SCIENTIFIC REPORTS

OPEN

Epidemic cholera spreads like wildfire

Manojit Roy[1,2], Richard D. Zinck[1]*, Menno J. Bouma[3] & Mercedes Pascual[1,2]

[1]University of Michigan, Department of Ecology and Evolutionary Biology, 2045 Kraus Natural Science Bldg., 830 N. University, Ann Arbor, MI 48109-1048, USA, [2]Howard Hughes Medical Institute, Chevy Chase, MD, USA, [3]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, University of London, London, UK.

SUBJECT AREAS:
ECOLOGICAL EPIDEMIOLOGY
POPULATION DYNAMICS
BACTERIAL INFECTION

Received
6 August 2013

Accepted
16 December 2013

Published
15 January 2014

Correspondence and requests for materials should be addressed to M.P. (pascual@umich.edu)

* Current address: BMW Car IT GmbH, Munich Labs, Petuelring 116, 80809 Munich, Germany.

Cholera is on the rise globally, especially epidemic cholera which is characterized by intermittent and unpredictable outbreaks that punctuate periods of regional disease fade-out. These epidemic dynamics remain however poorly understood. Here we examine records for epidemic cholera over both contemporary and historical timelines, from Africa (1990–2006) and former British India (1882–1939). We find that the frequency distribution of outbreak size is fat-tailed, scaling approximately as a power-law. This pattern which shows strong parallels with wildfires is incompatible with existing cholera models developed for endemic regions, as it implies a fundamental role for stochastic transmission and local depletion of susceptible hosts. Application of a recently developed forest-fire model indicates that epidemic cholera dynamics are located above a critical phase transition and propagate in similar ways to aggressive wildfires. These findings have implications for the effectiveness of control measures and the mechanisms that ultimately limit the size of outbreaks.

Cholera remains a public health threat in many countries around the world where outbreaks occur sporadically and punctuate periods of disease extinction or fade-outs[1]. This epidemic behaviour is characterized by dramatic variation in the size of individual outbreaks including large intermittent and unpredictable events. It corresponds to the so-called 'type III' epidemics in the literature on childhood diseases such as measles[2,3], and differs fundamentally from the better studied population dynamics of endemic cholera in foci of persistent disease with recurrent seasonal transmission[4–7]. Historically, epidemic cholera was common in the vast stretches of land surrounding the Ganges and Brahmaputra in former British India, upriver from the estuary and the homeland of the disease in Bengal[8,9]. In recent decades, epidemic cholera has appeared in Africa, South and Central America, and the Asian Subcontinent[1].

Although the population dynamics of some individual epidemics have been addressed[10–12], a characterization of ensemble properties for multiple events over time is lacking for any region. Consideration of the statistical properties of multiple events is of relevance to address what mechanisms determine and limit the size of outbreaks. In particular, the apparent stochasticity implied by disease fade-outs does not tell us *per se* whether cholera is above or below the 'emergence/elimination' threshold at which an infected individual replaces itself by less than one secondary infection on average. Similarly, it does not inform us on how the size of outbreaks will vary as we approach this threshold from below or above, as environmental, socio-economic or intervention conditions change in time. These questions are closely related to the current interest in critical transitions in nonlinear systems in general[13].

Results

To examine ensemble properties, we have analysed an extensive historical data set for cholera mortality from the districts of Punjab and Assam in former British India that span about six decades (1880–1939) (Fig. 1), as well as a compilation of contemporary weekly cholera cases reported to the World Health Organization (WHO) for countries in Africa from 1990 to 2006. We compute epidemic (or "event") sizes from the time series data by aggregating all case numbers (for Africa) or mortality numbers (for India) between consecutive fade-outs, with a fade-out defined as no reported infection, or no death, over a one month period. This duration is long enough to ensure that events are independent, in the sense of one event going extinct before the other is initiated (since the pathogen is known to rapidly loose culturability and to suffer predation by bacteriophages in aquatic environments, outside the human host[14], where its mean lifespan is typically considered less than a week[10,11]). Figure 1d illustrates the frequency distribution of epidemic sizes for one particular historical district. This distribution is fat-tailed, with the frequency of the larger epidemics over-represented relative to what would be expected for an exponential or a bell-shape distribution with a characteristic, most frequent, size. This pattern is similar to those described for the size distributions of natural disasters such as earthquakes and wildfires[15–17]. This similarity is

