Bridging the gap: Using reservoir ecology and human serosurveys to estimate Lassa virus incidence in West Africa

Andrew J. Basinski[1], Elisabeth Fichet-Calvet[2], Anna R. Sjodin[3], Tanner J. Varrelman[4], Christopher H. Remien[1], Nathan C. Layman[3], Brian H. Bird[5], David J. Wolking[5], Corina Monagin[5], Bruno M. Ghersi[6], Peter A. Barry[7], Michael A. Jarvis[8], Paul E. Gessler[9], Scott L. Nuismer[3]

1 Department of Mathematics, University of Idaho, Moscow, ID 83844, USA

2 Department of Virology, Bernhard-Nocht Institute of Tropical Medicine, Hamburg 20359, Germany

3 Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA

4 Bioinformatics and Computational Biology, University of Idaho, Moscow, ID 83844, USA

5 One Health Institute, School of Veterinary Medicine, University of California, David, CA 95616, USA

6 Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

7 Center for Comparative Medicine, California National Primate Research Center, Department of Pathology and Laboratory Medicine, University of California, Davis, CA 95616

8 School of Biomedical and Healthcare Sciences, University of Plymouth, Devon PL4 8AA, UK

9 College of Natural Resources, University of Idaho, Moscow, ID 83844, USA

* snuismer@uidaho.edu

Abstract

Forecasting how the risk of pathogen spillover changes over space is essential for the effective deployment of interventions such as human or wildlife vaccination. However, due to the sporadic nature of spillover events and limited availability of data, developing and validating robust predictions is challenging. Recent efforts to overcome this obstacle have capitalized on machine learning to predict spillover risk. Past approaches combine infection data from both humans and reservoir to train models that assess risk across broad geographical regions. In doing so, these models blend data sources that separately describe pathogen risk posed by the reservoir and the realized rate of spillover into the human population. We develop a novel approach that models as separate stages: 1) the contributions of spillover risk from the reservoir and pathogen distribution, and 2) the resulting incidence of pathogen in the human population. Our methodology allows for a rigorous assessment of whether forecasts of spillover risk can reliably predict the realized spillover rate into humans, as measured by seroprevalence. In addition to providing a rigorous cross-validation of risk predictions, this methodology could shed light on human habits that modulate or amplify the resultant spillover. We apply our method to Lassa virus, a zoonotic pathogen that poses a high threat of emergence in West Africa. The resulting framework is the first forecast to quantify the extent to which predictions of spillover risk from the reservoir explain regional variation in human seroprevalence. We use predictions generated by the model to revise existing estimates for the annual number of new human Lassa infections. Our model predicts that between 935,200 – 3,928,000 humans are infected by Lassa virus each year, an estimate that exceeds current conventional wisdom.

Author Summary

The 2019 emergence of SARS-2 coronavirus is a grim reminder of the threat animal-borne pathogens pose to human health. Even prior to SARS-2, the spillover of so-called zoonotic pathogens was a persistent problem, with pathogens such as Ebola and Lassa regularly but unpredictably causing outbreaks. Machine-learning models that can anticipate when and where animal-to-human virus transmission is most likely to

occur could help guide surveillance effort, as well as preemptive countermeasures to pandemics, like information campaigns or vaccination programs. We develop a novel machine learning framework that uses data-sets describing the distribution of a virus within its host and the range of its animal host, along with human immunity data, to infer rates of animal-to-human transmission across a focal region. By training the model on data from the animal host, our framework allows rigorous validation of spillover predictions on human data. We apply our framework to Lassa fever, a viral disease of West Africa that is spread to humans by rodents, and update estimates of symptomatic and asymptomatic Lassa virus infections in humans. Our results suggest that Nigeria is most at risk for the emergence of new strains of Lassa virus, and therefore should be prioritized for outbreak-surveillance.

