Supplement 3 for Effect of Ebola Disease Progression on Transmission and Control in Liberia

Data analysis

Our data on Montserrado County, Liberia, was obtained from the Ministry of Health and Social Welfare (MoHSW). They included aggregated cumulative cases (13) as well as contact tracing data (Supplement 1). The data differentiates between three Ebola diagnostics: ‘confirmed’, ‘probable’, and ‘suspected’ cases. Cases are considered ‘confirmed’ if they were laboratory confirmed with a PCR test, ‘probable’ if Ebola-specific symptoms were diagnosed clinically, ‘suspected’ when there has been known contact with a ‘probable’ or ‘confirmed’ case, and in the presence of sudden high fever or three clinical symptoms of Ebola (32). Generally, these data are provided on a daily basis on weekdays, but not on weekends. The data also indicate the number of newly reported Ebola-related deaths. The contact tracing data are provided on a daily basis at an individual level that includes the date on which a case was reported as well as the dates of fatalities. For our modeling analyses, we considered any of the three categories to be a case of Ebola because under-reporting is likely. In addition, Pearson correlation between the cumulative number of ‘probable’, ‘confirmed’ or ‘suspected’ cases observed in both data sources was at least 97.5%. To determine the number of cases on a daily basis, we applied a linear interpolation to the cumulative incidence.

Primary contact tracing data in Montserrado County was provided by the Liberian Ministry of Health and Social Welfare between August 7th and August 26th. The data specifies the contacts of 246 infected individuals, among whom 113 were dead, traced since symptom onset. Over this timeframe, for each infected individual $i$ we used the number of contacts, $C_i$, as provided in the data, excluding one report of a person who declared 72 contacts, which is more than 10 standard deviations higher than the average. This contact data was also used to construct a probability distribution of the contacts per infected individual, $f(C)$, (Supplement 2). Outside To extrapolate beyond the timeframe of the contact-tracing data, we sampled the contact of infected individuals from this distribution, $C_i \sim f(C)$. For non-survivors during the late phase of infection, the number of contacts was drawn from
the distribution, but conservatively restricted between one and $\min(C_i, 5)$, as we assumed that during the late phase individuals are only in contact with household members (See Methods, Main text). See Supplement 2.

**Estimation of the rate ratio of transmission risk given contact**

To evaluate the daily infectiousness of an infected individual given contact assumptions, we integrated daily measures of VL through disease progression. The VL estimates used were based on daily RNA copy levels detected among survivors and non-survivors per day of infection during the 2000-2001 Ebola outbreak in Uganda (8). Consistent with previous studies (16,17), we assumed that for any given contact, each ten-fold increase in VL leads to an $r$-fold increase in infectiousness, i.e., $g(V^s(t)) = r \log(V^s(t))$. We evaluated values of $r$ based on Dowell et al. (9) who assessed the relative risk of becoming infected given three types of common contacts: 1) conversation, 2) sharing a meal, and 3) sharing a bed. This study on the 1995 DRC Ebola outbreak surveyed 173 individuals who had contacts with 27 Ebola patients, and specifies the relative risk of infection and 95% confidence intervals given contacts with Ebola patients in their early phase and late phase of infection (9).

For a specific contact, the ratio between the relative risk of infection in the early phase and the late phase is given by:

$$\frac{RR_{Late\ phase}}{RR_{Early\ phase}} = \frac{\frac{1}{r(t_2-t_1-1)\sum_{t=t_1+1}^{t_2} \log(V(t))}}{\frac{1}{r(t_1-t_0)\sum_{t=t_0}^{t_1} \log(V(t))}}$$

where $RR$ represents the relative risk of infection, $\tau_0$ represents the first day of symptom onset, $\tau_1$ the last day of the early phase, and $\tau_2$ the last day of the late phase. Rearranging,

$$r = \left( \frac{RR_{Late\ phase}}{RR_{Early\ phase}} \right) \left( \frac{1}{(t_2-t_1-1)\sum_{t=t_1+1}^{t_2} \log(V(t))} - \frac{1}{(t_1-t_0)\sum_{t=t_0}^{t_1} \log(V(t))} \right)^{-1}.$$
For each type of contact, we sampled values for $RR_{Late\ Phase}/RR_{Early\ Phase}$ (9), and used Eq. (2) to estimate the distribution of $r$ -fold increase in infectiousness for each ten-fold increase in VL (Supplement 4). In our analysis, we employed only samples that satisfied $RR_{Late\ Phase} > RR_{Early\ Phase} \geq 1$. This approach ensures that only realistic values are considered for the evaluation of $r$. Namely, an individual who had contact with a patient in the late Ebola phase would have a higher risk of becoming infected than if he had a similar contact with a patient in the early Ebola phase.