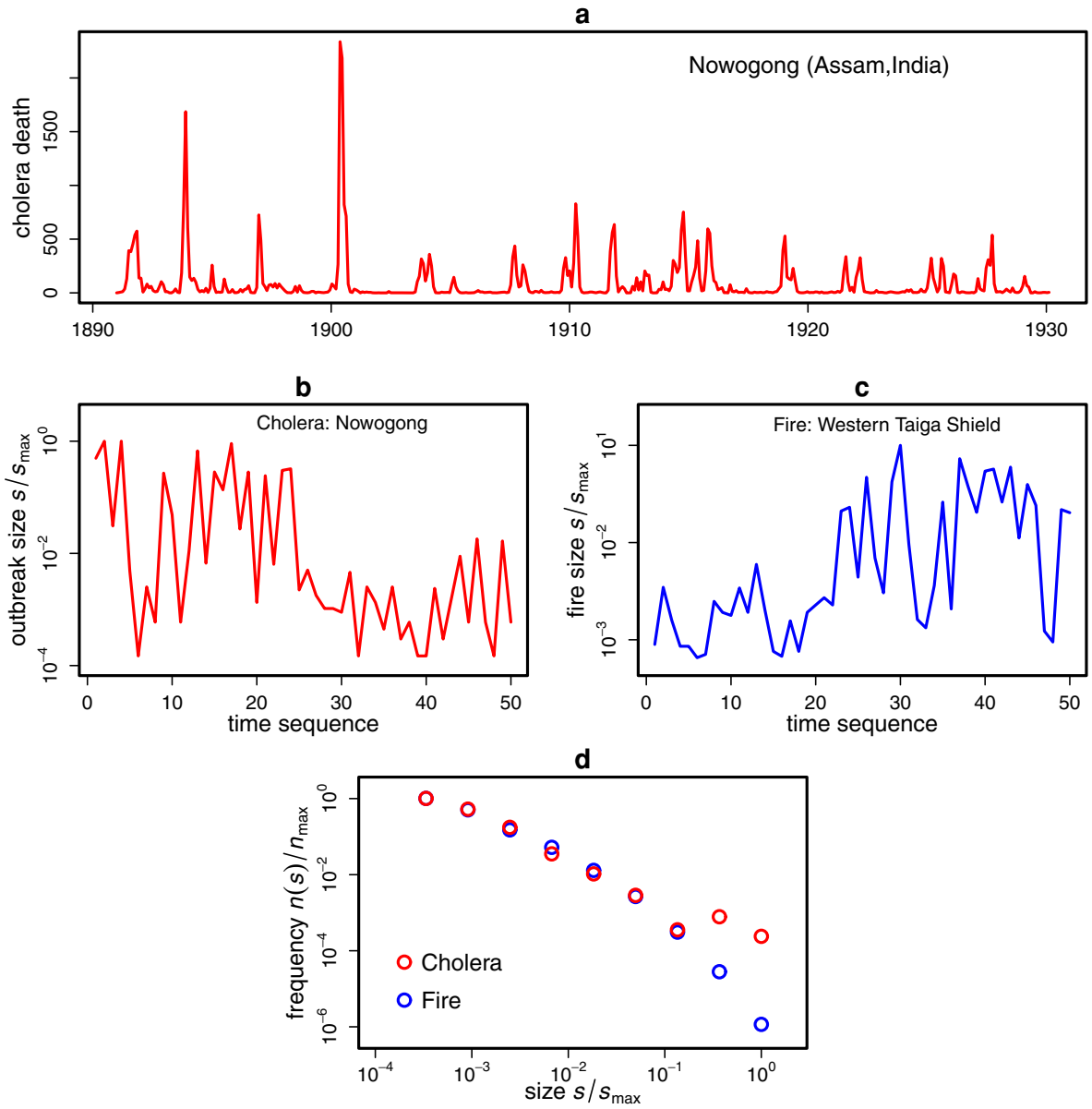

Figure 1 | Time series and size distributions exhibit similarities for cholera and wildfire. (a), Type III epidemic pattern in a representative time series for historical cholera mortality in the Nowogong district of Assam (India). (b & c), Outbreak sizes for the above cholera data and fire sizes from the Western Taiga Shield (Canada)[17] shown as time sequences. These are qualitatively similar. (d), The respective distributions for outbreak sizes and fire sizes show similar fat-tailed shapes, with comparable slopes. (The largest sizes typically deviate from the main distribution because of small samples).

particularly apparent when the size distribution for cholera epidemics is compared to that of fires for a region with aggressive wildfires such as the Boreal forest of the Northern hemisphere (Fig. 1d). Examination of these size distributions for the full data set (Fig. 2) reveals that there are regularities in these patterns across regions and historical periods, with outbreaks of all sizes over several orders of magnitude whose relative frequencies are not random but scale approximately as a power-law. A similar pattern characterizes the distribution of inter-epidemic intervals, although these range over fewer orders of magnitude (Supplementary Fig. S5).

This motivates us to better characterize these size distributions with a recently developed minimal forest-fire model that is both spatially explicit and stochastic[17]; this model represents an extension of the well-known Drossel-Schwabl model (hereafter DSM[18]). The original forest fire model was proposed in the context of self-organized criticality (hereafter SOC[19]) and applied to both wildfires in nature[15] and childhood diseases such as measles[20]. The extension we consider here was recently shown to better explain the observed variation in empirical size distributions of wildfires, in particular the values and full spread of the exponents of these power-law-like patterns observed across ecoregions[16,17].

In the model, transmission is localized by representing space as a two-dimensional lattice whose cells can be in one of three possible states: infected, recovered (and immune) and susceptible (non-immune). Disease dynamics is the result of three main processes affecting the state of these cells: the local birth-death process modelling the propagation of infection to near-neighbouring cells (at respective rates λ and μ), the random immigration of infection from outside the system (the injection or 'sparking' rate f), and the replenishment of susceptible individuals via new births or the loss of immunity (at rate p). Infection spreads at rate λ as long as the epidemic front has at least one non-isolated infected cell (one with susceptible cells in its local neighbourhood) (Methods). In the well-mixed or "mean-field" version of the model, the infected cell can infect a susceptible cell anywhere on the lattice. The temporal scales of these processes differ significantly, so that the spread of

