

Determinants of transmission risk during the late stage of the West African Ebola epidemic

Alexis Robert, W. John Edmunds, Conall H. Watson, Ana Maria Henao-Restrepo, Pierre-Stéphane Gsell, Elizabeth Williamson, Ira M. Longini Jr., Keïta Sakoba, Adam J. Kucharski, Alhassane Touré, Sévérine Danmadji Nadlaou, Boubacar Diallo, Mamamdou Saidou Barry, Thierno Oumar Fofana, Louceny Camara, Ibrahima Lansana Kaba, Lansana Sylla, Mohamed Lamine Diaby, Ousmane Soumah, Abdourahime Diallo, Amadou Niare, Abdourhamane Diallo and Rosalind M. Eggo

Correspondence to Dr. Rosalind M. Eggo, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT (e-mail: r.eggo@lshtm.ac.uk)

Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, UK (Alexis Robert, W. John Edmunds, Conall H. Watson, Adam J. Kucharski, and Rosalind M. Eggo); World Health Organization, Geneva, Switzerland (Ana Maria Henao-Restrepo and Pierre-Stéphane Gsell); Department of Medical Statistics, London School of Hygiene & Tropical Medicine, UK (Elizabeth Williamson); Department of Biostatistics, University of Florida, USA (Ira M. Longini Jr.); WHO Ebola vaccination team, Guinea (Keïta Sakoba, Alhassane Touré, Sévérine Danmadji Nadlaou Mamamdou Saidou Barry, Thierno Oumar Fofana, Louceny Camara, Ibrahima Lansana Kaba, Lansana Sylla, Mohamed Lamine Diaby, Ousmane Soumah, Abdourahamane Diallo, Amadou Niare and Adbourhamane Diallo); and Ministry of Health, Guinea (Boubacar Diallo).

FUNDING

AR acknowledges funding from the Norwegian Institute for Public Health “A randomised trial of ring vaccination to evaluate Ebola vaccine efficacy and Safety in Guinea, West Africa”. WJE acknowledges funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement EBOVAC1 (grant 115854). The IMI2 is supported by the European Union Horizon 2020 Research and Innovation Programme and the European Federation of Pharmaceutical Industries and Associations. AJK acknowledges funding from a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant Number 206250/Z/17/Z). RME acknowledges funding from an HDR UK Innovation Fellowship (grant MR/S003975/1).

Conflict of interest: none declared.

Running head: Risk factors for Ebola transmission in Guinea

ABSTRACT

Understanding risk factors for Ebola transmission is key for effective prediction and design of interventions. We used data on 860 cases in 129 chains of transmission from the latter half of the 2013-16 Ebola outbreak in Guinea. Using negative binomial regression, we determined characteristics associated with the number of secondary cases resulting from each infected individual. We found that attending an Ebola Treatment Unit was associated with a 38% decrease in secondary cases (Incident rate ratio (IRR) 0.62, 95%CI: 0.38, 0.99) in individuals that did not survive. Unsafe burial was associated with a higher number of secondary cases (IRR 1.82, 95%CI: 1.10, 3.02). The average number of secondary cases was higher for the first generation of a transmission chain (mean = 1.77), compared with subsequent generations (mean = 0.70). Children were least likely to transmit (IRR 0.35 (95%CI: 0.21, 0.57) compared with adults, whereas older adults were associated with higher numbers of secondary cases. Men were less likely to transmit than women (IRR 0.71 (95%CI:

0.55, 0.93)). This detailed surveillance dataset provided an invaluable insight into transmission routes and risks. Our analysis highlights the key role that age, receiving treatment, and safe burial played in the spread of EVD.

KEYWORDS

Ebola; Risk factors; Regression analysis; Multiple imputation; Guinea

ABBREVIATIONS

EVD Ebola virus disease

CI Confidence interval

IRR Incidence Rate Ratio

ETU Ebola Treatment Unit

IQR Interquartile Range

ORIGINAL UNEDITED MANUSCRIPT

Between December 2013 and June 2016, the largest Ebola virus disease (EVD) epidemic to date occurred in West Africa, causing more than 28,000 cases mainly in Liberia, Sierra Leone and Guinea (1). Of these, 3,804 confirmed cases and 2,536 deaths were in Guinea (1). There remains a pressing need to understand the transmission dynamics of this outbreak, so that interventions can be designed, and accurate forecasts can be made as outbreaks continue to occur (2,3).

During the 2013-16 epidemic in Guinea, intensive epidemiological investigation was made of cases, including assembling those individuals into chains of transmission, which link infected individuals to their descendant cases during case investigations (4). In contrast to studies of cases, which can give insight only on risk factors for acquisition of EVD (4–6), or genetic analysis, which has been used to reconstruct spatial dispersion of the disease in different regions (7), transmission chains allow detailed analyses of the risk factors for onward transmission (8,9). These data are invaluable for understanding the characteristics of individuals likely to have high onward transmission, but have so far been underused in analyses of this outbreak.

