***Epidemiology and Infection***

**Geospatial Analysis of Household Spread of Ebola Virus in a Quarantined Village — Sierra Leone, 2014**

Authors: B. L. GLEASON, S. FOSTER, G. E. WILT, B. MILES, B. LEWIS, K. CAUTHEN, M. KING, F. BAYOR, S. CONTEH, T. SESAY, S. I. KAMARA, G. LAMBERT, P. FINLEY, W. BEYELER, T. MOORE, J. GAUDIOSO, P. H. KILMARX, J. T. REDD

**Supplementary Figure S1 (online animation)**

An interactive animation of the outbreak in Village X, which includes dates of symptom onset for the index case in each case-household and dates of household and village quarantine, is available in the following formats: (1) in the zip folder within the supplementary materials as an interactive swf file, (2) as an avi file within the supplementary materials, or (3) online at: <http://www.fastswf.com/31WQktc>.

**Appendices**

1. Quarantine:In August–October, 2014 in Sierra Leone, household quarantine entailed the following: identifying households where a symptomatic individual with Ebola lived or visited; requiring household members to remain within their households for 21 days (i.e., the duration of the maximum incubation period for Ebola) after last contact with an individual with Ebola for daily monitoring; extending the quarantine if another person in the household developed Ebola; providing food provisions for the duration of the quarantine; providing security around the household; and demarcating the household with orange tape around its perimeter. The village-wide quarantine in Village X essentially prohibited villagers from exiting and visitors from entering, unless part of the public health response, by placing security guards at each entrance to the village and advising the villagers of the policy. Residents of non-quarantined households were still able to move about within the village.
2. Household characteristics collected by in-person interviews:
3. Number of people who lived in the household at the onset of the Ebola outbreak
4. Number of houses in the household
5. Number of rooms in the household
6. Number of toilets or latrines for the household
7. Date of onset of first person in household with Ebola
8. Date of quarantine initiation
9. Name of first person in household who became sick with Ebola
10. Age of first person in household who became sick with Ebola
11. Sex of first person in household who became sick with Ebola
12. Number of people in household who became sick from Ebola
13. Number of people in household who died from Ebola
14. Number of people in household currently in a holding centre or Ebola Treatment Unit (ETU)
15. Number of people in household who survived
16. Number of people who live in household at time of interview
17. Number of people in household who fled the household during the outbreak (e.g., ran away to the bush)
18. MoHS unique identifier for Ebola cases in household as captured in the surveillance database
19. Other information (free text field captured information such as households of village leaders, residents currently in holding centres, and hygiene practices)

Of note, questions 5–13 and 16 were not asked during interviews of households that did not have an Ebola case.

**Supplementary Table S1.** Household Characteristics by Ebola Case Status in Village X, Sierra Leone, August to October 2014

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Case-households** | | | | **Non Case-households** | | | | **T-Test comparing case- and non case-households** |
| **Independent**  **Variables** | **Mean** | **Median** | **SD** | **Range** | **Mean** | **Median** | **SD** | **Range** | **P-Value** |
| **# of people per household** | 16.7 | 10.0 | 15.1 | 2–64 | 11.1 | 9.0 | 7.8 | 1–35 | 0.085 |
| **# houses per household** | 1.6 | 1.0 | 1.0 | 1–5 | 1.4 | 1.0 | 0.7 | 1–4 | 0.370 |
| **# rooms in household** | 6.9 | 5.0 | 4.4 | 2–22 | 5.3 | 5.0 | 3.0 | 1–17 | 0.125 |
| **# of people per house** | 10.0 | 9.0 | 6.3 | 2–30 | 7.9 | 7.0 | 4.6 | 1–21 | 0.143 |
| **# of toilets/latrines per household** | 0.7 | 1.0 | 0.5 | 0–2 | 0.8 | 1.0 | 0.5 | 0–2 | 0.826 |
| **# of people in household who became sick with Ebola** | 4.8 | 3.0 | 6.4 | 1–30 | NA | NA | NA | NA | NA |
| **# of people in household who died** | 3.5 | 2.0 | 5.9 | 0–24 | NA | NA | NA | NA | NA |

1. Spatio-Temporal Analyses
2. Weighted K

Multi-Distance Spatial Cluster Analysis is a way to analyse the spatial pattern of incident point data by summarizing spatial dependence (feature clustering or feature dispersion) over a range of distances. K-function analysis focuses on the variance of interevent distances. It considers all combinations of pairs of points. It compares the number of observed pairs with the expectation at all distances, taking into consideration the density of points, the borders, and the size of the sample. The weighted K-function considers both location (x, y) and the value of a point (z), in this case whether the household had Ebola-infected cases (1) or not (0). The K-function evaluates spatial distribution of a selected factor (such as Ebola case-households) in relation to Complete Spatial Randomness (CSR), even when a weight field is provided. When the Observed L(d) value is larger than the Maximum L(d) value, spatial clustering of infected households for that distance is statistically significant. When the Observed L(d) value is smaller than the Minimum L(d) value, spatial dispersion of infected households for that distance is statistically significant.

1. SaTScan

SaTScan™ software using Kulldorf’s scan statistic method of retrospective space-time permutation, purely temporal, and purely spatial Poisson models was used to identify Ebola-infected household clusters. We used household population size as a covariate divided into three groups, based on a K means cluster analysis of the household populations:

1. Space-Time

Space-time cluster analysis revealed two statistically significant clusters:

• A primary one including five infected households in a 58 metre circle that includes 10 households over a 2-day period from Days 32 to 33 (10–11 September 2014) with a p-value of 0.013.

• A secondary one including five infected households in an 86 metre circle that includes 25 household over a 1-day period on Day 27 (5 September 2014) with a p-value of 0.047.

