Quantifying the seasonal drivers of transmission for Lassa fever in Nigeria

Andrei R. Akhmetzhanov, Yusuke Asai and Hiroshi Nishiura

Graduate School of Medicine, Hokkaido University, Sapporo, Hokkaido, Japan
ARA, 0000-0003-3269-7351; HN, 0000-0003-0941-8537

Lassa fever (LF) is a zoonotic disease that is widespread in West Africa and involves animal-to-human and human-to-human transmission. Animal-to-human transmission occurs upon exposure to rodent excreta and secretions, i.e. urine and saliva, and human-to-human transmission occurs via the bodily fluids of an infected person. To elucidate the seasonal drivers of LF epidemics, we employed a mathematical model to analyse the datasets of human infection, rodent population dynamics and climatological variations and capture the underlying transmission dynamics. The surveillance-based incidence data of human cases in Nigeria were explored, and moreover, a mathematical model was used for describing the transmission dynamics of LF in rodent populations. While quantifying the case fatality risk and the rate of exposure of humans to animals, we explicitly estimated the corresponding contact rate of humans with infected rodents, accounting for the seasonal population dynamics of rodents. Our findings reveal that seasonal migratory dynamics of rodents play a key role in regulating the cyclical pattern of LF epidemics. The estimated timing of high exposure of humans to animals coincides with the time shortly after the start of the dry season and can be associated with the breeding season of rodents in Nigeria.

This article is part of the theme issue ‘Modelling infectious disease outbreaks in humans, animals and plants: approaches and important themes’.

1. Introduction

Growing human population, urbanization and global warming increase the chance of human interaction with wildlife, resulting in elevated risk of zoonotic diseases. Despite numerous publications assessing the risk factors for viral spillovers, published studies have tended to miss out quantitative estimation of their frequency and characteristics and causal links among different host species. In this study, we provide quantitative analysis of external forcing of infection, as applied to Lassa haemorrhagic fever (LF), a widespread disease in West Africa. The proposed modelling framework can be further extended to other diseases with common characteristics and viewed as key in light of anticipated emergence of a novel infectious disease in the future—the ‘Disease X’ as it was recently coined by the World Health Organization.

LF virus was first discovered in 1969, but its presence can be traced back for centuries [4]. Current viral hotspots are focused on West African countries, namely, Nigeria, Guinea, Benin, Sierra Leone and Liberia. The animal reservoir of the
Using the first and second parts of the model, we also captured the seasonality in the numbers of infected rodents. The difficulties of collecting data in Nigeria (see the discussion in [18]) and the laboratory settings [15–17] were predicted to reach their maximal level in the first two to three months of the year, exceeding the baseline by approximately four times. This phenomenon can be captured by implementing a time-varied probability of exposure in our model. Two distinct time periods characterize a low and high risk of exposure to the virus. Importantly, the model fit allows the probable time boundaries of each period to be determined.

Published estimates of the case fatality risk (CFR) from LF of 1–2% have been proposed for previously healthy populations without underlying comorbidities [8]. However, this rate varies considerably depending on the context, being for example 2–5% for hospital-treated cases and increased to 20–60% for laboratory-confirmed cases or during nosocomial outbreaks [2,9,10]. Because higher estimates of the CFR are likely due to underreporting and ascertainment bias, we are able to account for this observational matter and estimate an underreporting factor. Similarly to our recent work [11], we assume that fatal cases are certainly reported in the surveillance system throughout the whole year, while less severe, non-lethal cases are likely to be missed. Such variation may especially be the case during low-risk periods, when LF infections are not frequently reported.

The establishment and maintenance of a surveillance system inevitably face many difficulties in the case of LF reporting. A clinical diagnosis of LF infection is often challenging owing to the similarities with other common diseases seen in the region, i.e. malaria or typhoid fever [12,13]. On average, four out of five cases are asymptomatic [8]. Furthermore, post-mortem examinations to confirm LF infection as the cause of death are recognized as taboos in some areas of Nigeria [1]. All of these aforementioned factors contribute to the limited accuracy and completeness of the passively reported surveillance data.

Here, we offer a unified epidemiological model that consists of two parts [14]. The first part describes the process of generating the incidence data in humans. We derived the risk of exposure to the virus throughout the year. The second part employs the so-called susceptible–infected–recovered (SIR) model, which captures the transmission dynamics of the virus in rodents. Based on existing observations of rodent populations both in the field and in laboratory settings [15–17] (see the discussion on the difficulties of collecting data in Nigeria in [18]), we have captured the seasonality in the numbers of infected rodents. Using the first and second parts of the model, we also estimated the relative contact frequency of humans with infected animals throughout the year.

Our study encompasses the analysis of surveillance data on LF incidence in humans along with the transmission dynamics in the rodent reservoir. Using this modelling framework, we aimed to estimate the impact of environmental factors on the annual fluctuations in the risk of LF infection in humans. Similar methodological applications can be performed for other zoonotic diseases that involve well-identified wildlife animal hosts.

2. Results

First, we estimated two distributions from existing datasets in humans, i.e. the incubation period and the time from illness onset to death. We used a model that rests on the renewal process of the viral exposure in humans, accounting for the time delay due to the incubation period and the time from illness onset to death. The formulation resulted in two Poisson- and binomially distributed likelihood functions for describing the LF incidence and mortality, respectively (electronic supplementary material, appendix B). We identified two distinct periods of a year which were separately analysed owing to considerably different levels of observed transmission activity: (i) a low-risk period with a weekly exposure rate of 6.4 (CrI: 0–32.5; maximum-likelihood estimate (MLE): 9.6) lasting from week 9 (26 February to 4 March in 2018, CrI: 5–13) to week 50 (10 December to 16 December in 2018; CrI: 47–52), and (ii) the rest of the year, characterized by a higher risk of contracting the virus, with the average weekly exposure rate reaching 24.7 (CrI: 0–111.7; MLE: 38.1), which exceeded the low-risk period by approximately four times. The CFR was estimated at 4.9% (CrI: 0–54.4; MLE: 8.9%). If we assume the unbiased value of the CFR at 2%, the reporting coverage can be estimated to be as low as 40%. The overlaid incidence data and variations in the parameter estimates are shown in figure 1. For comparison, an analogous model with a single (constant) exposure rate yielded a less plausible fit to the data (Akaike information criterion (AIC) values: 3679.2 for a single- and 2877.2 for a two-period model).

Subsequently, we assessed whether a seasonal increase in the number of infected rodents can solely explain the increase in the incidence of LF infections in humans in the high-risk period. Using the dynamic transmission model for LF infection in rodents (Material and methods section; electronic supplementary material, figure S2), the number of infected animals was predicted to reach its maximal level in the middle of May, while the minimal level was observed in December (electronic supplementary material, figure S3). The LF incidence in humans which is thought to correlate
with the LF level of infection in rodents controversially peaks in the first two to three months of the year. During that period, the number of infected rodents was smaller than for the rest of the year. We therefore propose that a factor other than seasonal fluctuations in the number of infected rodents may play a role in driving LF seasonal epidemics.

To search for such a factor, we analysed the correlative power of the LF incidence rates and climatological variables, such as rainfall, temperature, relative humidity, specific humidity and precipitable water. We found that only the rainfall seasonal pattern was significantly correlated with the seasonal incidence of LF in humans \( p < 0.01 \), figure 2). This was also confirmed by the start and end times of the high/low exposure periods, which approximately coincide with the start and the end of the rainy season (cf. figure 1b; electronic supplementary material, figure S4). Therefore, an event associated with rainfall patterns contributes to the incidence of LF in humans, e.g. this could be the change in seasons and seasonal migration or a change in the behaviour of the rodents.

Last, we calculated the relative frequency of contact of humans with infected rodents using the ratio of the estimated weekly exposure rate to the density of infected rodents predicted by our transmission model (figure 3). After assigning the average frequency of contact in the low-risk exposure period as 1, we determined that the relative risk of human exposure was 5.3 times higher on average during the high-risk exposure period. The maximal value of 6.7 was identified in the first week of the year.

3. Discussion

This study analysed the surveillance-based incidence data of LF in Nigeria from 2016 to 2018 and identified two different (high- and low-)risk periods. The high-risk period spanned the last month of the year to the second to third month of the year. The relative risk of acquiring LF infection during the high-risk period was five times greater than during the rest of the year. In our search for a possible explanation, we identified that the rainfall pattern was negatively and highly correlated with LF incidence. The rainfall does not affect the transmissibility of the virus directly, but it is noteworthy that rodents migrate to within close proximity to human settlements to breed and hibernate during the dry season. This in turn leads to an increase in the contact rate of humans with rodents, and as a consequence, a higher probability of acquiring LF infection.