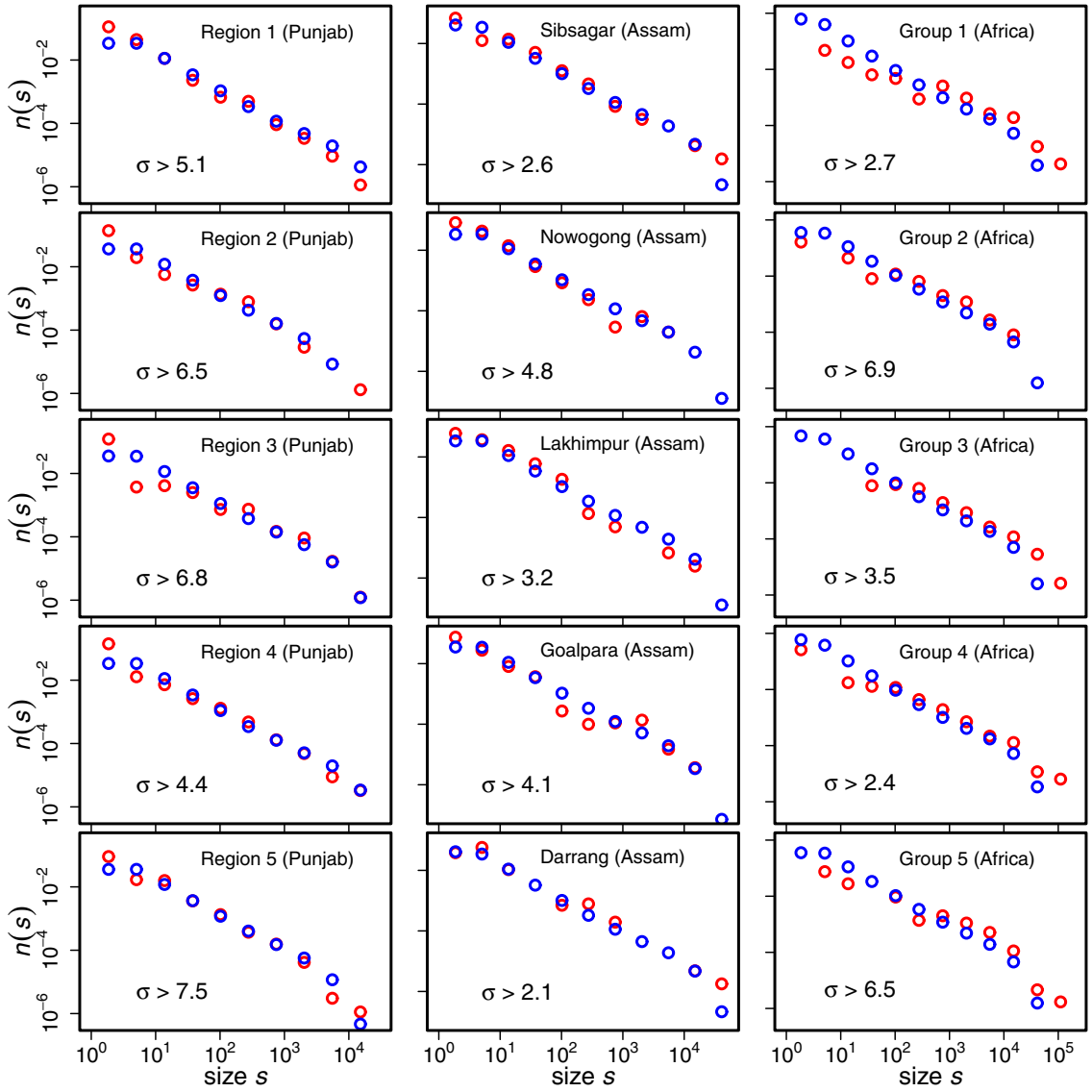

Figure 2 | The size distribution of cholera outbreaks is fat-tailed for both the historical and recent periods. Size distributions for the Punjab, Assam and Africa are shown respectively in the three columns from left to right (red circles). The size distributions obtained from the corresponding best-likelihood model are overlaid for comparison (blue circles). Lower bound estimates of the parameter σ place epidemic cholera in the highly contagious category for all data sets. See Methods for a description of the data.

infection occurs at a much faster rate than the re-growth of susceptibles, which in turn is much faster than the rate at which new infections enter the system. The model has effectively two parameters, $\sigma = \lambda/\mu$ corresponding to the (local) reproductive ratio $\mathbb{R}_0$ of the disease (up to a maximum value given by the neighbourhood size), and $\theta = p/f$ corresponding to the average susceptible replenishment number between consecutive immigration of infections. The critical value of $\sigma = 1$ separates two distinct dynamical regimes: (1) a sub-critical regime in which fade-outs are purely random and occur from demographic drift because the local spread of infection is too slow relative to recovery or death; and (2) a super-critical regime where epidemics are curtailed because infections have rapidly depleted their local pool of susceptible hosts.

We have fitted the model to determine how accurately it captures the empirical distribution of epidemic sizes and whether the dynamics are above or below the critical point, as well as their distance from this point (see Methods). For the model, the epidemic size is given by the number of infected cells as the epidemic progresses from start to end. Figure 2 illustrates the fit of the model to the epidemic size distributions for different locations in the full data set. The estimated value of σ for both the contemporary and historical data always exceeds 1, indicating a super-critical regime (Supplementary Table S1). This estimate is typically large ($\gg$1), corresponding to a region of parameter space for which the model behaves like the DSM with similar power law exponents (Fig. 3). Our maximum-likelihood estimation procedure can give us the lower bound of the confidence interval for σ but not a precise estimate itself, once the system is considerably above the critical point. This is because the size distribution changes little (Fig. 3), and so does the model likelihood (Supplementary Fig. S4), once σ is sufficiently far into the super-critical regime. A more precise estimate of a lower bound is possible with a larger number of epidemic events. To this end, we aggregated all regional events for each of the three data sets for Africa, Punjab and Assam (Supplementary Table S1); the aggregate estimates remain consistent with those of the component regions.