Using a large database of epidemiologically-reconstructed transmission chains we summarized information on cases reported in the late stages of the 2013-16 outbreak in Guinea. We used this information to identify characteristics of cases that were associated with increased onward transmission.

METHODS

Data from two databases were linked in this study: the transmission chain dataset (1012 cases) and the Guinean surveillance database of EVD cases. 152 individuals in the transmission chain dataset were participants in the *Ebola ça Suffit* ring vaccination trial (10,11). These cases were removed because of the likely impact of the trial on transmission, leaving 860 cases used in this analysis.

Transmission chain dataset

Data on cases and the epidemiological links between them were collected by the Ministère de la Santé et de l'Hygiène Publique of Guinea (Ministry of Health and Public Hygiene) during the epidemic. Field teams conducted interviews with cases (where possible) and their contacts, as part of epidemiological investigations. Based on contact with confirmed or probable cases, the most likely infector or infectors were assigned to each case. The chains of transmission were continually revised and updated during the EVD response in Guinea, and when new cases were confirmed those were added to the database and to the chains. This could result in changes to the likely infector or joining sub-trees together as new information became available. The chains therefore represent the best possible epidemiological linkage of cases to each other, made by trained field teams with access to cases, contacts, and contextual information. We restricted our analysis to confirmed and probable cases infected between September 2014 and November 2015 because transmission chains were available during this time period, and resources were available to digitize these data.

Variables were the age, gender, location (prefecture, sub-prefecture and village), survival status of the case, whether the burial was safe or unsafe, the epidemiologically-inferred source of infection, and the route of transmission (including household, nosocomial, neighbor, ETU), national identification number for each individual (Table 1). Geographic information, demographic variables and the probable routes of transmission were ascertained by the field epidemiology teams (Web Appendix 1). Safe burial means that it was safe and dignified burial conducted by a trained burial team. We used dates of: i) onset, ii) admission to Ebola treatment unit (ETU), iii) discharge from an ETU, iv) death, and v) burial. We deleted 10 implausible epidemiological links (for example, where the end of symptoms of the infectee is earlier than the start of symptoms of the named infector). When several infectors were reported and plausible for a case, we considered only one link in the transmission chain by random selection. We conducted sensitivity analysis on this selection.

Guinea surveillance database

The Guinean surveillance database is a line list of confirmed cases in Guinea from the national surveillance system. Each record contains the same information on each case as the transmission chain dataset, except for the transmission link, but completeness of other fields (such as dates) is higher. Therefore, we matched the transmission chain dataset to the surveillance database using national identification number, or name, location, age, and dates of infection. This increased the completeness of the data used in this analysis.

Matching cases

664 cases in the transmission chain dataset (77.2%) matched with a record in the surveillance database. 135 (15.7%) of these did not provide extra information on the case. We compared the features of the 529 remaining cases in each database to eliminate mismatches (Web Appendix 2, Web Figures 1-2) and used the surveillance database to supplement features of 380 (44.2%) cases. Among these cases, all reported variables matched for 71 individuals, and the surveillance database contributed information for 380 cases. In cases that were not perfect matched, the mismatches were minor and we assumed these differences were due to data entry errors as other variables matched (Web Appendix 2). For the other cases, we kept the features described in the transmission chain dataset. 196 of the 860 individuals in the transmission chain dataset could not be matched to cases in the Guinean surveillance database (22.8%). Table 1 shows reporting and values of each variable in the final dataset.

Classification by number of transmissions

We calculated the number of reported secondary cases for each individual, and categorized them as: i) high transmitters (more than 3 cases), ii) moderate transmitters (1 to 3 cases), iii) no onward transmission, and tested for associations with demographic characteristics (Web Appendix 3, Web Figure 3).

Statistical analysis

We used negative binomial regression to estimate the impact of characteristics of the cases on the number of secondary cases caused (17). We grouped the age of cases into: 0-14; 15-34; 35-54; 55-74; 75+ years. Conakry and Prefecture city centers were considered as urban areas, whereas other areas (villages, towns) were defined as rural areas. Cases without a known infector were defined as the first generation of a chain, and all others as subsequent generations of a chain.

We created a variable that combined the survival status, ETU attendance, and burial status of each case, called the "Outcome". In the data, all reported survivors had been admitted to an ETU and all non-survivors that had been admitted to an ETU had a safe burial. Therefore we used four unordered levels describing the outcome of each case: i) Survivor who attended an ETU, ii) Non-survivor who attended an ETU and was safely buried, iii) Non-survivor who did not attend an ETU and was safely buried, iv) Non-survivor who did not attend an ETU and was unsafely buried (Web Appendix 3, Web Figure 4).