We compared three different scenarios, which correspond to the three types of contact. Our analysis suggests that a less close type of contact (i.e., sharing a conversation rather than sharing a meal), resulted in higher $r$ estimates. While all individuals are more likely to transmit for closer types of contacts, the differential in risk of transmitting between individuals with relatively high VL and low VL decreases with the closeness of the contact, from conversation to sharing a meal to sharing a bed.

Additionally and conservatively, we did not assume that the risk of infection would increase with clinical symptoms such as bleeding or vomiting, which is more likely to occur during the late phase of infection.

**Simulation studies**

For every infected individual recorded in our data set, we independently sampled the incubation duration, the infection period, and early phase duration. In each iteration, we stored a vector that tracked the number of infectious individuals (Supplement 5, shown in yellow) and the number of individuals exposed to Ebola per day (Supplement 5, shown in blue). We also sampled the daily VL and the number of contacts in early and late phase infection (Main text, Table 1). For any given day, we evaluated the probability that a newly infected individual could have been infected by any of the currently infectious individuals (Main text, Eq. (1)). Based on these probabilities we assigned the source of infection for each exposed individual.
As of October 2, according to the US CDC (30), there were 3696 reported cases and 1998 reported deaths in Liberia. Calculating case fatality by dividing the number of deaths reported to date by the number of cases reported to date (i.e. ~54%), will be biased downwards, because a substantial proportion of the unresolved cases reported will still progress to fatalities. Following discussion with contact tracers, and consistent with a recent study (20), the reported fatality date of non-survivors represents the day of death while for survivors the date of case is typically reported the day after symptom onset, around which we consider uncertainty by varying incubation duration, and infectious duration (Main text, Table 1). Further, the actual number of infected individuals beyond the end date of our contact-tracing data (i.e., September 22nd) is unknown. Thus, in each iteration, we used only cases whose entire infectious period, as obtained from the stochastic process of transmission (Supplement 5), was within the model timeframe. Taking this adjustment into consideration, case fatality as obtained from the model was estimated to be 63% (95% CI 60-64%).

**Effectiveness of intervention strategies**

Due to logistical challenges in case detection (24) and shortage of isolation units, we evaluated the probability of achieving disease elimination from case isolation of 50% to 100% of the non-survivors compared with of all infected individuals, when the type of contact is dominated through conversation or by sharing a meal or by sharing a bed (Supplement 7). Our results indicate that isolation of 70-100% within four days from onset of their symptoms has a high chance of achieving disease elimination for all types of contacts.

Additionally, we assessed a more pragmatic strategy that included self-quarantine among 50-100% of the infected, considering contact reduction from symptom onset that varied between 0 to 100% (Supplement 8). Consistent with our main finding, our results indicate that a reduction in contacts of 60%-100% is likely to curtail transmission.
Legends

Supplement 1. Cumulative Incidence Reported From the Liberian Ministry of Health and Social Welfare (MoHSW). Individual contact tracing data was used between August 7th and August 26th, and aggregated cases data was used in the rest of the model analysis period.

Supplement 2. Contact Distribution of an Infected Individual During the Infectious Period. The dash line represents the upper bound of the number of contacts for non-survivors in the late phase.

Supplement 4. Distribution of the Rate Ratio of Transmission Risk for Three Types of Contact. A. Conversation B. Sharing a meal C. Sharing a bed. Each ten-fold increase in VL is assumed to lead to an r-fold rise in infectiousness.

Supplement 5. Visualization Transmission Patterns in One Stochastic Iteration. Blue represents the day of exposure, yellow represents infectious period. Darker colors represent higher transmissibility resulting from a combination of contacts and viral load. A. Visualization of incidence over time B. Visualization of aggregated incidence by date.

Supplement 7. Probability Of Disease Elimination From Case Isolation of Nonsurvivors and of All Infected for Different Transmission Routes. A. Conversation, non-survivors B. Conversation, all infected C. Sharing a meal, non-survivors D. Sharing a meal, all infected. E. Sharing a bed, non-survivors F. Sharing a bed, all infected.