective population size* [10]. Under the DSA interpretation of the SIR model, this estimate may be obtained as

$$\hat{N} = \frac{k_T}{\tau_T}, \tag{6}$$

where k_T is the count of observed infected over the time horizon T and τ_T is given by (5). Similarly, one may also estimate the final epidemic count K_∞ of all already observed and future infections by

$$\hat{K}_\infty = \hat{N}\tau, \tag{7}$$

where $\tau = \lim_{T\to\infty} \tau_T$ is the final epidemic size.