Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic

Max S. Y. Lau^a,b, Benjamin Douglas Dalziel^c,d, Sebastian Funk^d, Amanda McClelland^e, Amanda Tiffany^f, Steven Riley^g, C. Jessica E. Metcalf^h, and Bryan T. Grenfell^i,h

^aDepartment of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; ^bDepartment of Integrative Biology, Oregon State University, Corvallis, OR 97331; ^cDepartment of Mathematics, Oregon State University, Corvallis, OR 97331; ^dCentre for the Mathematicial Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; ^eInternational Federation of Red Cross and Red Crescent Societies, CH-1211 Geneva 19, Switzerland; ^fEpicentre, CH-1211 Geneva 6, Switzerland; ^gMedical Research Council Centre for Outbreak Analysis and Modelling, Department Infectious Disease Epidemiology, Imperial College London, London SW7 2AZ, United Kingdom; and ^hFogarty International Center, National Institutes of Health, Bethesda, MD 20892

Edited by David Cox, Nuffield College, Oxford, United Kingdom, and approved January 5, 2017 (received for review September 8, 2016)

The unprecedented scale of the Ebola outbreak in Western Africa (2014–2015) has prompted an explosion of efforts to understand the transmission dynamics of the virus and to analyze the performance of possible containment strategies. Models have focused primarily on the reproductive numbers of the disease that represent the average number of secondary infections produced by a random infectious individual. However, these population-level estimates may confound important systematic variation in the number of cases generated by infected individuals, particularly found in spatially localized transmission and superspreading events. Although superspreading features prominently in first-hand narratives of Ebola transmission, its dynamics have not been systematically characterized, hindering refinements of future epidemic predictions and explorations of targeted interventions. We used Bayesian model inference to integrate individual-level spatial information with other epidemiological data of community-based (undetected within clinical-care systems) cases and to explicitly infer distribution of the cases generated by each infected individual. Our results show that superspreaders play a key role in sustaining onward transmission of the epidemic, and they are responsible for a significant proportion (~61%) of the infections. Our results also suggest age as a key demographic predictor for superspreading. We also show that community-based cases may have progressed more rapidly than those notified within clinical-care systems, and most transmission events occurred in a relatively short distance (with median value of 2.51 km). Our results stress the importance of characterizing superspreading of Ebola, enhance our current understanding of its spatiotemporal dynamics, and highlight the potential importance of targeted control measures.

Ebola | superspreading | offspring distribution | Bayesian inference

Spatial and temporal dynamics of superspreading

To whom correspondence should be addressed. Email: msylau@princeton.edu.

The authors declare no conflict of interest.

For many infections, some infected individuals transmit to disproportionately more susceptibles than others, a phenomenon referred to as “superspreading.” Understanding superspreading can facilitate devising individually targeted control measures, which may outperform population-level measures. Superspreading has been described for a recent Ebola virus (EBOV) outbreak, but systematic characterizations of its spatiotemporal dynamics are still lacking. We introduce a statistical framework that allows us to identify core characteristics of EBOV superspreading. We find that the epidemic was largely driven and sustained by superspreading that were ubiquitous throughout the outbreak and that age is an important demographic predictor for superspreading. Our results highlight the importance of control measures targeted at potential superspreaders and enhance understanding of consequences of superspreading for EBOV.

Significance


www.pnas.org/cgi/doi/10.1073/pnas.1614595114

PNAS Early Edition | 1 of 6

PNAS Early Edition | 1 of 6
study localized transmissions of EBOV and complement formal surveillance by detecting cases that did not interface with clinical care. In this work, we built an age-specific spatiotemporal framework, which allowed us to explicitly infer the probability distribution of the number of new cases generated by each infected individual (hereafter, offspring distribution). This framework was applied to the community-based EBOV case dataset and deployed to infer transmission dynamics and identify superspreaders. Specifically, we used Bayesian inferential techniques to synthesize individual-level spatial data (i.e., GPS coordinates), age data, symptoms onset time, and burial time (Study Data), and to impute unobserved infection time and transmission network (Materials and Methods and SI Text).