To further explore the potential implications of these empirical patterns for the dynamics transitioning towards, or away from, criticality[13], we summarize in figure 3 the different size distribution patterns exhibited by the model as the parameter σ is progressively increased from 1. A power-law emerges at the (2nd order) critical

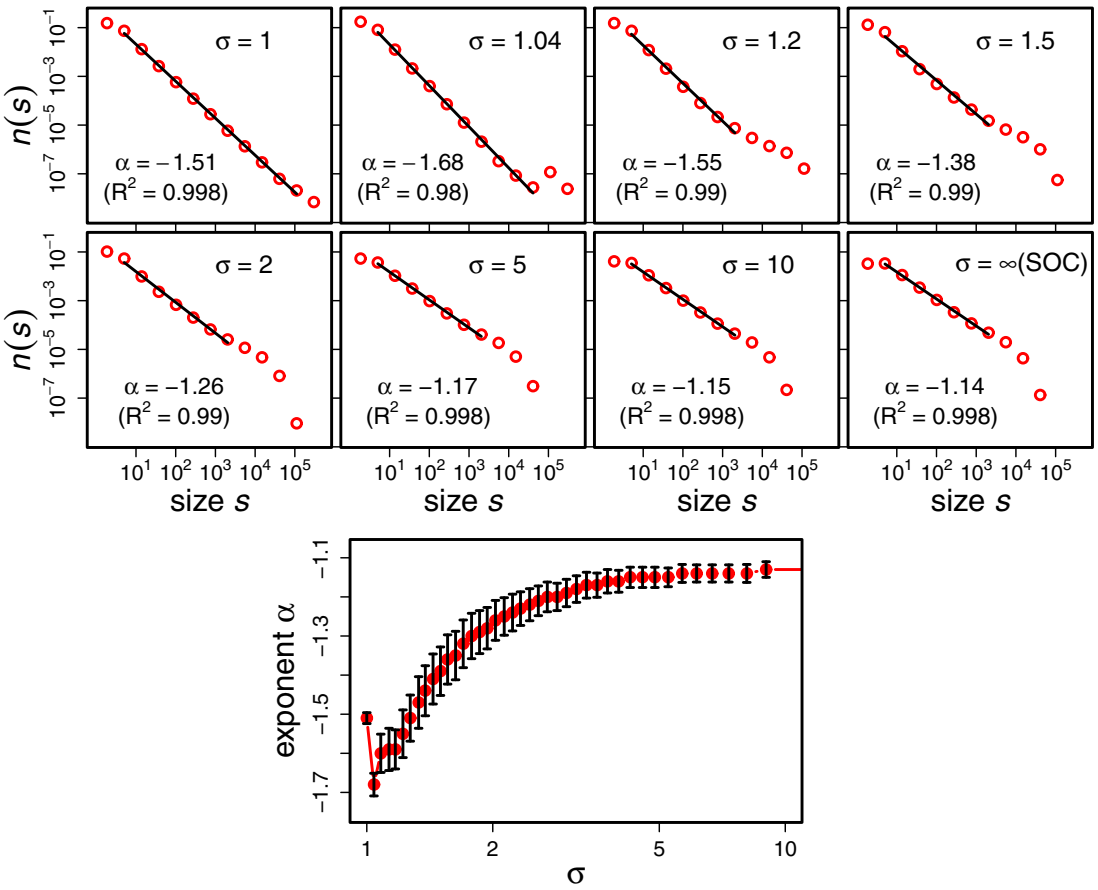

Figure 3 | Simulated size distributions from the model are fat-tailed for all values of σ above 1. Top panel, Size distributions simulated from the neighbourhood model are shown for increasing values of σ, from $\sigma = 1$ through ∞, together with the corresponding fit of the power-law function $n(s) \sim s^{-\alpha}$ (solid black line) (a finite-size scaling of the form $n(s) \sim s^{-\alpha}\exp(s/s_{max})$ applies to the distributions for $\sigma \geq 1.2$, where s_{max} denotes the largest outbreak size). The model bridges two critical phenomena, namely 2nd order criticality at $\sigma = 1$ (with $\alpha \approx -1.5$) and SOC at $\sigma = \infty$ (with $\alpha \approx -1.14$), and generates a spectrum of fat-tailed distributions for different values of a single parameter σ. Bottom panel, The plot shows the relationship between the estimated power-law exponent α and σ. As σ is lowered from ∞ towards 1, the exponent decreases slowly at first, and then more rapidly as it approaches the critical point $\sigma = 1$, undershooting the critical slope of -1.5 between $1 < \sigma \leq 1.3$.

phase transition at $\sigma = 1$ whose exponent equals –1.5. Below this value, the dynamics are essentially linear, with no depletion of local susceptibles; the event size distribution for this sub-critical regime shows a characteristic exponentially decreasing shape (Supplementary Fig. S6). An analytical expression for the size distribution of the birth-death process exists in the literature[21], which shows how the long-tail expands as σ approaches 1. In contrast, in the limit of infinite σ, the model reduces to the DSM with a power-law exponent –1.14. (A correction to this numerically observed DSM scaling has been proposed in the limit of very large size and θ[22,23], much larger than those considered here.) Hereafter we refer to the limit of infinite σ as the "SOC limit" of our model. Although models exhibiting the 2nd order phase transition and DSM have been independently studied quite extensively[17–23], the model presented here is to our knowledge the only one capable of bridging these two distinct phenomena by varying a single parameter, σ. This feature gives our model the ability to generate a family of size distributions that remain power-law like regardless of considerable variation in parameters.