Imputation of missing data

ORIGINAL UNEDITED MANUSCRIPT

Four of the variables included in the negative binomial regression analysis were incompletely reported: gender (1.7% missing), age (11.8% missing), survival status (4.5% missing), burial safety status (29.4% missing) and ETU admission status (13.6% missing) (Web Appendix 3, Web Figures 5-6). Since some of the variables were incomplete, and we aimed to retain the full population in the regression analysis, we used multiple imputation for missing values (12–14). In the imputation model, we considered all factors included in the regression analysis as explanatory variables (see below), plus four others from the database: the number of community, funeral, or nosocomial transmissions caused, and month of onset. The imputed variables were age (using predictive mean matching), survival status, burial safety status, and ETU admission status (logistic regression). We assumed that missing data were missing at random (15) (Web Appendix 4, Web Table 1, Web Figures 7-8). Forty datasets were generated using the MICE package in R version 3.5.0 (16). We used pooled coefficient estimates drawn from 40 imputed datasets (Web Appendix 4, Web Table 2, Web Figure 9), and performed sensitivity analysis on the multiple imputation (Web Appendix 5, Web Tables 3-4).

RESULTS

Chains of transmission

The proportion of the total cases represented in this dataset increases through time (Figure 1). There were 818 cases in 87 chains of transmission of 2 to 11 generations (Figure 1) and 42 individuals not linked to any infector or to subsequent cases. These first generation cases occurred throughout the study period (Web Figure 2). The largest chain of transmission included 78 cases, starting on 1st January 2015 and ending on 25th April 2015.

The mean serial interval (time between the date of onset of the infected case and onset in their infector) was 12.3 days (Figure 2A), calculated from 308 serial intervals. The serial interval did not vary through time (Figure 2A), by route of transmission, age, or by generation of the chain (Web Appendix 6, Web Figure 10).

Reproduction number

The unadjusted average number of secondary cases per individual was 0.89. Most cases did not result in subsequent transmission. 299 (34.8%) cases resulted in at least some transmission, and 62 (7.2%) individuals were deemed high transmitters and were responsible for 53.5% of the transmission events observed. The maximum number of observed secondary cases was 22. We fitted a negative binomial distribution to the number of secondary cases and found high dispersion (mean = 0.89, dispersion parameter = 0.31 (95% CI: 0.25, 0.37), Index of dispersion = 3.87 (95% CI: 3.41, 4.56)). We stratified first and subsequent generations and observed a higher reproduction number among first generation individuals (mean = 1.77) than among the subsequent generations (mean = 0.70) (Figure 3, Web Appendix 6, Web Figure 11).

Univariable description of transmitters

Older cases were more likely to be high transmitters, whereas children or young adults were less likely to transmit: 49.4% of the cases who did not cause any transmission are younger than 30, whereas only 11.3% of high transmitters falls into this age category (Figure 2B). We did not observe any change in the route of transmission between moderate and high transmitters (Figure 2C). The major route of transmission was through the household (57.9% of all cases and 75.9% of transmission events). We did not observe any difference between case characteristics (such as gender and location) and classification of high, moderate, or no transmission (Web Appendix 3, Web Figure 3).

Determinants of transmission

We found significant associations of gender, outcome (four unordered levels), age category and being the first generation of a chain, with the number of secondary cases generated (Table 2). In our multivariable negative binomial regression model, the estimated intercept was 0.69 (95% CI: 0.38, 1.28) and was defined as the mean number of secondary cases for women, aged 35-54, who did not survive, did not go to an ETU, and had a safe burial, in urban area and were not the first generation of a chain (Table 2).

Individuals younger than 35 caused significantly fewer secondary cases, and the first generation of chains caused significantly more secondary case than subsequent generations (incidence rate ratio (IRR) = 1.76 (95%CI: 1.27, 2.44)). Men caused significantly fewer secondary cases than women (IRR= 0.71 (95%CI: 0.55, 0.93)), which was not observed in the univariable analysis.

By comparing the coefficients for non-survivors who did and did not attend an ETU, but had safe burials, we determined that there was a significantly lower number of secondary cases in those who attended an ETU (IRR = 0.62 (95%CI: 0.38, 0.99)). We found that unsafe burial was associated with a large increase in transmission (IRR = 1.82 (95% CI: 1.10, 3.02)). We compared individuals who survived with non-survivors in those who attended an ETU and had a safe burial, and found that survival was associated with significantly fewer secondary cases than non-survival (IRR = 0.51 (95%CI: 0.31, 0.82)) (Web Table 3).

There was no significant association between urban or rural location and number of subsequent cases.