**Study Data**

We analyzed a community-based dataset collected from the Safe and Dignified Burials program conducted by the International Federation of Red Cross, between October 20, 2014, and March 30, 2015, in Western Area (which comprises the capital Freetown and its surrounding area) in Sierra Leone. These data contain GPS locations (collected by mobile phones) of where the bodies of 200 dead who tested positive for Ebola were collected (typically at their homes). Age, sex, time of burial (which was usually performed within 24 h of death), and symptom onset time were also recorded. Symptom onset time was reported retrospectively by next of kin.

**Results**

**Natural History Parameters.** We estimated that $R_0$ has median value 2.39, with 95% credible interval (C.I.) of [2.05, 2.84] (Fig. 1A). We also estimated the time-varying reproductive number $R_t$ (Fig. 1B). The incubation period was estimated to be 6.74 d [1.29, 16.21]. These estimates are broadly consistent with what have been reported (3, 12).

The mean of infectious period (i.e., duration from symptoms onset to death/burial) was estimated to be 3.9 d [3.75, 4.0]. Because the transmission tree and times of infection were imputed (Materials and Methods), we were also able to infer the mean generation time of EBOV, which was estimated to be 10.9 d [9.25, 13.01]. Both estimates were lower than that estimated from cases detected within the clinical care system [e.g., mean infectious period 8 d estimated for patients who received clinical care (13) and mean generation time 15.3 d estimated by the WHO (1)]. These discrepancies potentially highlight systematic differences between community-based cases and cases notified in clinical care systems, with terminal community-based cases progressing significantly more rapidly.

**Fig. 1.** Estimates of reproductive number. (A) Posterior distribution of the basic reproductive number, $R_0$. (B) Posterior distribution of the weekly effective reproductive number, $R_t$. Bars represent 95% C.I., and red line connects the medians.

**Fig. 2.** (A) Spatial distribution of mean number of offspring resulting from initial cases at the individual level. An infection is classified as an index case if it has a posterior probability of importation (i.e., not infected by any cases in the data) >0.5; otherwise, it is classified as a secondary case. Lat, latitude; Lon, longitude. (B) Distribution of mean number of offspring by different sources of infection. (C) Proportion of infected individuals who are direct and indirect descendants of the first five superspreaders (i.e., first five individuals with highest number of mean offspring; note that the choice of five is arbitrary here). “Any” includes superspreaders who were also the index cases (i.e., the roots of transmission trees).
Superspreading in Space and Time. Fig. 2. A and B show a clear asymmetry in the average number of “offspring” at the individual level, quantifying the impact of superspreading. In particular, it was observed that most secondary cases generated less than one offspring on average. Thus, the epidemic growth appeared to be fueled mostly by only a few superspreaders (i.e., the outliers in the boxplot). A common empirical measure of degree-of-transmission heterogeneity and superspreading is the dispersion parameter $k$, assuming that the offspring distribution is a binomial with variance $\sigma^2 = \mu(1 + \mu/k)$, where $\mu$ is the mean (9). Generally speaking, a lower $k$ represents a higher degree of transmission heterogeneity and superspreading; and $k < 1$ implies substantial superspreading (compared with a geometric distribution, for which $k = 1$). Our empirical estimate of $k$ of our inferred mean offspring distribution (including index and secondary) was 0.37, and it is higher (i.e., implies less heterogeneity) than an estimate from an observational study in which $k$ was estimated to be 0.16 (10, 11). This discrepancy in the estimate of $k$ suggests that our estimate of the degree of superspreading may be conservative (Sensitivity Analysis), although it should be noted our estimate was made based on a study in a different geographical region and time frame. By sampling probabilistically consistent transmission networks among infected individuals (Materials and Methods), we were able to identify whether a case was a descendent of superspreaders by performing a backward search of sampled transmission tree from the case—for each case, we first identified its (most recent) direct infector ($IF_1$) from the sampled tree, from where we could subsequently identify the infector of $IF_1$; We continued this backward searching until we reached an index case [i.e., the root of a (sub)tree]; a superspreader is an ancestor of this case if it happens to be one of the infectors during the backward searching. Fig. 2C shows that a few superspreaders (~3% of all of the cases) were responsible, either directly or indirectly, for a substantial proportion (with median 61%) of all of the cases generated, highlighting the key role of these superspreaders in driving the epidemic growth—had the superspreaders been identified and quarantined promptly, a majority of the infections could have been prevented.