Fichet-Calvet & Rogers [22] previously showed that the rainfall pattern was the main single abiotic factor contributing to LF. The migration of rodents was notified as a main factor in another report [23]. Our findings confirm the importance of the combined effect of these two factors on the seasonality of LF epidemics and offer a quantitative estimation of their strength, which has never to our knowledge been done in the past. When the rainy season ends in November, the breeding season for rodents starts about two months later. At that time, newly born offspring and scarcity of food on the ground force mature rodents to approach human-occupied areas. Consequently, this may lead to a rise in the
contact frequency between humans and infected rodents. This high-exposure frequency persists until the rainy season starts again the following year, at which point the rodents migrate back to the ground and we observe the subsequent decline in human cases of LF. These findings highlight the importance of seasonal ecology of animal hosts in explaining the seasonality of LF epidemics. Similar findings have been reported for many other diseases, not only for those of an epizootic nature, e.g. [24,25], but also for insect-vectored plant diseases [26]. This points to rather general applicability of our approach that may frame the spillovers of other pathogens with two epizootic transmission routes. An important achievement of the present study was that we were able to offer quantitative estimates of the impact of the two risk factors, rainfall pattern and migration of rodents, on the exposure probability of humans to LF infection.

This highlights the importance for preventive measures that aim to contain seasonal epidemics of LF to be specifically designed: (i) to control rodent populations and reduce the encounter frequency of rodents and humans, and (ii) to raise awareness among local residents. This can be envisioned as an eradication campaign, especially in rural areas with agronomic activities, and in public markets in urban areas, where rodents are frequently seen. Preventive measures may also include improved hygiene practices, hiding food from rodents during the night time, or designing educational campaigns that raise awareness of LF pathogenicity. The implementation of such programmes would be expected to lead to a decline in LF cases in the near future.

However, we were unable to measure the impact of awareness of the scale of epidemics using the available data. Even though occasional peaks in the number of recorded deaths may be indicative of periods of low awareness, our analysis of seasonal trends only showed that awareness probably does not involve a strict seasonal component. Nevertheless, the effect of awareness on the reporting of LF incidence may be indirectly evident from the geographical distribution of LF cases over the last 6 years (electronic supplementary material, figure S5). The geographical area with newly reported LF cases has expanded in the last few years, presumably because of better recognition of the epidemics and improved capability of the surveillance system.

Several technical limitations of our study must be noted. First, we did not consider a spatial component in our analysis which may involve geographical heterogeneity in reporting rates. Second, we used only the counts of suspected cases that included both laboratory-confirmed cases and cases that tested negative. However, the available dataset indicates that the majority of suspected cases involved specimens that tested negative. This becomes especially evident in low-risk exposure periods, and taking this into consideration in our analysis would further amplify the difference between high-risk and low-risk periods. Third, we did not account for temporal changes in the surveillance system and improvements in laboratory facilities for the detection of LF, both of which have been greatly improved in recent years. Fourth, we did not distinguish possible variations in population densities of rodents in habiting human houses and farm fields or, similarly, urban and rural areas.

In summary, we identified and quantified the effect of two factors that drive the seasonality of LF epidemics in Nigeria. Combining the fit of the mechanistic model to the observed counts of LF disease in humans with the predictive model of the transmission dynamics of LF in rodents, we quantified the annual change in relative contact frequency between humans and infected rodents. A high seasonal amplitude was identified, and our model indicated the first nine weeks of the season as the high-risk period for LF transmission.

4. Material and methods
   (a) Data collection

Data were routinely collected from the weekly epidemiological report of the Nigerian Centre for Disease Control (NCDC) for the period from 2016/week 4 to 2018/week 30. The extracted counts included newly suspected weekly cases for both positively and negatively testing specimens, as well as the number of fatal cases reported per week. For both cases and deaths, the corresponding week of the report represents the week in which the illness onset and death events occurred, respectively. (i.e. fatal cases were not reported as a function of the week of illness onset). We did not use the available data for previous time periods (2012–2016) owing to irregular reporting and removal.

Figure 2. Cross-map causality for shared seasonality of environmental variables with the Lassa fever incidence. Red circles show the unlagged cross-map skill. Box-plots show null distributions for cross-map skill expected from random surrogate time series that share the same seasonality as the true environmental variable. The single filled circle indicates that the measured causality is significantly better than the null expectation ($p < 0.01$). As for the use of convergent cross-mapping, please see literature [20,21]. (Online version in colour.)

Figure 3. Contact frequency of humans with infected rodents. Dashed vertical lines show the time boundaries separating the high-risk period from the low-risk period. The average contact frequency in the low-risk exposure period was set to 1. Solid black line indicates the median estimate, whereas light and dark shaded areas indicate 95 and 50% credible intervals for posterior estimates, respectively. (Online version in colour.)
of reports from an official website of the NCDC. However, we provide all of the accumulated data as electronic supplementary material.

To access weekly data on climatological variables (i.e. precipitation, temperature, relative and specific humidity, precipitable water), we used publicly available gridded NCEP/NCAR Reanalysis data [27,28]. The chosen reference point was set to coordinates 6.75° N 6.25° E, located in Edo state, Nigeria. Historically, Edo state was characterized by a high prevalence of LF infection. It was also marked as highly hazardous for LF transmission owing to an average annual rainfall of 1786 mm [22]. Additionally, we used monthly data records over a longer time period (1901–2015) for the same location that were provided by another source: a publicly available dataset of the University of East Anglia Climatic Research Unit [29]. While the former was used for correlative analysis of rainfall patterns and available incidence data from the last 6 years, the latter was used only to generate the historical characteristics of rainfall, as shown in electronic supplementary material, figure S4.

(b) Estimation of the incubation period and the time from illness onset to death

We fitted the distributions of the incubation period \( f \) and the time from illness onset to death \( h \), which were essential for describing the epidemiological process, to gamma distributions. The analysed dataset included the cases and transmission events during a nosocomial outbreak in the Evangel Hospital in Jos, Nigeria, in 1970 (fig. 1 in [19]). In total, the number of described cases was 23 with probable time intervals of exposure, the date of illness onset and the date of death. To determine the best-fit parameters for both distributions, we applied a Markov chain Monte Carlo (MCMC) method in a Bayesian framework. The 95% credibility interval (CrI) for each fitted parameter of the two distributions was determined as the 95% high-density interval [30].

(c) Epidemiological modelling

First, we accounted for two transmission routes of LF infection, i.e. the animal-to-human route and the human-to-human transmission route. We specified a total exposure rate in weeks \( w \) by a variable \( e_w \). This could be written as the sum of the two rates: \( e_w = a_w + h_w \), where \( a_w \) and \( h_w \) are weekly exposure rates via animal-to-human and human-to-human transmission routes, respectively. An important estimate previously reported in the literature [6] indicated that ca 19% of all infections are attributed to the human-to-human transmission route, leading to the following requirement: \( h_w (a_w + h_w) = r = 0.19 \). We can then express the total exposure rate \( e_w \) solely through the exposure rate \( a_w \) such that: \( e_w = a_w (1 - r) \). This expression was used in our framework to avoid a detailed modelling of the human-to-human transmission route. Owing to limited human-to-human transmission potential, we focused only on exposure to the virus through a contaminated environment.

Qualitatively, it has been noted that the incidence of infections is likely to rise in the first two to three months of each calendar year. We accounted for this in our model by incorporating seasonal variations in the weekly exposure rate \( a_w \) as a function of the calendar week \( w_k \). In its simplest form, the exposure rate \( a_w \) was modelled by a two-level step function:

\[
a_w(a_z, l_1z) = \begin{cases} 
  a_z, & \text{for } l_1z 
  l_2, 
  a_z, & \text{otherwise,}
\end{cases}
\]

where \( l_1z \) represents two time boundaries separating a period of high exposure from a period of low exposure (1 ≤ \( l_1z \) ≤ 52). We expected that \( a_z > a_z \).

Because all records in our dataset counted the weekly numbers \( w \) beginning from the first record, we assumed a functional dependence \( a_w = a_w(w) \) in the following formulations.

These assumptions allowed us to define the epidemiological process in a concise form. Here, two Poisson processes are considered. First, each newly reported non-fatal case was dealt with as a result of previous exposure delayed by the length of the incubation period. This implied a Poisson process with the expected number of new cases \( i_w \) at week \( w \) to be a result of the convolution of the exposure rate \( a \) and distribution of the incubation period \( f \), multiplied by the chance of survival \((1 - q)\) with \( q \) being the risk of death (or CFR):

\[
E(i_w | \theta) = (1 - q) \sum_{k=1}^{w-1} \frac{a_k (a_w, l_1z) \times k}{1 - r},
\]

where \( \theta = \{a_z, q, l_1z \} \) is a set of model parameters.