DISCUSSION

Using the largest set of epidemiologically-linked transmission trees available for EVD, we identified key patient characteristics associated with increased onward transmission and estimated their association with the number of secondary cases each case generated. By doing this we have been able to quantify the association between attending an ETU and safe burials on onward transmission, in the late stage of the epidemic in Guinea.

Attending an ETU was associated with a large decrease in the number of transmission events, and unsafe burial was associated with an almost twofold increase in number of transmissions. Our estimates emphasize the importance of ETU attendance and safe burials as control measures for Ebola, and are similar to values found in other studies (8,18–22).

These data are drawn from the late stage of the epidemic, and the same risk factors for transmission extend even late in the epidemic, when awareness of EVD transmission routes may have been higher.

In agreement with previous studies of EVD transmission, we found a highly skewed distribution of secondary cases (18,19,23–25). Indeed, the majority of cases did not transmit EVD at all, and only a small number had high numbers of transmission events (26).

Importantly, we were able to determine the case-characteristics associated with the number of transmissions. This information could be used in real-time prediction, by incorporating information on the case mix of incident cases.

Our analysis is of disjoint transmission chains, which are observations of a fully-connected transmission tree. This complete tree contains the entire outbreak, with each case linked together. Our findings relating to the first generation of each chain are therefore a measure of the impact of a case not having a traced link to a prior case, and not the absence of a true link to a prior case.

We found that the first generation of each chain was associated with a higher number of secondary cases than those identified later in the chain. The first generation of each chain is necessarily an individual who could not be epidemiologically-linked to any prior chains by the field epidemiology teams. First generation cases may have spent longer in the community (and therefore had a longer transmission window) because they either were not traced by contact-tracing and therefore did not know they were at risk (27); or they evaded contact-tracing (28). Alternatively, or additionally, there may be a bias toward detection of large transmission events in our data, whereby untraced contacts are more likely to be detected if they give rise to a larger cluster of cases, i.e., ascertainment bias.

We found that children and young adults had lower onward transmission, whereas older adults were more likely to result in large numbers of secondary cases. Several studies have determined that children were at lower risk of infection in previous outbreaks (6,29), and during the west African epidemic (4,5,30), and behavioral differences in caring roles are suggested as the reason (29–31). Our study adds to evidence (32,33) that children were also associated with lower risk of onward transmission, although not all analyses find this pattern (34).

Three quarters of reported transmissions were in the household, making it the most frequent transmission route. Studies from the early stages of the epidemic found a higher contribution from funeral and nosocomial transmission routes (8) than observed here. Our findings could be the result of public health interventions to increase safe and dignified burials, protect health care workers, and raise awareness of transmission risks (22,35).

There were several interventions occurring at the time of the study (11,18). We accounted for the largest of these – the *Ebola ça Suffit* ring vaccination trial – by removing participants of the trial. Other interventions could have affected transmission, although we did not detect an association between time and the number of secondary cases in the model. In addition, the study period is in the latter part of the epidemic, and there may be differences in inferred transmission risk in other time periods of the epidemic.

Although we did not find evidence for differences between cases in the transmission chains dataset and those not, it is possible that there are different characteristics in the number of secondary cases generated.

This study is limited to observed cases, and therefore the number of transmission events could be lower than the true value. Of note, there were no survivors who did not attend an ETU in our data. It is likely that these survivors remained undetected, and therefore we could not include transmission risk from these individuals in our analysis. Not attending an ETU could be associated with other transmission risks, or with community resistance to interventions (36).

Multiple imputation provides unbiased estimates under the assumption that missing data were missing at random: given all the information available, the missing values were similar in distribution to the observed values. If, for example, all individuals with an unknown burial status in fact had an unsafe burial, then this assumption might be violated, potentially leading to bias in the estimated regression coefficients.

Some of the links inferred by on-the-ground epidemiologists may be incorrect, which could affect our estimates of determinants of onward transmission. However, in contrast to other studies which retrospectively linked cases into transmission chains (32,37), the chains used in our study were generated in real time. Genetic data linking cases together could be used to test if there are incorrect links (38), but these data were not available for this study.

By the end of the epidemic, the chains of transmission represent the best possible record of epidemiological investigations of EVD cases in Guinea. During periods of high numbers of cases, the epidemiological teams may have been more stretched and therefore surveillance effort per case could have been lowered. It is possible that proposed infectors have been mis-specified during this time period, which could affect the findings, but, based on the characteristics of cases, we think it is unlikely that this would be a systematic error.

During the EVD outbreak in Guinea, detailed investigations were conducted around each case reported to surveillance, to inform the public health response. This enormous undertaking resulted in large quantities of data that provide an invaluable insight into the routes and risk of transmission. Recent outbreaks in the Democratic Republic of the Congo indicate the vital importance of epidemiologically-informed control measures in the control of Ebola (39,40). Analyses of these data reveal the key role that older individuals and those that did not seek treatment played in the spread of EVD.