The model's behaviour can be further characterized by comparison with that of its 'mean-field' counterpart, in which the locality of transmission is lost and an infection can randomly propagate to random sites throughout the lattice. For $\sigma = 1$, the neighbourhood and mean-field model exhibit as expected the same power law because spatial correlations result from the branching process itself[21] rather than from the neighbourhood interactions (Fig. 4a). As σ increases past 1 and enters the super-critical regime, these correlations quickly break down in the mean-field model. Because the infection can spread far and wide and in so doing, generate very large outbreak sizes quite distinct from the bulk of the distribution (of smaller outbreaks), the fat-tailed distribution breaks apart (Fig. 4b–d). By contrast, the spread of localized infection in the neighbourhood model can maintain spatial correlations at all scales, and the size distributions remain power-law like for all σ up to $\sigma = \infty$, the SOC limit (Figs. 3,4). These patterns do not depend on the number of neighbours (Supplementary Fig. S2): the spatial correlations determining the shape of these distributions remain unaffected for neighbourhood sizes much smaller than the size of the lattice. This observation indicates that the parameter estimates are insensitive to the particular neighbourhood used in the model as long as the condition relative to the total size of the system holds. It also follows that the well-mixed, temporal models developed for endemic cholera will fail to reproduce an important statistical property of empirical data in epidemic settings, the scaling evident in frequency distributions of epidemic size. Local spread and stochasticity appear essential to reproduce those patterns.

Discussion

To summarize, a minimal model of forest-fire dynamics captures the empirical patterns of cholera epidemics in widely separate regions, and suggests that epidemics spread like fire in the super-critical

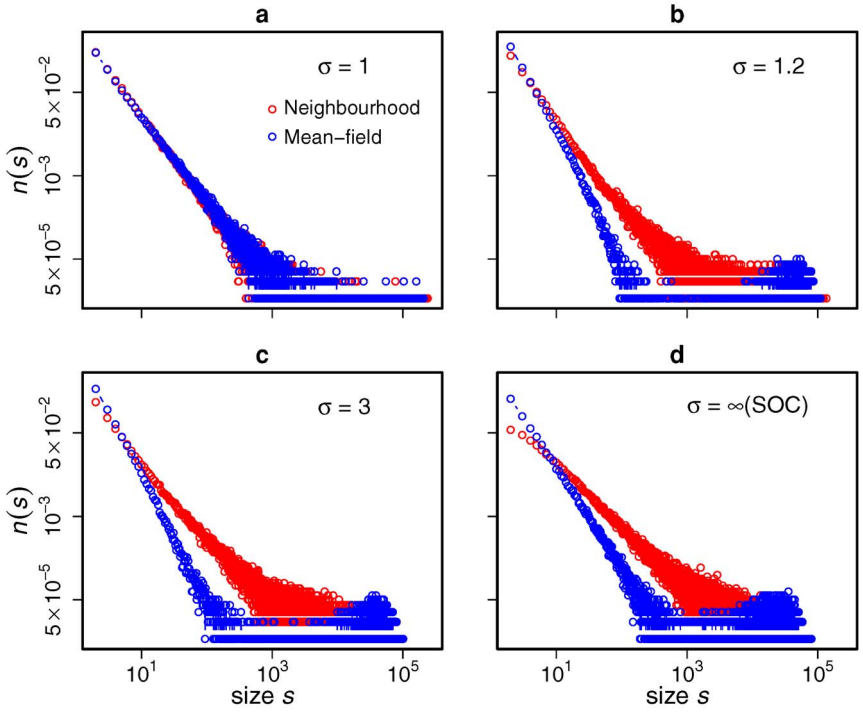

Figure 4 | Size distributions for the neighbourhood and mean-field models are compared. The outbreak size distributions for the neighbourhood (red circles) and mean-field (blue circles) model are compared for $\sigma = 1$, 1.2, 3 and ∞ (raw, or un-binned, distributions are shown to highlight differences). (a), Both models exhibit the same 2nd order criticality at $\sigma = 1$, and hence the two distributions coincide with identical slope $\alpha \approx -1.5$. (b–d), The slopes diverge for $\sigma > 1$: only the neighbourhood model continues to exhibit the fat-tailed distribution as σ is increased all the way to ∞; by contrast, the mean-field scaling quickly breaks apart into two distinct groups of small and large sizes (see text).

regime. This implies that the local depletion of susceptibles underlies the turnaround of epidemics, rather than demographic stochasticity as proposed with a model for a South African epidemic[24] or seasonality per se as in models for endemic cholera in Bangladesh[6,7]. This local depletion may result from local 'herd' immunity or even behavioural changes arising when individuals become aware of cholera's presence. An interesting consequence of the 2nd order criticality is that epidemics of all sizes tend to arise as the critical point $\sigma = 1$ is approached (Supplementary Fig. S1c). This implies that intervention efforts that push the system progressively towards the elimination boundary would increase the likelihood of extreme epidemic events.

Type III epidemics in isolated small populations were modelled previously with an elegant application of the DSM that was limited to the extremely fast (instantaneous) propagation of infection[20]. This resulted in the same size distribution (with a fixed exponent) regardless of the values of other parameters, as the system 'self-organized' to the same patterns characteristic of DSM (the $\sigma = \infty$ limit), a limitation that was resolved by invoking a different dimensionality of the transmission network[20]. At the opposite extreme of sub-critical behaviour, branching (birth-death) processes were previously applied to the 'stuttering' outbreaks of measles in London which have emerged with the progressive relaxation of vaccination[21]. Our model bridges these two regimes. As such, it can be applied to follow changes in σ over time as the result of intervention measures, and to anticipate emergence, or elimination, as an infectious disease approaches the critical point.

The analogy to forest fire dynamics is not limited to epidemics outbreaks in small individual populations in isolation[20]; it should apply more generally to epidemic infectious diseases in a metapopulation structure[25] depending on population size and distances from the populations acting as sources of infection. The key aspect would be the relative time scales of the three basic processes, and in particular how the introduction rate is influenced by the connectivity of the metapopulation network. In this sense, the heavy-tailed shape of the epidemic distribution in a large city may break down not because of the large population size *per se*, but due to the frequent movement among relevant spatial subunits, which would violate the condition of the double separation of time scales. An estimate of θ for an epidemic of another diarrhoeal disease, rotavirus, in Melbourne (Australia) suggests evidence for such an inadequate time-scale separation[26]. Investigation of the metapopulation dynamics of the model is underway, including the effect of hierarchical patterns of connectivity (within and between populations as a function of size) and associated patterns of synchrony.