In a similar manner, we introduced another Poisson process to describe the reporting of fatal cases. LF deaths result from exposure to the virus in previous weeks postponed by a probabilistically distributed time period between the exposure event and death. The latter was denoted by \( g \) and defined as a convolution of the incubation period \( f \) and the distribution of time from illness onset to death \( h \). Then the expected number of fatal cases \( d_w \) in week \( w \) was determined by a convolution of the exposure rate \( a \) and distribution \( g \), multiplied by the risk of death \( q \):

\[
E(d_w | \theta) = q \sum_{k=1}^{w-1} \frac{a_k (a_w, l_1z) \times k}{1 - r}.
\]

Then, the likelihood function for describing the total number of reports (i.e. cases plus deaths) was:

\[
L_i (\theta | \{i_w + d_w\}) = \prod_w \left( \frac{E(i_w + d_w | \theta)^{iw + dw} \exp(-E(i_w + d_w | \theta))}{(i_w + d_w)!} \right),
\]

where \( \{i_w + d_w\} \) is a set of all available data records with a given independently reported incidence that includes both non-fatal and fatal cases, \( E(i_w + d_w | \theta) = E(i_w | \theta) + E(d_w | \theta) \). The second likelihood function for describing the number of deaths was derived from the binomial sampling process:

\[
L_d (\theta | \{i_w + d_w\}) = \prod_w \left( \frac{(i_w + d_w)^{i_w} d_w^{d_w}}{E(i_w + d_w | \theta)^{i_w}} \right) \left( \frac{1 - E(d_w | \theta)}{E(i_w + d_w | \theta)} \right)^{d_w},
\]

where the terms in the first set of parentheses in the product denote the binomial coefficient, \( \{d_w\} \); on the left-hand side is a set of all available data records for weekly counts of deaths.

We first defined the model parameters \( \theta \) by using a Bayesian approach and performing MCMC iterations. Second, we used the maximum-likelihood method to obtain point estimates. For the maximization procedure, we used the total (composite) likelihood of the form:

\[
L_2 (\theta | \{i_w + d_w\}) = L_i (\theta | \{i_w + d_w\}) L_d (\theta | \{i_w + d_w\}).
\]

The likelihood \( L_2 \) was maximized, respectively, to each parameter in \( \theta \) by equating its first partial derivatives to zero.

(d) Transmission dynamics of LF infection in rodents

To describe the transmission dynamics of LF infection in rodents, we used an SIR model (electronic supplementary material, figure S3 and more details in appendix C). The birth rate of rodents was estimated as model parameter estimates previously derived in another field study on the same rodent species in Tanzania [16,32]. To adjust for the differences in climatological profile between the two countries, we shifted the dynamics

\[
\text{Phil. Trans. R. Soc. B} \quad 374, \quad 20180056
doi:10.1098/rspb.2018.0056
fitted to the Tanzania data by a time lag equal to the difference between the starting times of the dry season in the two countries. Specifically, we defined this time lag by comparing historical rainfall averages for the time period 1901–2015 (electronic supplementary material, figure S4). We then implemented known measurements of the LF prevalence in rodents captured in Nigeria in the dry and rainy seasons separately, and we accounted for a non-zero probability of vertical transmission of the virus and its antibodies, as previously observed in rodents (see [16] for details).

(e) Association of seasonal LF dynamics and climatological variables

To test for the significance of a causal relationship between the observed LF incidence and climatological variables, we employed an empirical dynamic modelling that is based on the convergent cross-mapping skill [20,21]. This method has previously been used to demonstrate the causal link between influenza transmission risk and humidity or temperature. The LF incidence was tested, searching for the significance of the association with one of the climatological variables, i.e. time series versus seasonal surrogates. The threshold of acceptance was chosen to be 0.01.

Data accessibility. We provide all of the accumulated data as electronic supplementary material.

Authors’ contributions. Conception of the study: A.R.A. and H.N.; model formulation: A.R.A., Y.A. and H.N. Data analysis and interpretation: A.R.A. Drafting the manuscript: A.R.A.; comment on the early version of the manuscript: A.R.A., Y.A. and H.N.

Competing interests. We declare that we have no conflict of interest.

Funding. H.N. was financially supported by the JSPS KAKENHI (grant numbers 16K10130, 17H04701, 17F05808, 16H06429 and 16K21723), Japan Agency for Medical Research and Development (JP18fk0108050) and the Japan Science and Technology Agency (JST) CREST Program (JPMJCR1413). The funders had no role in study design, data collection and analysis, preparation of the manuscript or the decision to publish.

Acknowledgements. We thank Kate Fox, DPhil, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

1. Khan SH et al. 2008 New opportunities for field research on the pathogenesis and treatment of Lassa fever. Antiviral Res. 78, 103–115. (doi:10.1016/j.antiviral.2007.11.003)
2. Yun NE, Walker DH. 2012 Pathogenesis of Lassa fever. Viruses 4, 2031–2048. (doi:10.3390/v4102031)
3. Andersen KG et al. 2015 Clinical sequencing uncovers origins and evolution of Lassa virus. Cell 162, 738–750. (doi:10.1016/j.cell.2015.07.020)
4. Fichet-Calvet E, Blischlager S, Strecker T, Kövögul L, Becker-Ziaja B, Camara AB, Soropogui B, Magassouba N, Günther S. 2016 Spatial and temporal evolution of Lassa virus in the natural host population in Upper Guinea. Sci. Rep. 6, 21977. (doi:10.1038/srep21977)
5. Childs J, Mills JN, Glass GE. 1995 Rodent-borne hemorrhagic fever viruses: a special risk for mammalogists? J. Mammal. 76, 664–680. (doi:10.2307/1382739)
6. Lo Iacono G et al. 2015 Using modelling to disentangle the relative contributions of zoonotic and anthroponotic transmission: the case of Lassa fever. PLoS Negl. Trop. Dis. 9, e3398. (doi:10.1371/journal.pntd.0003398)
7. Baize S et al. 2014 Emergence of Zaire Ebola virus disease in Guinea. N. Engl. J. Med. 371, 1418–1425. (doi:10.1056/NEJMoa1405455)
8. Richmond JK. 2003 Lassa fever epidemiology, clinical features, and social consequences. BMJ 327, 1271–1275. (doi:10.1136/bmj.327.7426.1271)
9. Ingebenede U, Ökösné J, Ingebenede J. 2010 Prevention of Lassa fever in Nigeria. Trans. R. Soc. Trop. Med. Hyg. 104, 51–54. (doi:10.1016/j.trstmh.2009.07.008)
10. Asogun DA et al. 2012 Molecular diagnostics for Lassa fever at Iruna Specialist Teaching Hospital, Nigeria: lessons learnt from two years of laboratory operation. PLoS Negl. Trop. Dis. 6, e1839. (doi:10.1371/journal.pntd.0001839)

11. Asai Y, Nishiusu H. 2017 Joint estimation of the transmissibility and severity of Ebola virus disease in real time. J. Biol. Syst. 25, 587–603. (doi:10.1142/S0218339017400022)
12. Akhmerokhan O et al. 2017 Prevalence of Lassa virus disease (LVD) in Nigerian children with fever or fever and convulsions in an endemic area. PLoS Negl. Trop. Dis. 11, e0005711. (doi:10.1371/journal.pntd.0005711)
13. Okohkere P et al. 2018 Clinical and laboratory predictors of Lassa fever outcome in a dedicated treatment facility in Nigeria: a retrospective, observational cohort study. Lancet Infect. Dis. 18, 684–695. (doi:10.1016/S1473-3099(18)30121-X)
14. Lo Iacono G et al. 2016 A unified framework for the infection dynamics of zoonotic spillover and spread. PLoS Negl. Trop. Dis. 10, e0004957. (doi:10.1371/journal.pntd.0004957)
15. Borremans B, Reijniers J, Hughes NK, Godfrey SS, Gyseels S, Makundi RH, Leirs H. 2017 Nonlinear scaling of foraging contacts with rodent population density. Oikos 126, 792–800. (doi:10.1111/oik.03623)
16. Borremans B, Reijniers J, Hess N, Leirs H. 2017 The shape of the contact–density function matters when modelling parasite transmission in fluctuating populations. R. Soc. Open Sci. 4, 171308. (doi:10.1098/rsos.171308)
17. Fichet-Calvet E, Becker-Ziaja B, Kövögul L, Günther S. 2014 Lassa serology in natural populations of rodents and horizontal transmission. Vector-Borne Zoonotic Dis. 14, 665–674. (doi:10.1089/vbz.2013.1484)
18. Swanepoel LH et al. 2017 A systematic review of rodent pest research in Afro-Malagasy small-holder farming systems: are we asking the right questions? PLoS ONE 12, e0174554. (doi:10.1371/journal.pone.0174554)
19. Carey DE, Kemp GE, White HA, Pinneo L, Addy RF, Fom ALMD, Stoth G, Casals J, Henderson BE. 1972 Lassa fever epidemiological aspects of the 1970 epidemic, Jos, Nigeria. Trans. R. Soc. Trop. Med. Hyg. 66, 402–408. (doi:10.1016/0035-9203(72)90271-4)
20. Sugihara G, May R, Ye H, Hsieh C-H, Doyle E, Fogarty M, Munch S. 2012 Detecting causality in complex ecosystems. Science 338, 496–500. (doi:10.1126/science.1227079)
21. Doyle ER, Maher MC, Hernandez RD, Busa S, Sugihara G. 2016 Global environmental drivers of influenza. Proc. Natl Acad. Sci. USA 113, 13081–13086. (doi:10.1073/pnas.1607747113)
22. Fichet-Calvet E, Rogers DJ. 2009 Risk maps of Lassa fever in West Africa. PLoS Negl. Trop. Dis. 3, e388. (doi:10.1371/journal.pntd.0000388)
23. Marien J, Koeuromu F, Magassouba N, Leirs H, Fichet-Calvet E. 2018 Movement patterns of small rodents in Lassa fever-endemic villages in Guinea. EcoHealth 15, 348–359. (doi:10.1007/s10393-018-1331-8)
24. Hampson K, DuShoff J, Cleaveland S, Haydon DT, Kaare M, Packer C, Dobson A. 2009 Transmission dynamics and prospects for the elimination of canine rabies. PLoS Biol. 7, e1000553. (doi:10.1371/journal.pbio.1000553)
25. Dudas G, Carvalho LM, Rambaut A, Bedford T. 2018 MERS-CoV spillover at the camel-human interface. elife 7, e32127. (doi:10.7554/eLife.32127)
26. Yang Y, Huang M, C. Beattie GA, Xia Y, Ouyang G, Xiong J. 2006 Distribution, biology, ecology and transmission of D. pseudotalpae (MERS-CoV spillover at the camel-human interface). eLife 7, e31257. (doi:10.7554/eLife.31257)
27. Yang Y, Huang M, C. Beattie GA, Xia Y, Ouyang G, Xiong J. 2006 Distribution, biology, ecology and control of the pseudovirus D. pseudotalpae in camels. BMJ 334: 20180268
28. Kalyan E et al. 1996 The NCEP/NIH 40-year Reanalysis Project. Bull. Am. Meteorol. Soc. 77, 437–471. (doi:10.1175/1520-0479(1996)077
29. NOAA. 2017 NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. Boulder, CO: Physical Sciences Division, NOAA National Center for Environmental Prediction. See
29. University of East Anglia Climatic Research Unit, Jones PD, Harris IC. 2008 Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena. Didcot, UK: NCAS British Atmospheric Data Centre. See https://catalogue.ceda.ac.uk/uuid/3f8944800cc48e11bc29a5ee12d8542d (accessed August 2017).