ACKNOWLEDGEMENTS:

Author affiliations: Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, UK (Alexis Robert, W. John Edmunds, Conall H. Watson, Adam J. Kucharski, and Rosalind M. Eggo); World Health Organization, Geneva, Switzerland (Ana Maria Henao-Restrepo and Pierre-Stéphane Gsell); Department of Medical Statistics, London School of Hygiene & Tropical Medicine, UK (Elizabeth Williamson); Department of Biostatistics, University of Florida, USA (Ira M. Longini Jr.); WHO Ebola vaccination team, Guinea (Keïta Sakoba, Alhassane Touré, Sévérine Danmadji Nadlaou Mamamdou Saidou Barry, Thierno Oumar Fofana, Louceny Camara, Ibrahima Lansana Kaba, Lansana Sylla, Mohamed Lamine Diaby, Ousmane Soumah, Abdourahamine Diallo, Amadou Niare and Adbourhamane Diallo); and Ministry of Health, Guinea (Boubacar Diallo).

We acknowledge the work of field epidemiologists in Guinea, and Arja Huetis and Megan O'Driscoll for data entry.

This work was supported by funding from the Norwegian Institute for Public Health “A randomised trial of ring vaccination to evaluate Ebola vaccine efficacy and Safety in Guinea, West Africa” (Mr. Robert). In addition, it was supported by funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement EBOVAC1 (grant 115854) (Prof. Edmunds). The IMI2 is supported by the European Union Horizon 2020 Research and Innovation Programme and the European Federation of Pharmaceutical Industries and Associations. The work was also funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant Number 206250/Z/17/Z) (Dr. Kucharski). It was also supported by funding from an HDR UK Innovation Fellowship (grant MR/S003975/1) (Dr. Eggo). The views expressed are those of the authors and not necessarily those of the funders. The funders had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Conflict of interest: none declared.

ORIGINAL UNEDITED MANUSCRIPT

FIGURE LEGENDS

Figure 1. Characteristics of the chains data. A) Time series of the daily incidence in Guinea. Light grey is total incidence, and dark grey area shows cases included in this analysis. B) Distribution of the number of generations per chain. C) Example of a chain with 9 generations. Squares mark women, and circles mark men. Lighter shade is under 35 years of age, and darker shade is over 35 years of age.

Figure 2. Characteristics of transmission. A) Serial interval through time, overall mean = 12.3 days. Number of cases in each group: 61, 163, 61, 23 respectively. B) Distribution of the transmitter status of the cases depending on the age of the cases. C) Distribution of the transmitter status depending on the route of transmission.

Figure 3. Distribution of secondary cases. A) Distribution of the number of secondary cases per case, and fit to a negative binomial distribution (blue curves). B) Fits of the first or the subsequent generations only (total: Mean = 0.89, sd = 0.31 (95%CI: 0.25, 0.37); 1st gen: Mean = 1.77, sd = 0.88 (95%CI: 0.53, 1.23), subsequent generation: Mean = 0.70, sd = 0.25 (95%CI: 0.19, 0.30)).

References

1. World Health Organization. Ebola Situation Report March 30, 2016. 2016;(January):1–16. (http://apps.who.int/iris/bitstream/10665/161976/1/roadmapsitrepre_15Apr2015_eng.pdf?ua=1&ua=1). (Accessed February 4, 2019)
2. World Health Organization. Ebola virus disease – Democratic Republic of the Congo. 2018;(http://www.who.int/csr/don/11-october-2018-ebola-drc/en/). (Accessed February 4, 2019)
3. Barry A, Ahuka-Mundike S, Ali Ahmed Y, et al. Outbreak of Ebola virus disease in the Democratic Republic of the Congo, April–May, 2018: an epidemiological study. *Lancet (London, England)*. 2018;392(10143):213–221.
4. Team WER. Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* 2014;371(16):1481–95.)
5. Glynn JR. Age-specific incidence of Ebola virus disease. *Lancet*. 2015;386(9992):432.
6. Brainard J, Hooper L, Pond K, et al. Risk factors for transmission of Ebola or Marburg virus disease: A systematic review and meta-analysis. *Int. J. Epidemiol.* 2016;45(1):102–116.
7. Dudas G, Carvalho LM, Bedford T, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*. 2017;544(7650):309–315.
8. Faye O, Boëlle P-Y, Heleze E, et al. Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect. Dis.* 2015;15(3):320–326.
9. Muoghalu I-S, Moses F, Conteh I, et al. The Transmission Chain Analysis of 2014–2015 Ebola Virus Disease Outbreak in Koinadugu District, Sierra Leone: An Observational Study. *Front. Public Heal.* 2017;5(July):1–10.
10. Henao-Restrepo AM, Longini IM, Egger M, et al. Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *Lancet*. 2015;386(9996):857–866.
11. Ebola ca suffit ring vaccination trial consortium. The ring vaccination trial: a novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola. *BMJ*. 2015;351:h3740.
12. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am. J. Epidemiol.* 2008;168(4):355–357.
13. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*. 2009;338:b2393.
14. Rubin DB. Multiple imputation for nonresponse in surveys. New York, NY: J.Wiley and Sons, Inc; 1987.

15. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York, NY: John Wiley & Sons, Inc; 2002.
16. Buuren S Van, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. Los Angeles, CA: University of California; 2010 1--68 p.
17. Lloyd-Smith JO, Schreiber SJ, Kopp PE, et al. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005;438(7066):355–359.
18. Agua-Agum J, Ariyarajah A, Aylward B, et al. Exposure Patterns Driving Ebola Transmission in West Africa: A Retrospective Observational Study. *PLoS Med*. 2016;13(11):1–23.
19. Lau MSY, Dalziel BD, Funk S, et al. Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proc. Natl. Acad. Sci*. 2017;114(9):2337–2342.
20. Lindblade KA, Nyenswah T, Keita S, et al. Secondary infections with Ebola virus in rural communities, Liberia and Guinea, 2014–2015. *Emerg. Infect. Dis*. 2016;22(9):1653–1655.
21. Hunt L, Gupta-Wright A, Simms V, et al. Clinical presentation, biochemical, and haematological parameters and their association with outcome in patients with Ebola virus disease: An observational cohort study. *Lancet Infect. Dis*. 2015;15(11):1292–1299.
22. Tiffany A, Dalziel BD, Kagume Njenge H, et al. Estimating the number of secondary Ebola cases resulting from an unsafe burial and risk factors for transmission during the West Africa Ebola epidemic. *PLoS Negl. Trop. Dis*. 2017;11(6):e0005491.
23. Althaus CL. Ebola superspreading. *Lancet Infect. Dis*. 2015;15(5):507–508.
24. Muoghalu I-S, Moses F, Conteh I, et al. The Transmission Chain Analysis of 2014–2015 Ebola Virus Disease Outbreak in Koinadugu District, Sierra Leone: An Observational Study. *Front. Public Heal*. 2017;5(July):1–10.
25. Ajelli M, Parlamento S, Bome D, et al. The 2014 Ebola virus disease outbreak in Pujehun, Sierra Leone: epidemiology and impact of interventions. *BMC Med*. 2015;13(1):281.
26. Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS One*. 2007;2(2):e180.
27. Rosello A, Mossoko M, Flasche S, et al. Ebola virus disease in the Democratic Republic of the Congo, 1976-2014. *Elife*. 2015;4(November 2014):e09015.
28. Agua-Agum J, Allegranzi B, Ariyarajah A, et al. After Ebola in West Africa — Unpredictable Risks, Preventable Epidemics. *N. Engl. J. Med*. 2016;
29. Dowell SF. Ebola hemorrhagic fever: Why were children spared? *Pediatr. Infect. Dis. J*. 1996;15(3):189–191.

30. Bower H, Johnson S, Bangura MS, et al. Exposure-specific and age-specific attack rates for Ebola virus disease in Ebola-affected households, Sierra Leone. *Emerg. Infect. Dis.* 2016;22(8):1403–1411.
31. WHO Ebola Response Team. Ebola Virus Disease among Children in West Africa. *N. Engl. J. Med.* 2015;372(13):1274–1277.
32. Dowell SF, Mukunu R, Ksiazek TG, et al. Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidémies à Kikwit. *J. Infect. Dis.* 1999;179 Suppl:S87–S91.
33. Lindblade KA, Kateh F, Nagbe TK, et al. Decreased Ebola Transmission after Rapid Response to Outbreaks in Remote Areas, Liberia, 2014. *Emerg. Infect. Dis.* 2015;21(10):1800–1807.
34. Fang L-Q, Yang Y, Jiang J-F, et al. Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc. Natl. Acad. Sci.* 2016;113(16):4488–4493.
35. Coltart CEM, Lindsey B, Ghinai I, et al. The Ebola outbreak, 2013–2016: old lessons for new epidemics. *Philos. Trans. R. Soc. B Biol. Sci.* 2017;372(1721):20160297.
36. Carrión Martín AI, Derrough T, Honomou P, et al. Social and cultural factors behind community resistance during an Ebola outbreak in a village of the Guinean Forest region, February 2015: A field experience. *Int. Health.* 2016;8(3):227–229.
37. Francesconi P, Yoti Z, Declich S, et al. Ebola Hemorrhagic Fever Transmission and Risk Factors of Contacts, Uganda. *Emerg. Infect. Dis.* 2003;9(11):1430.
38. Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science (80-)*. 2014;345(6202):1369–1372.
39. Wise J. WHO is “cautiously optimistic” about Ebola ring vaccination programme in DRC. *BMJ.* 2018;361(0959-8138).
40. World Health Organization. Ebola vaccination begins in North Kivu. *World Heal. Organ.* 2018;(http://www.who.int/news-room/detail/08-08-2018-ebola-vaccination-begins-in-north-kivu). (Accessed February 4, 2019)