In Fig. 3A, we show the time dependence of superspreading, illustrating that superspreading becomes relatively more important over time (i.e., within ~100 d after the epidemic peak). This figure suggests that, after the initial period of fast growth of the epidemic (i.e., time before peak), superspreaders may be crucial to sustaining and fueling epidemic growth and also prolonging the epidemic duration. Near the end of the epidemic (period 5 in Fig. 3A), most cases did not spread, and superspreading was non-significant, as reflected by $k > 1$. Fig. 3B shows that most of the transmission (including superspreading) occurred over relatively short distances (median 2.51 km), indicating that transmission tends to take place at the local community level.

Heterogeneity of Infectiousness by Age. Although superspreading in EBOV was evident and may be partly attributed to unsafe burial practice during the early stage of the outbreak (14), other drivers (e.g., social contact pattern) of this process remain unclear. In Fig. 4A, as expected, the infectious period had a clear positive relationship with mean offspring number. Despite the clear relationship between infectious period and the magnitude of superspreading, this covariate cannot be used as a predictor of superspreading, because it is not known a priori. More importantly, there is a significant difference in instantaneous infectious hazard exerted by different age groups (Fig. 4B)—cases <15 and >45 appear to have higher instantaneous transmissibility. Our
Sensitivity Analysis. Underreporting is a ubiquitous feature of epidemiological data (17, 18). In this section, we explore the effect of underreporting on our analysis under two probable scenarios: (i) All unreported cases were circulating in the community and not hospitalized; and (ii) all unreported cases were hospitalized and therefore not reported in our database. In both scenarios, we tested with constant underreporting rates, across the whole study period and region, ranging from a very low (10%) to a very high one (90%). Doing so allowed us to investigate the probable lower and upper bound of our estimates. We also tested with time-varying underreporting rates in both scenarios. Details of how to include underreported cases are provided in Materials and Methods.

We focused on investigating the effect on $k$, $R_0$ and transmission distance. Fig. 5A shows that, in general, superspread-
Discussion
Superspreading is a core process for the transmission of many infections (7, 8). However, the importance of superspreading in driving epidemics varies with context. For instance, its impact depends on how it persists over the course of an epidemic. Quantifying superspreading and identifying scenarios where it is more likely to occur can facilitate refining future epidemics predictions and help in devising targeted intervention strategies that may outperform population-level control measures (9). To date, a systematic understanding of how EBOV has been (super)spreading in the recent outbreak in Western Africa is lacking, particularly in terms of individual-level covariates, and across the spatiotemporal setting. The key contributions of this work are to highlight and quantify the importance of superspreading and to show that it is in some senses systematic.

Community-based surveillance data offer a valuable opportunity to study superspreading, by focusing on nonhospitalized cases that may have been involved in superspreading events and not detected by formal surveillance. Here, we introduce a continuous-time spatiotemporal model that integrates individual spatial information with other epidemiological information of community-based cases and deploy it to quantify superspreading and its drivers for EBOV. Our framework enabled us to sample likely realizations of the unobserved transmission network among cases from which the offspring distribution of each case could be inferred, providing explicitly a machinery for understanding superspreading in space and time.

Our analysis is broadly consistent with previous work, indicating values of $R_0$ of 2.39 [2.05, 2.84] for the outbreak in Sierra Leone (in particular, close to the 2.53 estimated in ref. 22). Our results show that EBOV exhibited a prominent superspreading pattern shared by SARS and MERS (7, 8, 23) [e.g., $k$ was estimated to be 0.16 for SARS (9)], which reinforces the finding that superspreading occurred during the recent EBOV outbreak (10).