Introduction

Emerging infectious diseases (EIDs) pose a deadly threat to public health. Approximately 40% of EIDs are caused by pathogens that circulate in a non-human wildlife reservoir (i.e., zoonotic pathogens) [1]. Prior to full scale emergence, interaction between humans and wildlife creates opportunities for the occasional transfer, or spillover, of the zoonotic pathogen into human populations [2]. These initial spillover cases, in turn, represent newly established pathogen populations in human hosts that are subject to evolutionary pressures and may potentially lead to increased transmission among humans [2, 3]. Consequently, a key step in preempting the threat of EIDs is careful monitoring of when and where spillover into the human population occurs. However, because the majority of EIDs from wildlife originate in low and middle income regions with limited health system infrastructure, accurately estimating the rate and geographical range of pathogen spillover, and therefore the risk of new EIDs, is a major challenge [1].

Machine learning techniques have shown promise at predicting the geographical range of spillover risk for several zoonotic diseases including Lassa fever [4–6], Ebola [7, 8], and Leishmaniases [9]. Generally, these models are trained to associate environmental features with the presence or absence of case reports in humans or the associated reservoir. Once inferred from the training process, the learned relationships

between disease presence and the environment can be extended across a region of interest. Using these techniques, previous studies of Lassa fever (LF) have derived risk maps that assess the likelihood of human LF cases being present in different regions of West Africa [4, 5]. Fitted risk maps are often assessed, in turn, by evaluating the ability of a model to discriminate between case data and background data that was left out of the training process [5, 7]. Though such models have demonstrated impressive discrimination abilities when evaluated by such binary classification ability, these forecasts are not explicitly vetted on their ability to predict the magnitude of pathogen spillover from the reservoir into humans. As a result, the extent to which predicted risk explains the realized variation in human exposure to the pathogen is unclear.

To address this need, we develop a multi-layer machine learning framework that accounts for the differences between how data involving a wildlife reservoir, and data from human serosurveys, can simultaneously inform spillover risk in people and rigorously assess whether predicted risk quantifies the rate of new cases in humans. Our approach uses machine learning algorithms that, when trained on data from the wildlife reservoir alone, estimate the likelihood that the reservoir and the zoonotic pathogen are present in an area. These predictions are then combined into a composite estimate of spillover risk to humans. Next, our framework uses estimates of human pathogen seroprevalence, as well as estimates of human population density, to translate the composite risk estimate into a prediction of the realized rate of zoonotic spillover into humans. Omitting human seroprevalence data from the training process of the risk-layer has several advantages. First, in the case of LF, due to modern transportation and the longevity of Lassa virus antibodies in humans, a general concern is that the reported location of human disease or Lassa virus antibody detection is not the site at which the infection occurred [10–12]. Training the risk layer on reservoir data alone helps avoid these biases. Secondly, in our framework, human seroprevalence estimates provide an ultimate test of the risk layer's ability to correlate with cumulative human exposure to the pathogen.

We apply our framework to Lassa virus (formally *Lassa mammarenavirus* [LASV]), a negative sense, bi-segmented, single-stranded ambisense RNA virus in the *Arenaviridae* family and the causative agent of LF in West Africa [11, 13]. Though LASV can transmit directly between humans and often does so in hospital settings [14],

rodent-to-human transmission is believed to account for the majority of new LASV infections [11, 15]. LASV spreads to humans from its primary reservoir, the multimammate rat *Mastomys natalensis*, through food contaminated with infected rodent feces and urine, as well as through the hunting of rodents for food consumption [16]. Because *M. natalensis* have limited dispersal relative to humans, direct LASV detection in the rodents is likely to indicate actual areas of spillover risk.