Table 1:

Pre-imputation characteristics of the dataset.

Variable	Number	%	Transmission status (%)		
			None	Moderate	High
Age					
Children (0-14 years old)	137	15.9	88.3	10.2	1.4
Adults (15-99 years old)	621	72.3	64.1	27.3	8.7
Unknown	102	11.8	41.2	51.9	6.9
Gender					
Male	391	45.5	69.6	24.3	6.2
Female	454	52.8	62.3	29.3	8.3
Unknown	15	1.7	40.0	60.0	0
EVD status					
Confirmed	661	76.9	72.3	23.4	4.2
Probable	199	23.1	41.7	41.2	17.1
Number of reported infectors					
First generation of a chain	133	15.4	31.6	55.6	12.8
1 infector	690	80.2	71.0	22.8	6.2
2 infectors	16	1.9	68.8	25.0	6.2
3 infectors	21	2.5	85.7	9.5	4.8
Route of infection^a					
Household transmission	217	75.9			
Nosocomial transmission	30	10.5			
Funeral transmission	36	12.6			
Other transmission	74	26.2			
Outcome					
Survivor, ETU+	235	27.3	85.5	13.2	1.3
Non-survivor, ETU+	232	27.0	71.6	24.6	3.9
Non-survivor, ETU-, safe burial	63	7.3	57.1	38.1	4.8
Non-survivor, ETU-, unsafe burial	70	8.1	35.7	40.0	24.3
Unknown	260	30.2	51.2	37.3	11.5
Location					
Rural area	458	53.3	60.0	31.9	8.1
Urban area	402	46.7	71.1	22.6	6.2

ETU: Ebola treatment unit.

^aAs more than 1 route was specified for some cases, we did not compute the transmission status for this variable, 286 cases caused transmission. “Survivor, ETU+”: survivor who attended an ETU; “Non-survivor, ETU+”: Non-survivor who attended an ETU and was safely buried; “Non-survivor, ETU-, safe”: Non-survivor who did not attend an ETU and was safely buried; “Non-survivor, ETU-, unsafe”: Non-survivor who did not attend an ETU and was unsafely buried.

ORIGINAL UNEDITED MANUSCRIPT

Table 2:

Results of regression analysis.

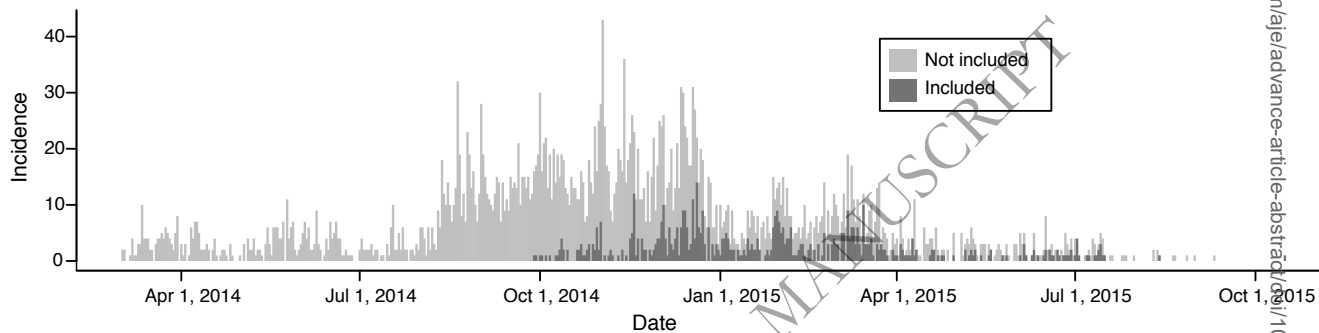
Variable	Secondary cases	IRR	95% CI	P-value
Intercept	0.69		0.38, 1.28	0.239
Gender				
Female		1	Referent	
Male		0.71	0.55, 0.93	0.012
Outcome				
Alive, ETU+		0.31	0.19, 0.51	$P < 10^{-3}$
Dead, ETU+, safe burial		0.62	0.38, 0.99	0.046
Dead, ETU-, safe burial		1	Referent	
Dead, ETU-, unsafe burial		1.82	1.10, 3.02	0.018
Location				
Urban		1	Referent	
Rural		1.18	0.90, 1.54	0.224
Age (years)				
0-14		0.35	0.21, 0.57	$P < 10^{-3}$
15-34		0.68	0.49, 0.93	0.015
35-54		1	Referent	
55-74		0.94	0.63, 1.40	0.757
75-99		1.47	0.55, 3.91	0.438
Generation number				
First generation		1.76	1.27, 2.44	0.001
Subsequent generations		1	Referent	

IRR: incident rate ratio.