Because of the way the data was collected, we have considered here cholera outbreaks assembled over large regions. Thus, the definition of an event cannot differentiate between a genuine single event and one that aggregates cases or mortality from two or more events that are sufficiently synchronous to overlap in time. A genuine event would be one that is initiated by a single 'spark' within the area under consideration. We have seen however that aggregation over different spatial extents does not alter our conclusions (Supplementary Fig. S7). Future work should seek to address better delineated spatial scales for individual events and consider the dynamics of the epidemic model in a regional metapopulation context. Data sets with higher spatial resolution would be invaluable to better understand how the stochastic dynamics of transmission are reflected in epidemic patterns, and how these patterns are themselves influenced by the spatial scale of aggregation. The effects of seasonality and climate induced extrinsic forcing of the birth-death process should also be addressed[27].

Our lower-bound estimates of σ suggest that the reproductive number of cholera tends to be considerably larger in epidemic than in endemic settings. For comparison, estimates of $\mathbb{R}_0$ reported in the literature for endemic cholera range from 1 to 17, with a majority of estimates not much larger than 1[28–30]. These estimates do not represent however measures of local spread, and the population-wide $\mathbb{R}_0$ can poorly apply to disease propagation in a spatio-temporal context

where it is neither necessary nor sufficient for its value to exceed 1 for an epidemic to occur[12]. Our estimates ($\gg$1) of the local $\mathbb{R}_0$ suggest that epidemic cholera not only spreads 'like fire' but does so like aggressive wildfires in nature. A raging wildfire in a similar regime cannot be easily brought under control and its final size largely depends on the exhaustion of available fuel. In the case of cholera, intervention saves lives, and as such, it should clearly be a priority. However, the final size of a cholera epidemic, once it is initiated, especially in less developed regions with a large vulnerable population, might be largely insensitive to intervention efforts. This observation reinforces the importance of ultimately controlling cholera, a well-known disease of poverty, by eliminating the conditions that make its regional persistence possible and its local reappearance likely. Treatment and vaccination during an epidemic, despite their clear value to individuals, are unlikely to achieve the well-known historical results of basic sanitation and access to clean water.

Methods

Data. The weekly data for epidemic cholera outbreaks in African countries during 1990–2006 has been assembled from the records of the World Health Organisation (http://apps.who.int/globalatlas/). The monthly district-wise cholera mortality data for former British India has been assembled from the Annual Sanitary (Public Health) Reports for the Provinces of Assam and Punjab. Because the number of epidemic events obtained from each of the data sets in Africa and Punjab is low, we consider ensembles of events from several geographically contiguous regions to generate size distributions, and obtain more reliable statistics (Supplementary Fig. S3). This grouping makes the implicit assumption that outbreaks are uncorrelated in space, in the sense that each of them in a given district or country starts with a single introduction of infection from outside the system, rather than from multiple initiation points generating concurrent epidemics in time within the given area (see Model formulation). The grouping also considers that aggregation over different or randomized spatial subunits within a region does not affect the shape of the distributions, an assumption we have verified in our analyses (Supplementary Fig. S7).

Specifically for Africa, five regional groups are defined by aggregating epidemic events as follows: Group 1 comprises Burundi, Malawi, Uganda, Rwanda, United Republic of Tanzania, Somalia, Kenya, Zanzibar; Group 2 comprises Chad, Sudan, Mali, Niger, Dijibouti, Burkina Faso; Group 3 comprises Mozambique, Swaziland, Zambia, Zimbabwe, Madagascar, Comoros; Group 4 comprises Sierra Leone, Liberia, Togo, Benin, Gambia, Cote de Ivoire, Ghana, Senegal, Guinea Bissau, Nigeria, Mali, Niger, Mauritania, Burkina Faso; and Group 5 comprises Congo, Central African Republic, Zaire, Gabon, Cameroon, Democratic Republic of Congo, Equatorial Guinea, Sao Tome and Principe, Angola. Similarly for Punjab, five regions are considered by aggregating epidemic events from different districts as follows: Region 1 comprises Attock, Rawalpindi, Jhelum, Gujrat, Shahpur, Mianwali; Region 2 comprises DGKhan, Mooltan, Muzaffaragarh, Jhang, Lyallpur, Montgommery; Region 3 comprises Gujranwala, Sialkot, Sheikhupura, Lahore, Amritsar, Gurdaspur; Region 4 comprises Ferozpore, Jullundur, Hoshiarpur, Kangra, Ludhiana, Hissar; and Region 5 comprises Gurgaon, Delhi, Rohtak, Karnal, Simla.

Model formulation. Each cell of the two-dimensional grid can be in one of three possible states, namely, infected, recovered (and immune) or susceptible (non-immune), and are labelled as I, R or S respectively. Infection is introduced randomly in the system at a slow rate f, and can initiate an epidemic only if it hits a susceptible cell. Empty cells become susceptible at a rate $p(\gg f)$, which ensures substantial susceptible growth between consecutive introductions. The infection spreads to susceptible cells locally, among 4 or 8 near neighbours, at a rate λ, and an infected cell recovers at a rate μ. The rates λ and μ are chosen such that $p \ll T^{-1}(s)$, where $T(s)$ is the duration of a typical outbreak of size s, which further ensures that susceptibles do not reappear until the epidemic dies out. The parameter $\sigma = \lambda/\mu$ corresponds to the localized basic reproductive ratio $\mathbb{R}_0$. The double separation of time scale $f \ll p \ll T^{-1}$ is similar to the necessary condition for generating fat-tailed size distributions in SOC models[17–19]. Our model additionally incorporates a finite propagation rate of the epidemic via rates $\lambda, \mu < \infty$, whereas in SOC the propagation is instantaneous; it reduces to the SOC model in the limit $\sigma \to \infty$.