We evaluate each layer of our framework for its ability to predict different attributes of LASV spillover into humans. Our model demonstrates a clear ability to predict spillover risk as measured by the spatial distribution of the LASV pathogen and reservoir, and a more moderate correlation between the predicted risk and human seroprevalence. We also use our machine learning framework to estimate the total LASV spillover into humans. Data from longitudinal serosurveys has been used to estimate that between 100,000 and 300,000 LASV infections occur each year, and that between 74 – 94% of LASV infections result in sub-clinical febrile illness or are asymptomatic [17]. Though these estimates are often used to describe the magnitude of LASV spillover into humans [11, 18, 19], their generality is unclear because they are based on extrapolation from serosurveys conducted in the 1980's in Sierra Leone [17]. More recent estimates indicate that as many as 13 million LASV infections may occur each year [20].

Data

Our response data-set contains three types of data: 1) capture-locations of genetically confirmed *M. natalensis*, as well as occurrence locations of non-*M. natalensis* murids; 2) locations and outcomes of LASV surveys conducted in *M. natalensis*; and 3) locations and measured seroprevalence of human LASV serosurveys. The first two data-sets generate response variables for the model layers that predict LASV risk. The human seroprevalence data are used to evaluate the combined LASV risk layer for its ability to predict LASV spillover in humans and are also used to calibrate the stage of the model that predicts human LASV spillover. Our full data-set and the script files used to fit the models are available in a github repository [21].

Mastomys natalensis presence data and background

We collected data on documented presences of *M. natalensis* using the African Mammalia database [22], supplemented with additional presences found in the literature. Because *M. natalensis* is morphologically similar to other rodents in the area (e.g., *Mus baoulei*, *Mastomys erythroleucus*), we only include those presences that have been confirmed with gene sequencing methods. Each presence was verified with the help of a rodent expert (E.F.C).

Fitting the model requires supplementing the presence-only data with background points, also called pseudo-absences [23, 24]. Background points serve as an estimate of the distribution of sampling effort for the organism being modeled [25]. We used background points chosen from capture locations of members within the Muridae family (i.e., rodents) in West Africa from the Global Biodiversity Information Facility (GBIF) website [26]. We only included background points that: 1) document the location of a species other than *M. natalenis*, 2) fall outside of any pixel that contains a documented *M. natalensis* capture, and 3) are within the study region.

These locations were used to categorize a subset of the pixels in West Africa into two exclusive categories: those in which one or more *M. natalensis* has been captured (capture-positive), and those with only non-*M. natalensis* occurrences. In total, our data-set classified 184 unique pixels as capture-positive for *M. natalensis*, and 897 pixels as background.

Surveys of *Mastomys natalensis* for Lassa virus

We compiled a data-set from published studies that sampled *M. natalensis* rodents for indicators of LASV. The majority of the studies used were found using Table 2 of Fichet-Calvet and Rogers (2009). For each study, we found the sampling location for each tested rodent (either latitude/longitude or a village name for which coordinates could be obtained). In total, we compiled ten rodent studies [17, 27–34]. To this data-set, we added documented occurrences of LASV in *M. natalensis* from the GenBank website [35]. We searched for the keywords Lassa, Lassa fever, Lassa virus, Lassa fever virus, Lassa arenavirus, and Lassa mammarenavirus. From the results, we collated any sources that tested confirmed *M. natalensis* for LASV and contained

latitude/longitude coordinates.

With these data, each 0.05°x0.05° pixel (approximately 5 km by 5 km) in West Africa that contained a *M. natalensis* LASV survey was classified into the categories "Lassa positive" or "Lassa negative." Specifically, a pixel was defined as Lassa positive if, at some point, a *M. natalensis* rodent was captured within the pixel, and the rodent tested positive for LASV using a RT-PCR assay. Pixels were classified as Lassa negative if five or more *M. natalenis* rodents in total were tested for LASV infection by RT-PCR, or tested for any previous arenavirus exposure using a serological assay, and all rodents tested were negative. This procedure allowed us to classify 74 unique pixels in total: 36 were classified as Lassa negative, and 38 were classified as Lassa positive (Figure 1).