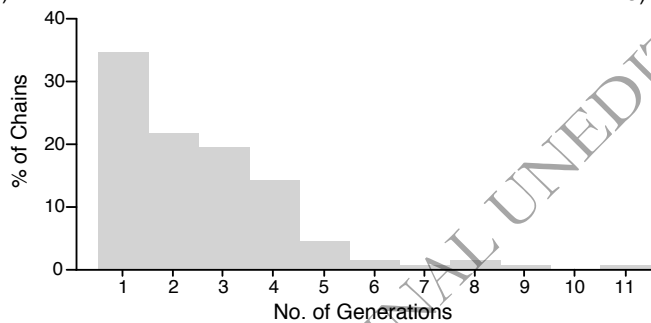
CI: Confidence interval.

ETU: Ebola treatment unit.

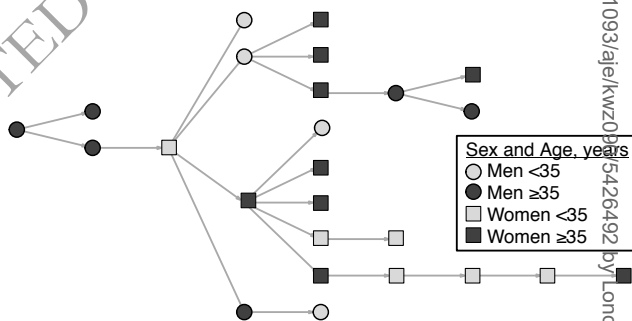
A)



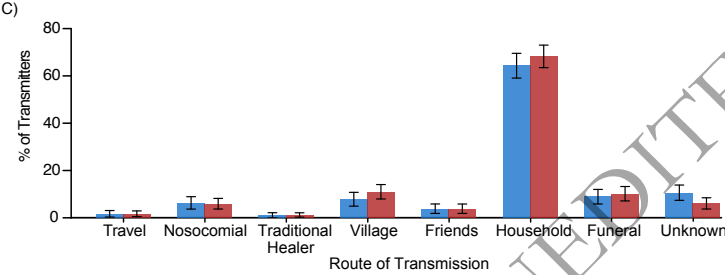
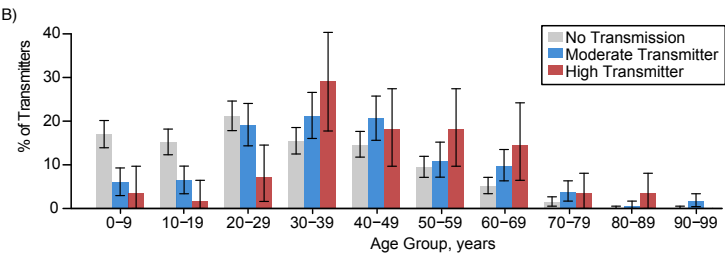
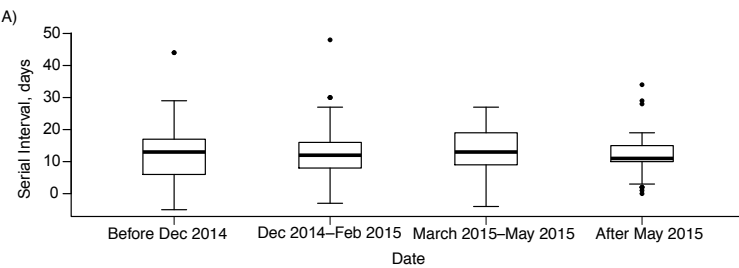
B)



C)

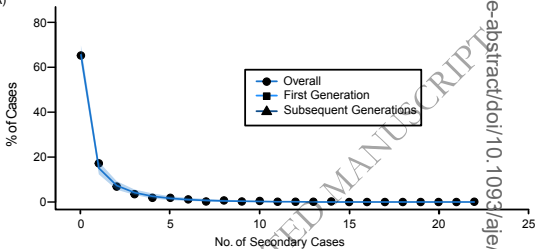


ORIGINAL UNEDITED MANUSCRIPT



ORIGINAL UNEDITED MANUSCRIPT

A)



B)