30. Hyndman RJ. 1996 Computing and graphing highest density regions. Am. Stat. 50, 120 – 126. (doi:10.1080/00031305.1996.10474359)

31. Peel AJ et al. 2014 The effect of seasonal birth pulses on pathogen persistence in wild mammal populations. Proc. R. Soc. B 281, 20132962. (doi:10.1098/rspb.2013.2962)

32. Sluydts V, Davis S, Mercelis S, Leirs H. 2009 Comparison of multimammate mouse (Mastomys natalensis) demography in monoculture and mosaic agricultural habitat: implications for pest management. Crop Prot. 28, 647 – 654. (doi:10.1016/j.cropro.2009.03.018)
Appendix for
Quantifying the seasonal driver of transmission for Lassa fever in Nigeria
Andrei R. Akhmetzhanov, Yusuke Asai, Hiroshi Nishiura

Contests

Supplementary figures

A. Analysis of a nosocomial outbreak in Jos, Nigeria, in 1970
   Reconstructing the timeline of the outbreak
   Inference of model parameters using the Bayesian approach
   MCMC iterations

B. Model-based inference using the human case (incidence) data
   Maximum likelihood estimation
   Fitting procedure using the Bayesian approach
   MCMC iterations

C. Dynamic modelling of LF transmission in rodent populations
   Model formulation
   Periodic conditions for the change in population size
   Susceptible-infected-recovered (SIR) model
   Inference of model parameters

D. Computer simulations and code sharing

Appendix references

Appendix figures

Appendix code snippets
**Fig S1:** Estimated gamma probability density functions for the incubation period \( f \) and the time from illness onset to death \( h \) in humans. Solid black line indicates the median estimate, whereas light and dark shaded areas indicate 95% and 50% credible intervals for posterior estimates, respectively. Background light-red bins on the bottom panel indicate the available data counts.
| Parameter                                      | Variable | Value     | Reference                        |
|------------------------------------------------|----------|-----------|----------------------------------|
| Average population size                        | $N$      | 60 H$^{-1}$ | lower estimate available in [16] |
| Scaling parameter of a birth pulse             | $\kappa$ | 17.9 year$^{-1}$ | [31] $^*$                        |
| Birth pulse synchrony                          | $s$      | 1.26      | estimated                        |
| Phase of birth pulse                           | $\phi$   | 1.71      | [16]                             |
| Mortality rate                                 | $\mu$    | 7 year$^{-1}$ | [16]                             |
| Probability of vertical transmission of the virus | $\varepsilon$ | 50%      | [17]                             |
| Probability of vertical transmission of antibodies to the virus | $\omega$ | 20%      | [17]                             |
| Recovery rate                                  | $\gamma$ | 4.06 year$^{-1}$ | [17]                             |
| Rate of losing the immunity                    | $\lambda$ | 3.04 year$^{-1}$ | [17]                             |
| Rate at which $S$ becomes $I$ once in contact with $I$ | $\nu$ | 16.9      | estimated                        |

Fig S2: Summary of the modelling framework for Lassa fever transmission dynamics in rodents: (a) model flow diagram; (b) birth-rate function; (c) description of model parameters. The total population of size $N$ consists of susceptible ($S$), infected ($I$), and recovered ($R$) rodents. See Appendix Section C for details. References are as in the main text.

$^*$ Birth pulse is given by the function: $B(t) = \kappa \exp(-s \cos^2(\pi t - \phi))$, where we chose the parameter: $\kappa = \mu e^{s/2}/I_0(s/2)$, to ensure a seasonal periodicity in the population size of rodents. Here $I_0(\cdot)$ is the complete Bessel function of the first kind.
Fig S3: Population dynamics of Lassa fever in rodents. (a) Modelled dynamics of the rodent population (solid), including the annual change in the number of infected rodents (dashed). Grey shading indicates the annual intensity of the rainfall. (b) Relative change in the total population size of rodents (blue), including infected (solid red) and immune rodents (dashed red). (c, d) Determination of the best-fit value of the rate $\nu$ at which $S$ becomes $I$ once in contact with $I$ using $\chi^2$-statistics. The shaded area with dashed boundaries indicates possible values of the prevalence level observed throughout the season. Vertical dashed line indicates the best-fit value of $\nu$. 

\[ \chi^2 \]
Fig S4: Monthly rainfall pattern in 1901–2015 in Edo state, Nigeria. (a) Shows the average monthly rainfall as the mean and one standard error. The red and blue points denote the beginning and the end of the rainy season respectively using the means. The dashed line indicates the threshold level of 60 mm for the rainy season according to the Köppen classification. (b) Shows the distribution of the starting and ending weeks of the rainy seasons (light red and light blue, respectively) over the whole time period, 1901–2015. Vertical axis is the number of times when the beginning or the end of the rainy season fell on that particular week. Darker bars indicate the events for recent years 2000–2015.
Fig S5: Geographical distribution of Lassa fever cases reported across Nigeria and visually reconstructed from weekly NCDC reports. Colour coding consists of three categories: 1–4, 5–9, and 10+ cases, respectively, to colour gradation. Empty sites indicate that no cases were registered.
A. Analysis of a nosocomial outbreak in Jos, Nigeria, in 1970

Reconstructing the timeline of the outbreak

We used the available event time data of a nosocomial outbreak in Jos, Nigeria, in 1970, to estimate the incubation period and the time from illness onset to death. In total, the outbreak involved 23 cases excluding the index case. Each case record \( j = \{1 \ldots 23\} \) contained a lower \( t_j^{e-} \) and an upper time boundary \( t_j^{e+} \) of the probable date of exposure to Lassa fever (LF) virus as well as two other event times, i.e., the date of illness onset \( t_j^o \) and the date of death \( t_j^d \) (Appendix Fig 1).

Inference of model parameters using the Bayesian approach

In our Markov Chain Monte Carlo (MCMC) iterations, we employed a normal distribution for the prior distribution of each exposure time, \( t_j^e \):

\[
t_j^e \sim \text{dnorm}\left(\text{mean} = \frac{t_j^{e+} + t_j^{e-}}{2}, \text{sd} = \frac{t_j^{e+} - t_j^{e-}}{2/1.96}\right).
\]

The mean and standard deviation (sd) were used to ensure that the majority of values belonged to the interval between \( t_j^{e-} \) and \( t_j^{e+} \). The normal distribution was preferred to the uniform prior \( \text{dunif}(\text{lower} = t_j^{e-}, \text{upper} = t_j^{e+}) \) to avoid the edge effects from the boundaries [1], see also discussion in [2].