Human seroprevalence data

Lastly, we collected human arenavirus seroprevalence data. To ensure that the measured seroprevalence was representative of a larger village population, we required that the individuals tested for the study were chosen at random from a village. This criterion excluded nosocomial outbreaks, for example, as well as case-studies that detected arenavirus antibodies in individual missionaries. Each datum contains latitude and longitude of where the serosurvey took place, the number of individuals tested, and the number of individuals determined to have arenavirus antibodies. In total, we collected 101 serosurveys from nine studies (Fig 1) [17, 36–43]. These serosurveys were conducted between 1973 and 2019 and are located in six countries in West Africa.

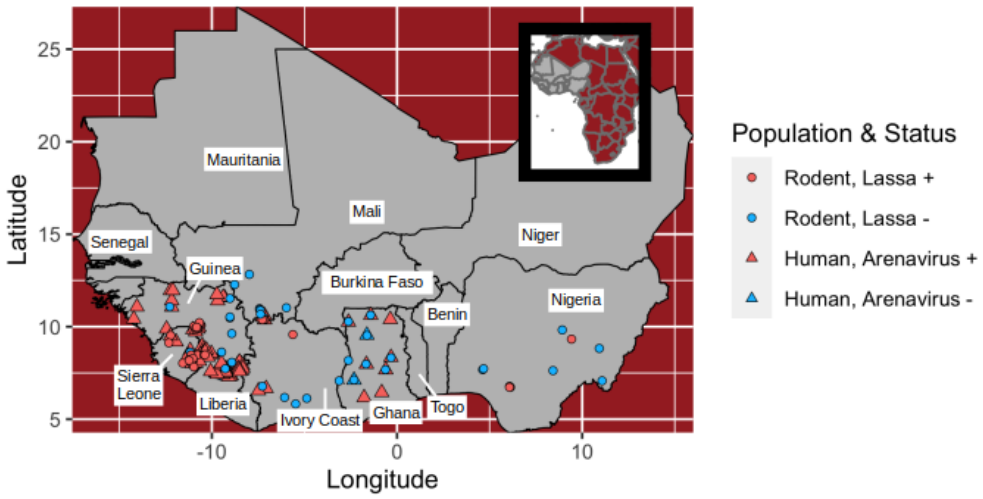

Fig 1. Locations at which Lassa virus or arenavirus antibodies have been sampled in rodents or humans. Each rodent point shows the outcome of a serological or PCR test. Each human population point shows the location of a village serosurvey.

Predictors

We include predictors that are broadly hypothesized to influence the distributions of *M. natalensis* and LASV. *M. natalensis* is widely distributed across sub-Saharan Africa in savanna and shrubland environments. Within such environments, *M. natalensis* is commonly associated with small villages and is considered a serious agricultural pest [44, 45]. To allow the model the possibility to learn these relationships, we include predictors that describe MODIS land cover features as predictors, and also include human population density within each pixel. We also include elevation in meters. Because climate seasonality and crop maturation affect the breeding season of *M. natalensis*, we include various measures of the seasonality of the vegetative index (NDVI), precipitation, and temperature [46]. See S1 Appendix for a complete list of environmental variables. LASV is often associated with *M. natalensis*, so we use the same set of predictors for the pathogen layer.

Methods

We developed a model that predicts the rate of LASV infection in humans within individual 0.05°x0.05° pixels across West Africa. This focal region is chosen as the intersection of West Africa and the International Union for Conservation of Nature (IUCN) range map for *Mastomys natalensis* [45]. Our *M. natalensis* capture data, as well as all of the LASV survey data, originate from within this region, thus providing a discrete bound on the area of West Africa in which the learned relationships of the model apply.