Model implementation. All I cells are kept in a queue and processed sequentially in the order they are added to the queue (random processing of the queue does not affect the conclusions presented here). A random number in (0,1) is drawn to decide whether the first I cell in the queue recovers (I $\to$ R transition) with probability $\mu/(\lambda + \mu)$, or whether a random S neighbour of this I cell becomes infected (S $\to$ I transition) with probability $\lambda/(\lambda + \mu)$. If there is no S neighbor, the next I cell in the queue is chosen and so on, until one infection occurs. The recovered I cell is removed from the queue, and the newly created I cell is added to the end of the queue, with these steps repeated until the last I cell in the queue recovers (i.e. the epidemic dies out). We present results obtained with the 4-neighbor model (the 8-neighbor model exhibits similar patterns; Supplementary Fig. S2).

Model simulation. Simulations were run on a grid of $L^2 = 500 \times 500$ cells and rigid boundary conditions are used. (For this large grid size, the alternative choice of periodic boundaries does not affect the outcome of the simulations.) A neighbourhood size of 4 (Von Neumann neighbourhood) is used throughout, unless stated otherwise (as in Supplementary Fig. S2). Once the densities reached a stationary state, 100,000 epidemics were sampled to obtain the frequency distribution for each combination of $\sigma = \lambda/\mu$, $\theta = p/f$. Size distributions were constructed by binning the outbreak size data according to (log-equally spaced) size classes. The values for σ were selected such that $\lambda + \mu = 1$ with λ bounded between (0.4,1) in steps of 0.01 ($\lambda = 1$ corresponds to $\sigma = \infty$, which is the SOC limit for the model). For the SOC regime, larger grids require longer transients to reach the stationary scaling[31]; we therefore calibrated the 500×500 grid for transient length by making sure the model exhibits SOC scaling at $\sigma = \infty$ (Figs. 3, 4d) after 100,000 epidemics, and used this transient length for all simulations with this grid size. A further calibration is needed for the parameter $\theta = p/f$ because a finite-size effect relates the grid size to the inoculation rate f. For a given growth rate p and grid size, a larger f and hence smaller θ will trigger outbreaks before susceptibles are sufficiently numerous to develop adequate spatial connectivity, resulting in the absence of large cluster sizes in the distribution[31]. Similarly for a given θ, smaller grids will effectively raise f by increasing the inoculation risk of a random S cell, and trigger outbreaks prematurely. For our model with a 500×500 grid, we found $\theta \geq 750$ to be the range that exhibits SOC scaling at $\sigma = \infty$, and used a representative value $\theta = 1000$ for all model simulations in figures 3, 4. However, for fitting the data sets (fig. 2), θ is allowed to move freely, and is estimated together with σ. The values of θ are well-determined by the data as indicated by narrow confidence intervals (Supplementary Table S1). The maximum-likelihood estimates indicate a clear separation of time scales for the processes of inoculation from outside (at rate f) and susceptible regrowth (at rate p) for both Assam and Africa. The separation is less pronounced for the Punjab data sets, with values around 100 that effectively constrain the tail of the distribution and potentially reflect a smaller population (Supplementary Table S1).

Parameter estimation and model validation. For fitting the cholera data with our model, we estimated the parameter σ together with θ (Supplementary Table S1). This was done by maximizing the log likelihood, $\log L = \sum_i n_i \log p_i$, of the model given the data, where n_i is the number of observed epidemics in bin i and p_i is the probability that an epidemic falls in that bin based on the model[17]. The uncertainty range of σ was calculated as the 95% confidence interval via likelihood profiling, with $\log L(\sigma_\mp) = \log L(\sigma) \mp 1.92$[32] (and similarly for θ).

The estimation procedure described above presupposes the model is a valid description of the data, which may not always be true[33]. We therefore independently addressed model validity by computing a goodness-of-fit using the KS (Kolmogorov-Smirnov) statistic $D = \max_{s \geq s_{\min}} |C_{\text{data}}(s) - C_{\sigma,\theta}(s)|$, where $s_{\min}$ denotes the smallest outbreak size ($\geq$5) considered for the distribution, and C_{data} and $C_{\sigma,\theta}$ refer to the CDF (cumulative distribution function) of event sizes for the data and model[33]. This quantity provides a measure of the distance between the size distribution from the data and that from a particular model (given σ and θ). We estimated the p-value that the model cannot be excluded given the data, by first generating a distribution of D values by using C_{sample} in place of C_{data} (in the expression) for a large number of samples from the simulations, and then comparing the D value based on the empirical data to this distribution. The model is valid at $p \geq 0.05$ level if the data D falls within 5–95% of the distribution. For each cholera data set, we obtained the p-value of a series of models with different σ (and θ fixed at the respective MLE; see Supplementary Table S1). This yields a lower bound of σ below which the model is excluded (with $p \leq 0.05$); this bound is consistent with the lower bound for the MLE of σ (not shown).