Then we defined two main distributions for the incubation period and the time from illness onset to death, and employed the gamma distribution for each:

\[
t_j \sim \text{dgamma}\left(\text{shape} = \frac{m_i}{sd_i}^2, \text{rate} = \frac{m_i}{(sd_i)^2}\right),
\]

\[
t_j^{o,d} \sim \text{dgamma}\left(\text{shape} = \frac{m_d}{sd_d}^2, \text{rate} = \frac{m_d}{(sd_d)^2}\right).
\]
where hyper-parameters, the means $m_i$ and $m_d$, and standard deviations $sd_i$ and $sd_d$, have non-informative positively defined priors given by a half flat (positive) distribution with arbitrarily small shape and scale parameters:

$$m_i, m_d, sd_i, sd_d \sim \text{dhalfflat}(\ )$$

The observed times of illness onset $t_j^o$ and death $t_j^d$ were subsequently inferred using sampling from normal distributions with a fixed standard deviation 0.5, which correlates with a half of the time scale in the timeline as a source of measurement error:

$$t_j^o \sim \text{dnorm}(\text{mean} = t_j^e + t_j^i, \text{sd} = 0.5),$$

$$t_j^d \sim \text{dnorm}(\text{mean} = t_j^e + t_j^i + t_j^{od}, \text{sd} = 0.5).$$

MCMC iterations

We performed MCMC iterations with 500,000 iterations, plus 20,000 iterations as a burn-in period. The thinning parameter of $m_i, m_d, sd_i, sd_d$ was equal to 50, and the exposure time $t_j^e$ was set at 10. The resulting trace plots for the mean and variance complemented with autocorrelation plots shown in Appendix Figures 2–5 demonstrate the sufficient convergence power of the iterative process.

B. Model-based inference using the human case (incidence) data

Maximum likelihood estimation

As stated in the main text, we used maximum likelihood estimation (MLE) to obtain point estimates of model parameters $\theta = \{a_{\pm}, q, l_{1,2}\}$. Specifically, we maximized the total (composite) likelihood of the form:

$$L_2(\theta \mid \{i_w + d_w, d_w\}) = L_i(\theta \mid \{i_w + d_w\}) L_d(\theta \mid \{i_w + d_w, d_w\}).$$
with respect to three varied parameters \(\{a_\pm, q\}\) for a fixed pair of indices \(l_{1,2}\). Then we performed a grid search over all possible values \(l_{1,2}\), spanning the range \(\{1,2,...,52\}\) to determine the global maximum of the likelihood \(L_2\). Each point estimate can be complimented with 95% confidence intervals, e.g., based on the likelihood profile.

Appendix Figure 6 shows the resulting fit using the MLE procedure. For comparison, the fit of the analogous model with a (single) constant exposure rate is shown in Appendix Figure 7. The single exposure rate model yielded a greater Akaike information criterion value (3679.2 in contrast to 2877.2 with two exposure model).

**Fitting procedure using the Bayesian approach**

We found the level of uncertainty in each model parameter estimate by conducting MCMC simulations for Bayesian inference. Here, we describe the sampling procedure for each estimate by inferring an underlying posterior distribution. Specifically, we were interested in estimating the posterior of the exposure rate \(a\) for each high/low-risk period, and the CFR \(q\).

The exposure rate \(a_w\) at week number \(w\) was sampled from the Gamma distribution:

\[
a_w(l_{1,2}) \sim \text{dgamma}\left(\text{shape} = \frac{m^a(w_c(w), l_{1,2})^2}{sd^a(w_c(w), l_{1,2})}, \text{scale} = \frac{m^a(w_c(w), l_{1,2})}{(sd^a(w_c(w), l_{1,2})^2)}\right),
\]

where \(w_c\) is a corresponding calendar week, whereas a pair of hyper-parameters \(m^a\) and \(sd^a\) are given by the following expressions:

\[
m^a(w_c, l_{1,2}) = \begin{cases} m_1^a, & \text{if } l_1 \leq w_c \leq l_2, \\ m_2^a, & \text{otherwise}, \end{cases}
\]

\[
sd^a(w_c, l_{1,2}) = \begin{cases} sd_1^a, & \text{if } l_1 \leq w_c \leq l_2, \\ sd_2^a, & \text{otherwise}. \end{cases}
\]

We supplemented each with the following non-informative prior distributions: \(m_1^a, sd_1^a \sim \text{dhalfflat}(\cdot)\), that are half flat (positive) distributions. The time boundaries \(l_{1,2}\) had
the uniform prior distributions: $l_{1,2} \sim \text{dunif}(\text{lower} = 0, \text{upper} = 53)$, which ensure: $0 < l_{1,2} < 53$. We also imposed constraints: $m^q_1 < m^q_2$ and $l_1 + 3 < l_2$.

The risk of death $q$ is sampled from a Gamma distribution to ensure the positive range of its values:

$$q \sim \text{dgamma} \left( \text{shape} = \left( \frac{m^q}{sd^q} \right)^2, \text{scale} = \frac{m^q}{(sd^q)^2} \right),$$

where two hyper-parameters have the non-informative prior distributions of the form: $m^q, sd^q \sim \text{dhalflat}(\cdot)$, that are analogous to the priors for the exposure hyper-parameters used above.

To infer parameters, we sampled the number of cases and the number of deaths from Poisson and Binomial distributions, respectively:

$$i_w + d_w \sim \text{dpois} \left( \text{rate} = \lambda^i_w + \lambda^d_w \right),$$

$$d_w \sim \text{dbinom} \left( \text{size} = i_w + d_w, \text{prob} = \frac{\lambda^d_w}{\lambda^i_w + \lambda^d_w} \right),$$

where the rates $\lambda^i_w$ and $\lambda^d_w$ resulted from convolution sums. The first being defined by the formula:

$$\lambda^i_w = \sum_{k<w} (1 - q) \cdot \frac{a_{w-k} \times f_k}{1 - r},$$

where $f_w$ denotes the incubation time distribution, see Appendix A. Whereas, the second rate was given by the following formula:

$$\lambda^d_w = \sum_{k<w} q \cdot \frac{a_{w-k} \times g_k}{1 - r},$$

where $g_w$ is the distribution of time periods from exposure to death. In turn, we convoluted the distribution $f_w$ with the distribution of time periods from illness onset to death $h_w$ obtained earlier in Appendix A, i.e.:
\[ g_w = \sum_{k<w} f_{w-k} \times h_k. \]

**MCMC iterations**

To obtain posterior distributions of model parameters \( m_{1,2}, sd_{1,2}, m^q, sd^q, l_{1,2} \), we compiled MCMC iterations with 18 chains, each characterized by 100,000 iterations and a burn-in period consisting of 20,000 iterations (Appendix Code Snippets). The thinning parameter for all parameters was set to 100 to avoid correlation effects in the chain. The resulting trace and density plots, as well as the autocorrelation plots for each pair of parameters, are shown in Appendix Figures 8–16. As shown, a sufficient level of convergence of the iterative algorithm was obtained.

**C. Dynamic modelling of LF transmission in rodent populations**

**Model formulation**

To model the transmission dynamics of LF in rodents, we adopted the modelling framework reported by Peel et al. [3]. We let \( N(t) \) be the population size of rodents at time \( t \) (\( 0 \leq t < 1 \) – where the season length is scaled to one). Then, the dynamics could be represented as follows:

\[
\frac{dN(t)}{dt} = (B_0(t) - \mu)N(t),
\]

\[
N(t) = N(t + 1), \quad 0 \leq t < 1,
\]

where \( \mu \) is the mortality rate per capita, and the growth rate per capita \( B_0(t) \) is given by a periodic Gaussian function: \( B_0(t) = \kappa \exp(-s \cos^2(\pi t - \phi)) \). The mortality rate for rodent species *Mastomys natalensis* is remarkably high: \( \mu = 7 \) year\(^{-1} \), whereas the two parameters \( s \) and \( \phi \) were previously identified from 20-year observations of rodent species in Tanzania: \( s = 2.7, \phi = 0.4 \), see [4] and Supplementary Figure 1 therein. To adjust the model for a situation in Nigeria, we considered a lower average population density of 60 rodents per Ha compared to 80 rodents per Ha as in [4]. Whereas the seasonality parameter \( s \) will be the subject of the model fit.
Due to apparent differences in climate between Tanzania and Nigeria, such as changes in the timing and length of the rainy season, a time shift was imposed to the function $B_0(t)$ to obtain the seasonal population dynamics in Nigeria. Specifically, we first identified the end of the rainy season in both countries based on historical averages of rainfall over the time period 1901–2015: $t_{TZ} = 0.35$ for part of the year for Tanzania, $t_{NG} = 0.83$ for part of the year for Nigeria (Appendix Figure 16). Then we calculated the difference between these two times as $\tau = t_{NG} - t_{TZ} = 0.48$ for part of the year, and accounted for the shift in the birth rate function of this value: $B(t) = B_0(t - \tau)$ (Fig S3b).