1. Temporal

Purely temporal cluster analysis revealed one statistically significant cluster, including fifteen infected households over an eight day period from Days 27 to 34 (5–12 September 2014) with a p-value of 0.003.

The fact that the space-time clusters are contained within the more significant temporal cluster and that there are no spatial clusters lends evidence to the hypothesis that that space-time cluster are subsets of the more statistically significant temporal cluster and not due to the spatial relationships between the households.

IV. Detailed Description of Distance Metrics

1. For each household we created a matrix of distances measuring the distance between that given household and each case-household that had symptom onset for its index case at least 2 days prior to that given household (i.e., each potential source-household). Because a given household had more than one potential source-household, we calculated the average distance of these distances from each potential source-household. Because households became case-households (and thus potential source-households) on different days throughout the study period, the distance metric accounts for the fact that the average distances to case-households were dynamic over the course of the outbreak. Therefore, each household has multiple distance metrics corresponding to the days when new households became case-households.

i. We also conducted alternate distance metric calculations (listed below in Appendix IV.9) including the minimum distance and sum of distances between a household and multiple case-households, but found that average distance metric was the most appropriate representation of our data.

ii. Because we hypothesize that a smaller distance increases risk, we found that it worked better to average rather than sum the distances.

iii. The minimum distance index, which we could make binary by picking a cutoff value, did not reflect the full degree of risk either because it only quantified the risk posed by the closest household rather than several nearby households.

2. The number of records for each household corresponds to a time-step (measured in days). The time-step has a value that represents every day in this study period when new households become case-households. As new case-households appeared on 17 of the 46 days in the study period, there were 17 time-steps and therefore 17 distance values were calculated for each household in the distance metric. As households became infected, they were removed from the next day’s non-infected household list and added to the potential source-household list.

3. Calculations of average distances incorporated dates of household infection to determine which case-households were potential source-households to each uninfected household. Integrating date of onset for each case-household into distance calculations allowed us to address temporality and spatial factors of household transmission within a single distance metric.

4. For the GEE analysis, each household maintains multiple records used in the modeling that include the average distance metric and the number of days between events.

5. The number of days and distances in the analysis that does not account for quarantine represents the time interval between time-steps and the average distance between a household and new case-households.

6. An additional distance metric (average quarantine distance metric) was developed to conduct an analysis of the effect of household quarantine. In the quarantine analysis, any case-household that was quarantined contributed days and distances to the average quarantine distance metric until the date of quarantine.

7. A third distance metric (average prompt quarantine distance metric) was developed to conduct an analysis of the effect of prompt household quarantine. In the prompt quarantine distance analysis, only case-households quarantined within 48 hours of date of symptom onset were categorized as undergoing prompt quarantine, while all other case-households were categorized as not quarantined. Case-households with prompt quarantine do not contribute days or distances to the distance metric after the date of quarantine.

8. For the prompt quarantine analysis we compared the odds ratio (OR) of the average distance metric to the OR of the prompt quarantine average distance metric.

9. The following distance metrics were explored to determine the most appropriate metrics for our data (selected metrics are identified by \*):

* + - 1. Total distance
      2. Minimum distance
      3. Median distance
      4. Average distance\*
      5. Total quarantine distance
      6. Minimum quarantine distance
      7. Median quarantine distance
      8. Average quarantine distance\*
      9. Total prompt quarantine distance
      10. Minimum prompt quarantine distance
      11. Median prompt quarantine distance
      12. Average prompt quarantine distance\*

1. Description of GEE

We developed generalized estimating equations (GEE) to describe risk factors for household acquisition of Ebola virus while incorporating repeated measures on households over time. GEE allowed for correlations of each household’s repeated measures. Our primary outcome variable, *case-household*, describes a household’s infection status on any given day during the study period (i.e., 0=uninfected, 1=infected); therefore, we modeled the probability that an uninfected household would become infected (i.e., a case-household) each day.

We modeled the effects of the average distance metric, average quarantine distance metric, and average prompt quarantine distance metric separately in the univariate analysis and compared resulting odds ratios to determine variations in the average quarantine distance metric and average prompt quarantine distance metric effects. Ultimately, the average quarantine distance metric was excluded from GEE based on non-significant univariate analysis results and given that the average prompt quarantine distance metric provided a mechanism to assess effect of household quarantine that had been implemented in an appropriate timeframe. We used the R package geepack with an exchangeable correlation structure. There were 64 clusters with the largest cluster having 17 values.

1. Predictors

The predictors entered into the first GEE were time, number of persons who were living in the household at the onset of the outbreak, and the average distance metric. The predictors entered into the second GEE were time, number of persons who were living in the household at the onset of the outbreak, and the average prompt quarantine distance metric. These predictors were used to maintain consistency and interpretability across the various model structures.

The GEE structure allows the intercept to vary by household. This gave each household a different intercept that represents its own baseline propensity for infection, thus adjusting for unmeasured household-level variables that may affect probability of infection. Furthermore, the model followed the assumption that repeated measurements made on the same household are correlated with one another.

1. Time

Days since 9 August 2014. The study period runs from day 1 (10 August 2014) to day 46 (24 September 2014). All case households drop out of the study after their day of infection.

1. Average distance metric

Average distances to case-households were calculated as described in Appendix IV. The distance from each household to potential source-households was calculated and the mean value was used for analyses.

1. Average prompt quarantine distance metric

The average prompt quarantine distance metric was calculated as described in Appendix IV similarly to the average distance metric; however, if the household was quarantined within 48 hours, it was excluded from the metric to test the assumption that prompt quarantines were effective in preventing household transmission of Ebola virus.