1. Harris, J. B., LaRocque, R. C., Qadri, F., Ryan, E. T. & Calderwood, S. B. Cholera. *Lancet* **379**, 2466–2476 (2012).
2. Bartlett, M. S. The critical community size for measles in the U. S. *J. R. Stat. Soc. A* **123**, 37–44 (1960).
3. Grenfell, B. T., Bjørnstad, O. N. & Finkenstädt, B. F. Dynamics of measles epidemics: scaling noise, determinsism, and predictability with the TSIR model. *Ecol. Monogr* **72**, 185–202 (2002).
4. Pascual, M., Rodó, X., Ellner, S. P., Colwell, R. & Bouma, M. J. Cholera dynamics and El Nino-Southern oscillation. *Science* **289**, 1766–1769 (2000).
5. Rodó, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. ENSO and cholera: a nonstationary link related to climate change? *Proc of the Natl Acad Sci U S A* **99**(**20**), 12901 (2002).
6. Koelle, K., Rodo, X., Pascual, M., Yunus, M. & Mostafa, G. Refractory periods and climate forcing in cholera dynamics. *Nature* **436**, 696–700 (2005).
7. King, A. A., Ionides, E. L., Pascual, M. & Bouma, M. J. Inapparent infections and cholera dynamics. *Nature* **454**, 877–881 (2008).
8. Glass, R. I. *et al.* Endemic cholera in rural Bangladesh, 1966–1980. *Am J Epidemiol* **116**, 959–970 (1982).
9. Longini, I. M. *et al.* Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *J. Infect. Dis.* **186**, 246–251 (2002).
10. Bertuzzo, E. *et al.* On the space-time evolution of cholera epidemic. *Water Resour Res* **44**, W01424 (2007).
11. Rinaldo, A. *et al.* Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections. *PNAS* **109**, 6602–6607 (2012).

12. Gatto, M. *et al.* Generalized reproduction numbers and the prediction of patterns in waterborne disease. *PNAS* **109**, 19703–19708 (2012).
13. Scheffer, M. *et al.* Anticipating critical transitions. *Science* **338**, 344–348 (2012).
14. Nelson, E. J., Harris, J. B., Morris Jr, J. G., Calderwood, S. B. & Camilli, A. Cholera transmission: the host, pathogen and bacteriophage dynamics. *Nat. Rev. Microbiol.* **7**, 693–702 (2009).
15. Malamud, B. D., Morein, G. & Turcotte, D. L. Forest fires: an example of self-organized critical behavior. *Science* **281**, 1840–1842 (1998).
16. Malamud, B. D., Millington, J. D. A. & Perry, G. L. W. Characterizing wildfire regimes in the United States. *PNAS* **102**, 4694–4699 (2005).
17. Zinck, R. D., Pascual, M. & Grimm, B. Understanding shifts in wildfire regimes as emergent threshold phenomena. *Am Nat* **178**, E149–E161 (2011).
18. Drossel, B. & Schwabl, F. Self-organized critical forest-fire model. *Phys. Rev. Lett.* **69**, 1629–1632 (1992).
19. Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality. *Phys. Rev. A.* **38**, 364–374 (1988).
20. Rhodes, C. J., Jensen, H. J. & Anderson, R. M. On the critical behaviour of simple epidemics. *Proc. R. Soc. Lond.* **B 264**, 1639–1646 (1997).
21. Jansen, V. A. *et al.* Measles outbreaks in a population with declining vaccine uptake. *Science* **301**, 804 (2003).
22. Pastor-Sartorras, R. & Vespignani, A. Corrections to scaling in the forest-fire model. *Phys. Rev. E* **61**, 4854–4859.
23. Grassberger, P. Critical behaviour of the Drossel-Schwabl forest fire model. *N. J. Phys.* **4**, 17.1–17.15 (2002).
24. Azaele, S., Marien, A., Bertuzzo, E., Rodriguez-Itrube, I. & Rinaldo, A. Stochastic dynamics of cholera epidemics. *Phys. Rev. E* **81**, 051901 (2010).
25. Keeling, M. & Rohani, P. *Modeling infectious diseases in humans and animals.* (Princeton University Press, 2008).
26. Lara-Sagahón, A., Govezensky, T., Méndez-Sánchez, R. A. & José, M. V. A lattice-based model of rotavirus epidemics. *Phys. A* **359**, 525–537 (2006).
27. Pueyo, S. Self-organized criticality and the response of wildland fires to climate change. *Climatic Change* **82**, 131–161 (2007).
28. Hartley, D. M., Morris Jr, J. G. & Smith, D. L. Hyperinfectivity: a critical element in the ability of V. cholerae to cause epidemics? *PLoS Med* **3**, e7 (2006).
29. Mukandavire, Z. *et al.* Estimating the reproductive numbers for the 2008–2009 cholera outbreaks in Zimbabwe. *PNAS* **108**, 8767–8772 (2011).
30. Mukandavire, Z., Smith, D. L. & Morris Jr, J. G. Cholera in Haiti: reproductive numbers and vaccination coverage estimates. *Sci. Rep* **3**, 997 (2013).
31. Turcotte, L. Self-organized criticality. *Rep. Prog. Phys.* **62**, 1377–1429 (1999).
32. Hilborn, R. & Mangel, M. *The ecological detective.* (Princeton University Press, 1997).
33. Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Rev* **51**, 661–703 (2009).

Acknowledgments

We thank Andy Dobson for comments on the manuscript. This work was supported by a Centennial fellowship of the J.S. McDonnell foundation and a grant from NOAA (Oceans and Health Program NA08NOS4730321). M.P. is an investigator of the Howard Hughes Medical Institute.

Author contributions

M.R., R.Z. and M.P. conceived this work and drafted the paper. M.R. and R.Z. implemented the model, conducted the simulations and analysed the data. All authors contributed to the interpretation of the results and the final writing of the paper. M.B. assembled, and provided expertise on, the historical records.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Roy, M., Zinck, R.D., Bouma, M.J. & Pascual, M. Epidemic cholera spreads like wildfire. *Sci. Rep.* **4**, 3710; DOI:10.1038/srep03710 (2014).