**Periodic conditions for the change in population size**

We chose a parameter $\kappa$ to satisfy the periodicity condition for $N(t)$. We required: $N(0) = N(1)$, which translates to:

$$\ln \frac{N(1)}{N(0)} = \int_0^1 (B(t) - \mu) \, dt = \int_0^1 \kappa e^{-s \cos^2(\pi t - \phi)} \, dt - \mu = 0.$$

This rewrites as follows:

$$\kappa \int_0^1 \exp \left(-s \frac{\cos((\pi t - \phi) + 1)}{2} \right) \, dt - \mu = \kappa e^{-\frac{s}{2\pi}} \int_0^1 \exp \left(-s \frac{\cos(2\pi t - 2\phi)}{2} \right) \, d(2\pi t - 2\phi) - \mu = \frac{\kappa e^{-\frac{s}{2\pi}}}{2\pi} \int_0^{2\pi} \exp \left(-s \frac{\cos \psi}{2} \right) \exp \left(-s \frac{\cos \phi}{2} \right) \psi \, d\psi - \mu = \frac{\kappa e^{-\frac{s}{2\pi}}}{\pi} I_0(s/2) - \mu = 0,$$

to obtain: $\kappa = \mu e^{s/2} / I_0(s/2)$: Here, $I_0(x) = \pi^{-1} \int_0^\pi \exp(x \cdot \cos(\phi)) \, d\phi$ denotes the Bessel function of the first kind. Thus, we did not need to assign an estimate for parameter $\kappa$.

We also note that the average population density can be defined as follows:

$$\bar{N} = N(0) \int_0^1 \exp \left( \int_0^t (B(\tau) - \mu) \, d\tau \right) \, dt.$$
Susceptible-infected-recovered (SIR) model

The flow diagram (Fig S3a) translates into the following dynamical equations:

\[
\frac{dS(t)}{dt} = B(t)[S(t) + (1 - \varepsilon)I(t) + (1 - \omega)R(t)] - \frac{\beta S(t)I(t)}{N(t)} + \lambda R(t) - \mu S(t),
\]

\[
\frac{dI(t)}{dt} = \varepsilon B(t)I(t) + \frac{\beta S(t)I(t)}{N(t)} - (\gamma + \mu)I(t),
\]

\[
\frac{dR(t)}{dt} = \omega B(t)R(t) + \gamma I(t) - (\lambda + \mu)R(t),
\]

for the three components: susceptible \( S(t) \), infected \( I(t) \), and recovered \( R(t) \), respectively. All rates are defined as per capita. \( \gamma \) and \( \lambda \) denote transition rates between compartments \( S \) and \( I \), and \( I \) and \( R \), respectively. The last two rates \( \varepsilon \) and \( \omega \) are the rates of vertical (parental) transmission of the LF virus and its antibodies, respectively. The transmission coefficient is set to a variable \( \beta = \nu \cdot c \), which consists of the rate \( \nu \) at which a susceptible rodent becomes infectious when in contact with an infected rodent, and contact-density function \( c = c(N) \). The latter has been previously found by fitting the observed experimental data of contact pattern in rodents to the sigmoidal function \([4,5]\):

\[
c(N) = \frac{2.13}{1 + \exp(-0.05(N - 101.2))}.
\]

We translated the absolute quantities of infected and recovered rodents to their densities with respect to the total size of the population. We also introduced two new variables: \( x(t) = I(t)/N(t) \), and \( z(t) = R(t)/N(t) \), and redefined the system of dynamic equations written above as follows:

\[
\frac{dN(t)}{dt} = (B(t) - \mu)N(t),
\]

\[
\frac{dx(t)}{dt} = \nu c(N(t))x(t)(1 - x(t) - z(t)) - ((1 - \varepsilon)B(t) + \gamma)x(t), \tag{S1}
\]

\[
\frac{dz(t)}{dt} = \gamma x(t) - ((1 - \omega)B(t) + \lambda)z(t),
\]

13
The number of susceptible rodents is given by: \( S(t) = N(t)(1 - x(t) - z(t)) \).

**Inference of model parameters**

To predict the transmission dynamics of LF infection among rodents, we use documented facts on the course of infection [6]: (i) LF infection is asymptomatic in rodents; (ii) it is also transient, with the virus being cleared from the blood after the infection; (iii) LF virus can be secreted in urine up to 103 days; (iv) there is a probability of horizontal and vertical transmission of LF virus and its antibodies. This results in the following choice of parameters, see Fig S3c. In particular, we define the characteristic recovery time from the infection as 90 days, and the rate of recovery is therefore set to: \( \frac{1}{90} \text{ days}^{-1} = 4.06 \text{ year}^{-1} \). The characteristic time at which immunity is lost is assigned as 120 days, i.e., the respective rate is set to: \( \frac{1}{120} \text{ days}^{-1} = 3.04 \text{ year}^{-1} \).

However, we still do not know the rate \( v \) between susceptible and infected rodents that characterizes the transmission dynamics in the SIR model, the seasonality parameter \( s \) and possibly unknown average population size \( \bar{N} \). Identification of this becomes a subject of the model fit.

To determine the values of model parameters, we required the mean prevalence levels of LF infection among rodents in the dry and rainy seasons, defined as \( \bar{x}_{\text{dry}} \) and \( \bar{x}_{\text{rny}} \), to be close to the observed levels in the field experiments. Fichet–Calvet and colleagues [6] reported that a one-time measurement in the dry season revealed the mean prevalence to be at 29%, while five independent measurements (one in the dry season and four in the rainy season) showed the mean level of prevalence to be 43%. Hence, we adjusted the mean prevalence level separately for both the dry and rainy seasons by using a simple algebraic rule: if \( x_{\text{rny}} \) is the mean prevalence level observed in the rainy season, then it should satisfy the following condition for five independent measurements: \( \frac{1}{5} \times 29\% + \frac{4}{5} \times x_{\text{rny}} = 43\% \). This yields the value: \( x_{\text{rny}} = 46.5\% \). Thus, we required the values \( \bar{x}_{\text{dry}} \) and \( \bar{x}_{\text{rny}} \) predicted by our model to be close to 29% and 46.5%, respectively.
A fitting procedure was carried out by using the $\chi^2$ statistic. Specifically, we tried to minimize the quantity:

$$
\chi^2 = \frac{(29\% - \bar{x}_{\text{dry}})^2}{x_{\text{dry}}} + \frac{(46.5\% - \bar{x}_{\text{rny}})^2}{x_{\text{rny}}},
$$

where $\bar{x}_{\text{dry}}$ and $\bar{x}_{\text{rny}}$ are values obtained from the dynamics (S1). Appendix Figure 18 shows that the minimum of $\chi^2$ is reached on some one-dimensional manifold in the space of two parameters $\bar{N}$ and $s$. The fitted dynamics and the evaluated values of $v$ and $s$ for fixed $\bar{N} = 60$ Ha$^{-1}$ are shown in Figure S3.

**D. Computer simulations and code sharing**

All calculations were made using free, open-source statistical and programming environments (R Version 3.5.1, Python Version 3.6.6, and Julia Version 1.0.1). The results were tested in multiple computational environments to ensure the validity of the calculations and to avoid processing errors. MCMC simulations were performed using the R package NIMBLE Version 0.6-12 [7], and comparative analysis of different packages can be found elsewhere [8]. To run the significance test of association between the observed LF incidence and climatological variables, we used the R package rEDM [9]. The code for all calculations and to reproduce all of the figures are accessible from the open-shared repository: [http://tiny.cc/Lassa18Scripts](http://tiny.cc/Lassa18Scripts). This information can be used freely for non-commercial purposes.
Appendix References

1. Carpenter B. 2017 Computational and statistical issues with uniform interval priors. Available at: andrewgelman.com/2017/11/28/computational-statistical-issues-uniform-interval-priors/. Accessed on August 30, 2018.

2. Park SW, Dushoff J, Earn DJD, Poinar H, Bolker BM. 2018 Human ectoparasite transmission of the plague during the Second Pandemic is only weakly supported by proposed mathematical models. *Proc. Natl. Acad. Sci.* **115**, E7892–E7893. (doi:10.1073/pnas.1809775115)

3. Peel AJ, Pulliam JRC, Luis AD, Plowright RK, O’Shea TJ, Hayman DTS, Wood JLN, Webb CT, Restif O. 2014 The effect of seasonal birth pulses on pathogen persistence in wild mammal populations. *Proc. R. Soc. B Biol. Sci.* **281**, 20132962–20132962. (doi:10.1098/rspb.2013.2962)

4. Borremans B, Reijniers J, Hens N, Leirs H. 2017 The shape of the contact–density function matters when modelling parasite transmission in fluctuating populations. *R. Soc. Open Sci.* **4**, 171308. (doi:10.1098/rsos.171308)

5. Borremans B, Reijniers J, Hughes NK, Godfrey SS, Gryseels S, Makundi RH, Leirs H. 2017 Nonlinear scaling of foraging contacts with rodent population density. *Oikos* **126**, 792–800. (doi:10.1111/oik.03623)

6. Fichet-Calvet E, Becker-Ziaja B, Koivogui L, Günther S. 2014 Lassa Serology in Natural Populations of Rodents and Horizontal Transmission. *Vector-Borne Zoonotic Dis.* **14**, 665–674. (doi:10.1089/vbz.2013.1484)

7. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. 2017 Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. *J. Comput. Graph. Stat.* **26**, 403–413. (doi:10.1080/10618600.2016.1172487)

8. Li M, Dushoff J, Bolker BM. 2018 Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches. *Stat. Methods Med. Res.* **27**, 1956–1967. (doi:10.1177/0962280217747054)

9. Ye H, Clark A, Deyle E, Munch S, et al. In press. *rEDM: Applications of Empirical Dynamic Modeling from Time Series. R* package version 0.7.1.