We also extended previous analyses by showing that a substantial proportion of secondary cases were either direct or indirect descendants of a small number of superspreaders, underscoring the importance of superspreading in driving the epidemic—that is, had the superspreaders been identified and quarantined promptly, ~61% of the infections could have been prevented. Furthermore, we show that superspreaders may have particular importance in driving and sustaining the epidemic progression over the course of the outbreak. The increasing relative importance of superspreading over the later stages of the outbreak (Fig. 3A) is consistent with the rising availability of hospital beds—that is, later in the outbreak, most infected individuals were able to get a bed at an Ebola treatment center (ETC) and largely did not further transmit; as a result, those superspreaders in the community who did not make it to ETCS may have played an increasingly important role in sustaining the epidemic by generating more secondary cases. Our results also suggest that EBOV transmission may have disproportionately affected the local community, because we estimate a relatively short transmission distance. This estimated distance has implications for implementation of regional control measures. Identifying individuals who have the profile (socially or culturally) of being at greater risk of causing superspreading events is crucial for implementing targeted interventions.

We reveal that age-dependent social contact structure may play an important role in (super)spreading EBOV in the local community. Specifically, our results identify age groups that have higher instantaneous transmissibility and show that cases in the more infectious age groups tend to be superspreaders when combined with a relatively long infectious duration. One plausible explanation, from the social perspective, may be that the young and old are much more likely to have (and infect) lots of visitors, compared to other age groups; a parallel corollary is that the young and old might be more likely to have others caring for them. Also, our results highlight systematic differences between community-based cases and cases notified to the clinical care systems, with terminal community-based cases progressing significantly more rapidly. Our results stress the importance of characterizing superspreading of EBOV, enhance current understandings of its spatiotemporal dynamics, and highlight the potential importance of targeted control measures—for example, during the 2014–2015 EBOV epidemic, millions of dollars were spent implementing message strategies about Ebola prevention and control across entire countries; our results suggest that message strategies targeting individuals with higher risk may be useful to prevent superspreading events and the persistence of the outbreak.

There are limitations of our results. First of all, although community-based surveillance data complement formal surveillance by detecting cases that did not interface with clinical care, they contain only partial information about the epidemic, with hospitalized cases omitted. Also, it is possible that, by underreporting some community cases who generated subsequent cases, certain reported cases may be falsely attributed as sources of infection for those subsequent cases, overestimating the degree of superspreading. Accordingly, our sensitivity analysis evaluated the impact of these sources of underreporting, showing that our estimated degree of superspreading may in fact be conservative and represents a lower bound—superspreading in EBOV may be even more prominent in reality (Fig. 5). It is also worth noting that, by considering only safe burials, which tend to be less transmissible (relative to those that did not receive safe burials) among deaths (14), our estimate of superspreading may have been conservative. Conversely, because it was reported that individuals who eventually died might have a higher intrinsic transmissibility (24), our analysis might bias toward high transmitters by only using death data. Our methodology represents a transmission network-based approach that focused on constructing transmission trees among cases (25–28). Although such an approach captures contacts that caused infections, it does not account for “unsuccessful” contacts that correspond to escaped infections. Future theoretical work will need to include such contacts. Nevertheless, because unsuccessful contacts are not parts of the transmission chain, ignoring them has limited effect on the transmission tree or on many overall topological characteristics (e.g., average number of offspring of an infected case) (25, 28, 29). Finally, although our analysis reveals the importance of age as demographic determinants of superspreading, future work in linking them with virological factors (e.g., age-specific viral loads) may shed further light (16).

