# Web extra material to "Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study" by Faye et al, Lancet Infectious Diseases

## I . Ebola fever case definition

A suspected case is defined as any person, alive or dead, who has (or had) sudden onset of high fever and had contact with a suspected, probable or confirmed Ebola case, or a dead or sick animal OR any person with sudden onset of high fever and at least three of the following symptoms: headache, vomiting, anorexia/loss of appetite, diarrhoea, lethargy, stomach pain, aching muscles or joints, difficulty swallowing, breathing difficulties, or hiccup; or any person with unexplained bleeding OR any sudden, unexplained death. A probable case is any suspected case evaluated by a clinician OR any person who died from 'suspected' Ebola and had an epidemiological link to a confirmed case but was not tested and did not have laboratory confirmation of the disease. A probable or suspected case is classified as confirmed when a sample from that person tests positive for Ebola virus in the laboratory.

## II . Classifying exposure occurrences in Ebola fever cases

For each case, data were collected using a standard case investigation form. This included dates of symptom onset, hospitalization and death, occupation (HCW or not), an indicator of nosocomial transmission if the case was not a HCW, the identity of up to 5 contacts who had presented symptoms compatible with EVD before the case and dates of contact if known, and whether the case had attended the funeral of someone who had died with a disease compatible with EVD.

Extra information was obtained during the interviews and in depth investigation of familial contacts that helped identifying chains of transmission. In some cases, multiple contacts were listed as possible infectors.

Cases were classified as: "Nosocomial" if they were not HCWs and only had contacts with EVD cases while in the hospital (there were 3 of them); "HCWs"; or "Community" cases.

We defined three possible contexts of exposure ("Hospital", "Community" or "Funeral") as follows:

- Hospital exposure corresponds to contacts of nosocomial or HCW cases on days when they did not attend funerals.
- Community exposure corresponded to contacts of community cases on days when they did not attend funerals.
- Funeral exposure corresponded to contacts on days when the case attended funerals, and could affect nosocomial, HCWs or community cases.

## II. Estimating incubation period

Using the number of daily exposure occurrence in the three contexts for each case/contact pair, the incubation period was estimated at maximum likelihood as follows. For each case *i*, denote $d_i$ the day of symptom onset and $j = 1, \ldots, \mathrm{K}_i$ the index of the possible infectors, where $K_i$ is the number of possible infectors. For each pair (*i*, *j*), exposure was possible on day *t* starting on day $s_{i,j}$ and ending on day $e_{i,j}$. Existence of an exposure on day *t* for pair (i,j) was denoted x(i,j,t) = 1 and 0 otherwise.

The contribution of an exposure event to the likelihood was simply equal to the probability of the associated incubation period. The cumulative density function of the incubation period was F(.;θ), where θ corresponds to the shape and scale parameters of the Gamma distribution. The log-likelihood was finally written as:

$$\sum_i \log \frac{1}{K_i} \sum_{j=1}^{K_i} \sum_{t=s_{i,j}}^{e_{i,j}} \left( F\left(d_i - t; \theta\right) - F\left(d_i - (t+1); \theta\right) \right) x\left(i, j, t\right)$$

Maximization was done using Mathematica. We report the mean and standard error of the fitted distributions.

**III. Estimating serial interval**
Using the same notation as above, each pair (i, j) defined a serial interval d(i, j).

The log-likelihood was written as:

$$\sum_i \log \frac{1}{K_i} \sum_{j=1}^{K_i} \left( F\left(d(i,j)+1;\theta\right) - F\left(d(i,j);\theta\right) \right)$$

We report the mean and standard error of the fitted distributions.

**IV. Resolving the transmission tree**
In our baseline analysis, each exposure occurrence was given a probabilistic weight that was simply proportional to the probability of the associated incubation period. In this model, if a case was to be exposed in multiple contexts on the same day, each context would therefore have equal probability of being the true context of infection on that day.

In a sensitivity analysis, we explored a model where the probabilistic weight associated with an exposure occurrence also depended on the context, with a multiplicative factor equal to $\lambda_C = 1$ (reference), $\lambda_H$ and $\lambda_F$ for community, hospital and funeral transmission, respectively. Our baseline scenario therefore corresponds to $\lambda_C = \lambda_H = \lambda_F = 1$.

Table S1 shows estimates are robust to assumptions about $\lambda_H$ and $\lambda_F$. This is because only a small proportion of cases are exposed to multiple contexts/infectors.

**Table S1: Proportion of transmission occurring in the community, in hospital and at funerals for different assumptions about per-day relative risk of infection per context.**

| | **Proportion of transmission** | | |
|---|---|---|---|
| | **In the community** | **In hospital** | **At funerals** |
| $\{\lambda_C = 1, \lambda_H = 1, \lambda_F = 1\}$ (baseline) | 82% | 12% | 6% |
| $\{\lambda_C = 1, \lambda_H = 1, \lambda_F = 4\}$ | 80% | 12% | 8% |

For example, in the scenario $\{\lambda_C = 1, \lambda_F = 4\}$, if a case was to be exposed to both transmission in the community and at funerals on the same day, he/she would be four times more likely to be infected at funerals than in the community.

Our analysis cannot be used to estimate the force of infection, i.e. the probability an exposed contact becomes infected in a particular setting, because this would require both numerator data (i.e. data describing contacts that became infected) and denominator data (data describing contacts that were not infected), when we only have numerator data. Our analysis can therefore not be used to estimate coefficients $\lambda$.