10. Carey DE, Kemp GE, White HA, Pinneo L, Addy RF, Fom ALMD, Stroh G, Casals J, Henderson BE. 1972 Lassa fever Epidemiological aspects of the 1970 epidemic, Jos, Nigeria. *Trans. R. Soc. Trop. Med. Hyg.* **66**, 402–408. (doi:10.1016/0035-9203(72)90271-4)

11. Sluydts V, Davis S, Mercelis S, Leirs H. 2009 Comparison of multimammate mouse (Mastomys natalensis) demography in monoculture and mosaic agricultural habitat: Implications for pest management. *Crop Prot.* **28**, 647–654. (doi:10.1016/j.cropro.2009.03.018)
12. University of East Anglia Climatic Research Unit, Jones PD, Harris IC. 2008 Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena. NCAS British Atmospheric Data Centre. Assessed August 2017.
Appendix Figures

| # | Month | Day | December | January | February |
|---|-------|-----|----------|---------|----------|
| 1 (Index) | 25 | 1 | 8 | 15 | 22 | 29 | 5 | 12 | 19 |
| 1 | | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | | | | | |
| 4 | | | | | | | | | |
| 5 | | | | | | | | | |
| 6 | | | | | | | | | |
| 7 | | | | | | | | | |
| 8 | | | | | | | | | |
| 9 | | | | | | | | | |
| 10 | | | | | | | | | |
| 11 | | | | | | | | | |
| 12 | | | | | | | | | |
| 13 | | | | | | | | | |
| 14 | | | | | | | | | |
| 15 | | | | | | | | | |
| 16 | | | | | | | | | |
| 17 | | | | | | | | | |
| 18 | | | | | | | | | |
| 19 | | | | | | | | | |
| 20 | | | | | | | | | |
| 21 | | | | | | | | | |
| 22 | | | | | | | | | |
| 23 | | | | | | | | | |
| 24 | | | | | | | | | |

**Appendix Fig 1: Timeline of a nosocomial outbreak in Jos, Nigeria, in 1970, adapted from [10].** Shaded grey areas indicate the exposure period, while shaded yellow areas span the time period from illness onset to death.
Appendix Figure 2: Trace and density plots for the mean $m_t$ ("incubation_mean") and standard deviation $sd_t$ ("incubation_sd") of the incubation period distribution.
Appendix Figure 3: Autocorrelation plots for pairs of mean $m_i$ ("incubation_mean") and standard deviation $sd_i$ ("incubation_sd") values for the incubation period distribution.
Appendix Figure 4: Trace and density plots for the mean $m_d$ (“death_mean”) and standard deviation $sd_d$ (“death_var”) of the distribution of the time period between illness onset and death.
Appendix Figure 5: Autocorrelation plots for pairs of mean $m_d$ ("death_mean") and standard deviation $sd_d$ ("death_sd") values for the distribution of the time period between illness onset and death.
Appendix Figure 6: Model fit to the observed data of new cases (a) and fatal cases (b) using **maximum likelihood estimation**. Solid black line indicates the obtained average, whereas the shaded area shows the 95% profile-based confidence intervals. The fitted exposure rate as a function of calendar week is shown in the inset of (a).
Appendix Figure 7: Model fit to the observed data of new cases (a) and fatal cases (b) using maximum likelihood estimation for the alternative model with (single) constant exposure rate. Solid black line indicates the obtained average, whereas the shaded area shows the 95% profile-based confidence intervals. The fitted exposure rate as a function of calendar week is shown in the inset of (a).
Appendix Figure 8: Trace and density plots for $m_1^a$ and $sd_1^a$. 
Appendix Figure 9: Autocorrelation plots for the pair $m_1^a$ and $sd_1^a$. 
Appendix Figure 10: Trace and density plots for $m_2^a$ and $sd_2^a$. 
Appendix Figure 11: Autocorrelation plots for the pair $m_2^a$ and $sd_2^a$. 
Appendix Figure 12: Trace and density plots for $m^q$ and $sd^q$. 
Appendix Figure 13: Autocorrelation plots for the pair $m^q$ and $sd^q$. 
Appendix Figure 14: Trace and density plots for $l_1$ and $l_2$. 
Appendix Figure 15: Autocorrelation plots for the pair $l_1$ and $l_2$. 
Appendix Figure 16: Inference of the model parameters using MCMC iterations.
Appendix Figure 17: Difference in rainfall patterns between Nigeria (black) and Tanzania (green). The location used for Tanzania was 6°51'S, 37°38'E, which was the data collection site in the original study [11]. The closeness was restrained by gridded data point distribution in the global precipitation dataset [12]. Blue points indicate the starting points for the dry seasons.
Appendix Figure 18: Numerical minimization of the $\chi^2$ value over two parameters $s$ and $\bar{N}$. 
Appendix Code Snippet 1: Main R script including the NIMBLE module

```r
## Preamble
args = commandArgs(trailingOnly=TRUE)
set.seed(as.numeric(args[1]))
libraries = c("dplyr","magrittr","tidyr","readxl","nimble")
for(x in libraries) { library(x,character.only=TRUE,warn.conflicts=FALSE) }

'%' = function(x,y)paste0(x,y)
rho = 0.19

## Data for to describe Jos outbreak
onsetTimes = c(17, 19, 21, 21, 22, 22, 23, 23, 23, 25, 26, 26, 26, 27, 31, 39, 41, 41, 42, 44)
deathTimes = c(33, 48, 31, 47, 30, 41, 37, 32, 31, 35, 54, 30, 33, 33, 38, 39, 51, 48, 55, 60, 62)

## Prior knowledge
exposureTimesLower = c(5, 9, 5, 9, 5, 5, 13, 11, 12, 5, 11, 11, 5, 5, 12, 25, 5, 25, 25)
exposureTimesUpper = c(15, 18, 15, 18, 15, 18, 15, 18, 14, 15, 18, 15, 15, 18, 18, 18, 31, 18, 31, 31)

## Dataset of human incidence stored in Nigeria_raw.xlsx
yearMin = 2016
data = read_excel("../../data/Nigeria_raw.xlsx", sheet = "Incidence") %>%
  select(-one_of("Timeseries","Imputation","File in the repo"),-contains("URL")) %>%
  filter(Year>=yearMin) %>%
  group_by(Year) %>%
  mutate(Incidence_Reported = if_else(Week==1,Reported,Reported-lead(Reported)),
         Incidence_Deaths = if_else(Week==1,Deaths,Deaths-lead(Deaths))
  ) %>%
  ungroup

data %>% select(Year,Week,starts_with("Incidence")) -> Df
data.frame(Year=yearMin-1,Week=1:52,Incidence_Reported=NA,Incidence_Deaths=NA) %>%
rbind(Df %>% arrange(Year,Week)) %>%
rowwise %>%
mutate(Incidence_Reported_NA=ifelse(is.na(Incidence_Reported),rpois(1,30),NA),
       Incidence_Deaths_NA=ifelse(is.na(Incidence_Deaths),rpois(1,1),NA)) %>%
ungroup -> Df
(K = nrow(Df))

# Convolutions = calculation of initial values for MCMC simulations
# that are used as initial values for the followed inference
incubation_shape = 8.038
incubation_rate = 1/0.2278
death_shape = 3.3012
death_rate = 1/0.5968
incubation_probability = pgamma(1:K,shape=incubation_shape,rate=incubation_rate)-pgamma(1:K-1,shape=incubation_shape,rate=incubation_rate)
```

36
timeFromOnsetToDeath_probability = pgamma(1:K,shape=death_shape,rate=death_rate)-pgamma(1:K-1,shape=death_shape,rate=death_rate)
# time from Exposure event to Death is the convolution of two latter probabilities
timeFromExposureToDeath_probability = c(0)
for (x in 2:K) {
  timeFromExposureToDeath_probability = c(timeFromExposureToDeath_probability,
    sum(incubation_probability[1:(x-1)]*timeFromOnsetToDeath_probability[(x-1):1]))
}