Materials and Methods

Spatiotemporal Transmission Model. We developed a continuous-time spatiotemporal transmission model that allowed us to sample the transmission tree among cases, integrating observed spatial and temporal individual data. This approach allowed us to infer explicitly the mean offspring distribution of each case. Specifically, the total probability of individual $j$ becoming infected during time period $[t, t + dt]$ was given by

$$r_j(t, dt) = \alpha \sum_{i \in I(t)} \beta_i \times K(d_{ij}, \eta) \cdot dt + o(dt),$$

where $i(t)$ is the set of all infectious individuals at time $t$, $\alpha$ is the background rate of infection, and $\beta_i$ is the age-specific instantaneous infection hazard of a case in $i(t)$. We allowed five-level $\beta_i$ according to the age—that is, we had $\beta_1 = \beta_a$ for age between $[0, 15)$, $\beta_2$ for age between $[15, 30)$, $\beta_3$ for age between $[30, 45)$, $\beta_4$ for age between $[45, 60)$, and $\beta_5$ for age $[60, \infty)$. We also tested alternative parameterization of priors in Table S3, giving virtually identical results compared with those obtained in the baseline case (see also Materials and Methods).
\(K(\theta; \eta)\), also known as a dispersal kernel, characterized the dependence of the infectious challenge from infectious to \(j\) as a function of distance \(d_{ij}\) between them. Here, we have \(K(\theta; \eta) = \exp(-\eta d_{ij})\). After the infection, it was assumed that individual \(j\) would go through an incubation period (i.e., time from infection to symptoms onset) and an infectious period (i.e., time from onset to death). The incubation period was assumed to follow a gamma distribution \(\Gamma(a, b)\) distribution (where \(a\) and \(b\) are mean and SD, respectively), and the infectious period followed an exponential distribution with mean \(c\). We assumed the infectiousness started from the symptoms onset time. It was noted that unknown contacts corresponding to escaped infections were not taken into account in our framework, resulting in a likelihood function that accounted for only successful infectious contacts (SI Text)—that is, our approach essentially represented a transmission network-based inference, where the focus was to construct the transmission tree among infected individuals (25–28).

**Data Augmentation and Model Fit and Validation.** We estimated \(\theta\) (i.e., the parameter vector) in the Bayesian framework by sampling it from the posterior distribution \(P(\theta|x)\), where \(x\) is the observed data. Denoting the likelihood by \(L(\theta; x)\), the posterior distribution of \(\theta\) is \(P(\theta|x) \propto L(\theta; x)\), where \(\pi(\theta)\) is prior distribution for \(\theta\). Weak uniform priors for parameters in \(\theta\) were used (Table S4). Markov chain Monte Carlo (MCMC) techniques (30) were used to obtain the posterior distribution. The unobserved infection times and transmission network were imputed in the MCMC. Sampled transmission networks were recorded and used to infer the offspring distribution of each case. Details of the likelihood function and the MCMC algorithm are given in SI Text. Model fit was assessed by comparing the observed data with those simulated from the estimated model, suggesting a good fit (Fig. S4). Furthermore, for validating the implementation of our inference procedures, we generated multiple sets of pseudodata from the model process and demonstrated that we could successfully reestimate the model parameters (Fig. S5).

**Testing Underreporting.** We divided the observational period into many 3-d-wide intervals. Within each time interval, we had the total number of unreported cases \(n_i = n_i^{(1)} - n_i^{(2)}\), where \(n_i^{(1)}\) and \(n_i^{(2)}\) were the observed cases in the interval and the assumed underreporting rate, respectively. Burial times and symptoms-onset time of these unreported cases were drawn from the empirical distribution of the observed cases. Finally, these \(n_i\) cases were distributed spatially by using the empirical distribution of (normalized) population densities across the study area. We also tested an underreporting rate that decreases with time (Fig. S1). For the scenario that considers unreported hospitalized cases, we drew the time from symptoms onset to hospitalization from the truncated above (at 7 d) empirical infectious period distribution of observed cases, effectively resulting in a shorter infectious period for unreported cases. These artificially generated data were combined with the observed data and fitted with our model.

**ACKNOWLEDGMENTS.** This work was supported by Bill & Melinda Gates Foundation Grant OPP1091919; the RAPID program of the Science and Technology Directorate Department of Homeland Security and the Fogarty International Center, National Institutes of Health, and the UK Medical Research Council (MRC). S.F. was also supported by MRC Career Award in Biostatistics MR/K021680/1.