**V. Relationship between context specific reproduction numbers and the overall reproduction number**
Consider a scenario where a case infects on average 1 person in the community, 1 person in hospital and 1 person at funerals. The context-specific reproduction numbers are all equal to 1. The total number of persons infected by a case is equal to 3 on average (overall reproduction number).

## VI. Viraemia and community reproduction number

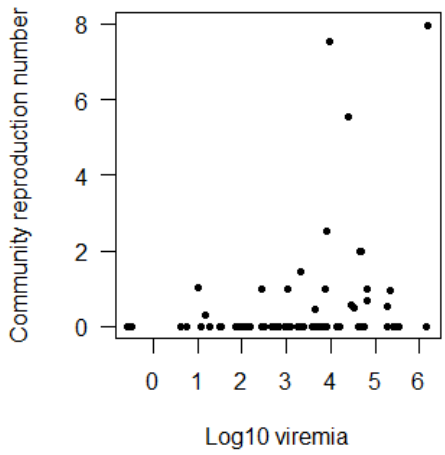
Viraemia was measured for 83 cases (55%) in the transmission tree. Samples were collected on average 5 days after symptom onset (inter-quartile range: 2, 7 days; range: 0, 21 days; SD: 3 days). For most cases with documented viraemia, only one sample was available. For the very few cases with multiple samples, we kept the value of viraemia in the earliest sample available.

A positive correlation between viraemia and the community reproduction number might be artificially generated for example if the two variables were positively correlated to the delay from onset to sample collection. To correct for this possible confounding, we adjusted for the delay from symptom onset to sample collection in our analyses. We also performed the analyses on restricted subsets of cases (samples collected in the 7 days following symptom onset, in the 10 days following symptom onset, all samples). Table S2 summarizes estimates of the relationship between the community reproduction number and viraemia.

**Table S2: Poisson regression of the community reproduction number with viraemia as explanatory variable, adjusting for the time period (March vs after March).** Different subsets of cases were considered. We also present analyses adjusted for the delay between symptom onset and sample collection.

| Subset of cases | Number of cases | Relative risk for 1 log10 unit increase in viraemia | |
|---|---|---|---|
| | | Unadjusted | Adjusted for the delay between symptom onset and sample collection |
| Samples collected in the 7 days following symptom onset | 63 | 1.9 (1.2, 3.1) | 1.8 (1.1, 3.0) |
| Samples collected in the 10 days following symptom onset | 76 | 2.6 (1.2, 5.4) | 2.5 (1.3, 4.8) |
| All samples | 83 | 2.5 (1.4, 4.7) | 2.3 (1.3, 4.3) |

**Figure S1: Community reproduction number as a function of log10 viraemia. (Pearsons's coefficient of correlation: 0.29, 95% CI: 0.07, 0.48, p=0.01).** Viraemia was measured for 83 cases (55%) in the transmission tree. Samples were collected on average 5 days after symptom onset (inter-quartile range: 2, 7 days; range: 0, 21 days). For most cases with documented viraemia, only one sample was available. For the very few cases with multiple samples, we kept the value of viraemia in the earliest sample available.



## VII. Family contacts in the transmission tree

In Figure S2, we reproduce Figure 2D, but this time with the information about transmission in the family.

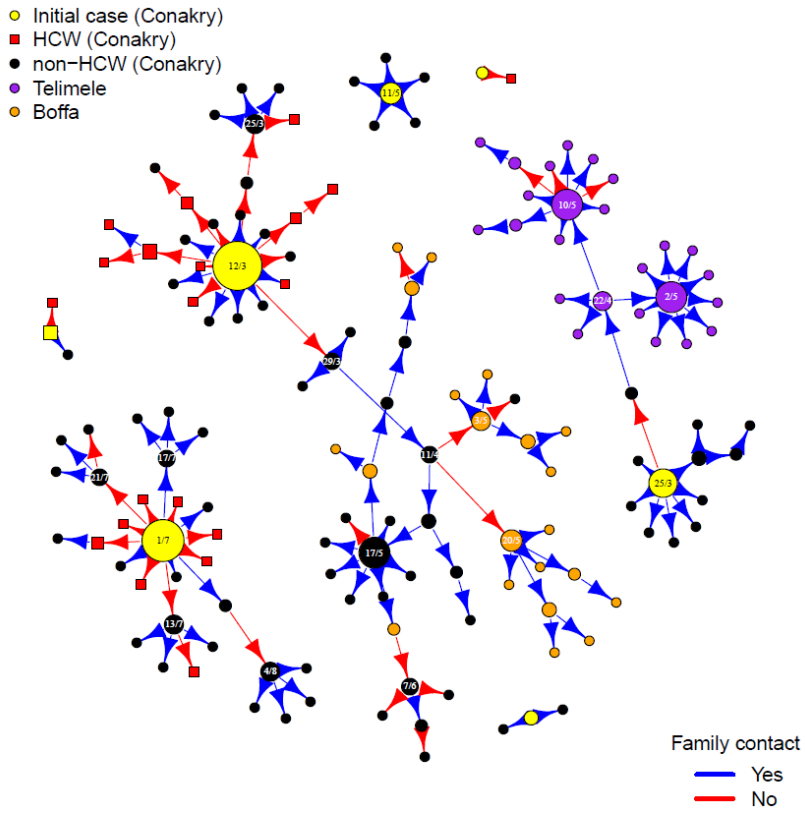



**Figure S2: A fully resolved transmission tree similar to Figure 2D, but with the information about transmission in the family. Each point represents an EVD case. The size of the point is proportional to overall reproduction number *R*. Dates of symptom onset are indicated in the figure for cases with *R*≥3. HCWs are represented by squares. The color of the arrow indicates whether transmission takes place in or out of the family.**