### Machinery for Nimble
# Convolution functions
nimConvolutionWithCFR = nimbleFunction(
  run = function(a = double(1), b = double(1), q = double(1)) {
    L <- dim(a)[1]
    res1 <- numeric(L)
    for(k in 1:L) {
      res1[k] <- q[k]*a[k]
    }
    ans <- inprod(res1[1:L],b)
    return(ans)
    returnType(double(0))
  }
)
nimConvolution = nimbleFunction(
  run = function(a = double(1), b = double(1)) {
    L <- dim(a)[1]
    ans <- inprod(a[1:L],b)
    return(ans)
    returnType(double(0))
  }
)

## Shift in l1 and l2 used to avoid edge effects
shift = 26

nCases = length(onsetTimes)

## Main script
nimData = list(# incubation and period to death
  onsetTime = onsetTimes,
  deathTime = deathTimes,
  # incidence
  infected = Df$Incidence_Reported,
  dead = Df$Incidence_Deaths,
  # for contraint
  one = 1)

# the following values are used for priors
incubation_mean_median = 13.16830
incubation_mean_lower = 10.99341
incubation_mean_upper = 15.52661
incubation_sd_median = 5.257898
incubation_sd_lower = 3.530271
incubation_sd_upper = 7.312072
death_mean_median = 14.21680
death_mean_lower = 10.83839
death_mean_upper = 17.70390
death_sd_median = 8.033101
death_sd_lower = 5.157413
death_sd_upper = 11.476860

nimConsts = list(nCases = length(onsetTimes),
                 incubation_uncertainty = 0.5,
                 death_uncertainty = 0.5,
                 exposure_mu = .5*(exposureTimesLower+exposureTimesUpper),
                 exposure_sd = (exposureTimesUpper-exposureTimesLower)/1.96/2,
                 r = rho,
                 K = K,
                 week = (Df$Week-shift-1)%%52+1,
                 incubation_mean_mean = (incubation_mean_upper+incubation_mean_lower)/2,
                 incubation_mean_sd = (incubation_mean_upper-incubation_mean_lower)/1.96/2,
                 incubation_sd_mean = (incubation_sd_upper+incubation_sd_lower)/2,
                 incubation_sd_sd = (incubation_sd_upper-incubation_sd_lower)/1.96/2,
                 death_mean_mean = (death_mean_upper+death_mean_lower)/2,
                 death_mean_sd = (death_mean_upper-death_mean_lower)/1.96/2,
                 death_sd_mean = (death_sd_upper+death_sd_lower)/2,
                 death_sd_sd = (death_sd_upper-death_sd_lower)/1.96/2)

nimInits = function()
  lambdaIncidence0 = rexp(K,1/3); CFR0 = runif(K,0.02,0.12);
  list(# incubation and period to death
       incubation_mean = runif(1,incubation_mean_lower,incubation_mean_upper),
       incubation_sd = runif(1,incubation_sd_lower,incubation_sd_upper),
       death_mean = runif(1,death_mean_lower,death_mean_upper),
       death_sd = runif(1,death_sd_lower,death_sd_upper),
       onsetExpectedTime = onsetTimes,
       deathExpectedTime = deathTimes,
       exposureTime = .5*(exposureTimesLower+exposureTimesUpper),
       timeToDeath = deathTimes-onsetTimes,
       # results
       incubationPeriod = incubation_probability,
       timeFromOnsetToDeath = timeFromOnsetToDeath_probability,
       timeFromExposureToDeath = timeFromExposureToDeath_probability,
       # incidence
       I1 = (runif(1,44,52)-shift-1)%%52+1,
       I2 = (runif(1,15,12)-shift-1)%%52+1,
       infected = Df$Incidence_Reported_NA,
       dead = Df$Incidence_Deaths_NA,
       lambdaIncidence = (1-CFR0)*lambdaIncidence0,
       lambdaDeath = CFR0*lambdaIncidence0,
       lambdaReported = lambdaIncidence0,
       pDeath = CFR0,
       exposure = lambdaIncidence0,
       CFR = CFR0,
       mean_a = c(runif(1,5,15),runif(1,5,3)),
sd_a = runif(2,0.1,1),
mean_q = runif(1,0.02,0.12),
sd_q = runif(1,0.1,1))}

nimCode = nimbleCode({
    ### incubation period and period from illness onset to death
    incubation_mean ~ dnorm(mean=incubation_mean_mean, sd=incubation_mean_sd)
    incubation_sd ~ dnorm(mean=incubation_sd_mean, sd=incubation_sd_sd)
    death_mean ~ dnorm(mean=death_mean_mean, sd=death_mean_sd)
    death_sd ~ dnorm(mean=death_sd_mean, sd=death_sd_sd)
    for (k in 1:nCases) {
        exposureTime[k] ~ dnorm(mean=exposure_mu[k],sd=exposure_sd[k])
        incubationTime[k] ~ dgamma(mean=incubation_mean, sd=incubation_sd)
        timeToDeath[k] ~ dgamma(mean=death_mean, sd=death_sd)
        onsetExpectedTime[k] <- exposureTime[k] + incubationTime[k]
        deathExpectedTime[k] <- onsetExpectedTime[k] + timeToDeath[k]
        onsetTime[k] ~ dnorm(onsetExpectedTime[k], sd=incubation_uncertainty)
        deathTime[k] ~ dnorm(deathExpectedTime[k], sd=death_uncertainty)
    }
    ### epidemiological model
    incubation_shape <- incubation_mean^2/incubation_sd^2
    incubation_rate <- incubation_mean/incubation_sd^2*7.0 #from days to weeks
    death_shape <- death_mean^2/death_sd^2
    death_rate <- death_mean/death_sd^2*7.0 #from days to weeks
    for (k in 1:K) {
        incubationPeriod[k] <- pgamma(k,shape=incubation_shape,rate=incubation_rate) -
                               pgamma(k-1,shape=incubation_shape,rate=incubation_rate)
        timeFromOnsetToDeath[k] <- pgamma(k,shape=death_shape,rate=death_rate) -
                                pgamma(k-1,shape=death_shape,rate=death_rate)
    }
    timeFromExposureToDeath[1] <- 0
    for (k in 1:(K-1)) {
        timeFromExposureToDeath[k+1] <- nimConvolution(incubationPeriod[1:k],timeFromOnsetToDeath[1:k])
    }
    ### exposure
    for (k in 1:K) {
        mean_a_realized[k] <- mean_a[1]+(mean_a[2]-mean_a[1])*equals(step(week[k]-l1),step(l2-week[k]))
        sd_a_realized[k] <- sd_a[1]+(sd_a[2]-sd_a[1])*equals(step(week[k]-l1),step(l2-week[k]))
        exposure[k] ~ dgamma(mean=mean_a_realized[k],sd=sd_a_realized[k])
        CFR[k] ~ dgamma(mean=mean_q,sd=sd_q)
    }
    for (k in 52:(K-1)) {
        lambdaIncidence[k+1] <- nimConvolutionWithCFR(exposure[1:k],incubationPeriod[1:k],1-CFR[1:k])/(1-r)
        lambdaDeath[k+1] <- nimConvolutionWithCFR(exposure[1:k],timeFromExposureToDeath[1:k],CFR[1:k])/(1-r)
        lambdaReported[k+1] <- lambdaDeath[k+1]+lambdaIncidence[k+1]
        pDeath[k+1] <- lambdaDeath[k+1]/lambdaReported[k+1]
    }
    for (k in 53:K) {
        infected[k] ~ dpois(lambdaReported[k])
        dead[k] ~ dbin(pDeath[k],infected[k])
    }
    ## Priors
    for(i in 1:2) {
        mean_a[i] ~ dhalfflat()}
sd_a[i] ~ dhalfflat()
}
mean_q ~ dhalfflat()
sd_q ~ dhalfflat()
11 ~ dunif(0,53)
12 ~ dunif(0,53)
one ~ dconstraint(12>11-3)
}

nimModel = nimbleModel(nimCode,
    constants = nimConsts,
    data = nimData,
    inits = nimInits())

## Checking that all variables are properly initialized
nimModel$initializeInfo()

nimConf = configureMCMC(nimModel, thin = 100, setSeed=TRUE)
nimConf$addMonitors(c("incubationTime", "timeToDeath", "lambdaIncidence", "lambdaDeath", "pDeath", "exposure", "CFR", "infected", "incubationPeriod", "timeFromOnsetToDeath", "timeFromExposureToDeath"))

## Model compilation
nimMCMC = buildMCMC()
compiledModel = compileNimble(nimModel, nimMCMC, resetFunctions = TRUE, showCompilerOutput = TRUE)

Niter = 1e5
Nburn = 2e4
compiledModel$nimMCMC$run(niter=Niter+Nburn, nburnin = Nburn)
compiledModel$nimMCMC$mvSamples %>% as.matrix %>% as.data.frame -> nimSamples

saveRDS(nimSamples, file = paste0("nimSamples-",args[1],".rds"))

compiledModel$nimMCMC$getTimes() %>% { sum(.)/60 }