An overview of the model framework is depicted in Fig 2. Outputs from the model are generated in two stages. The first stage uses environmental features to estimate different layers of LASV spillover risk. The layers of risk, in turn, are described by: 1) D_M, a classification score indicating the likelihood that a pixel contains the primary rodent reservoir, *M. natalensis*, and 2) D_L, a score indicating the likelihood that LASV circulates within the *M. natalensis* population. Depending on the layer, the response variable for this stage is generated from documented occurrences of *M. natalensis* (D_M layer), or evidence of past LASV infection in *M. natalensis* (D_L layer). The second stage of our framework uses a generalized linear model to regress the estimates of human arenavirus seroprevalence onto a composite layer made from D_M and D_L. Lastly, we used a susceptible-infected-recovered-susceptible (i.e., SIRS) model to derive human incidence from the predictions of seroprevalence.

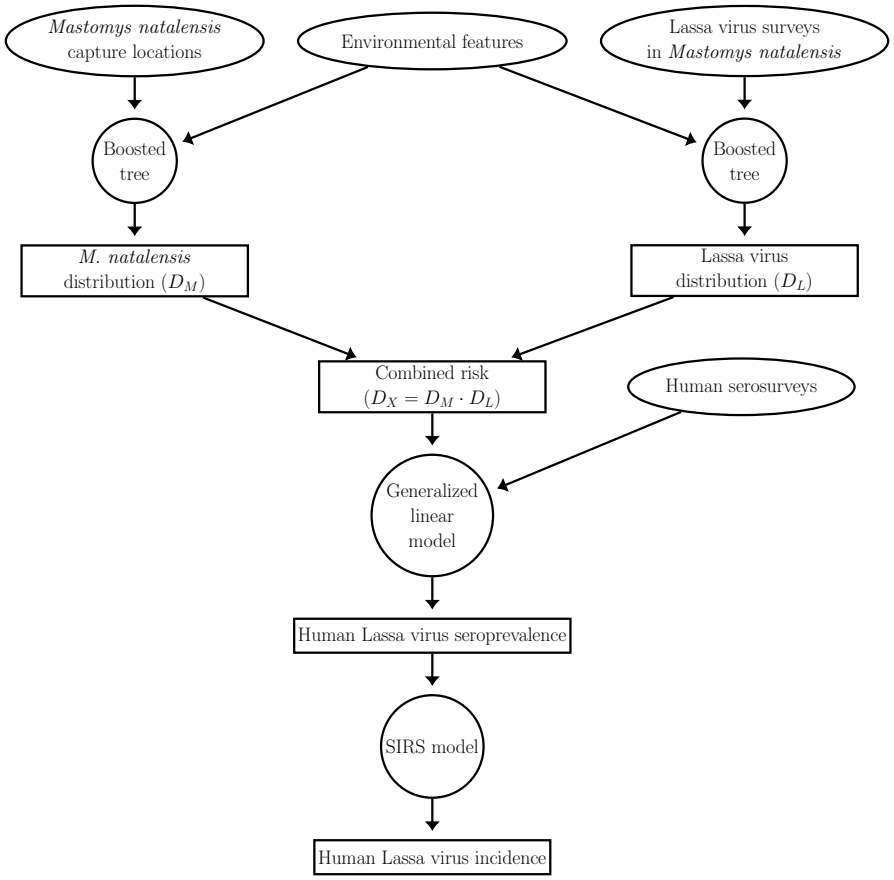

Fig 2. Overview of the model. Ellipses represent data-sets, circles represent models, and rectangles represent model predictions.

LASV risk layers

Each risk layer of the first stage is generated by a separate boosted classification tree (BCT). The BCT, in turn, uses environmental features within a pixel to infer a classification score, between zero and one, that indicates how likely it is that the pixel is positive for *M. natalensis* (D_M layer) or LASV in *M. natalensis* (D_L layer). BCTs use a stage-wise learning algorithm that, at each stage, trains a new tree model to the residuals of the current model iteration. Each newly fitted tree is added to the ensemble model, thereby reducing the residual deviance between the model predictions and a training set [47]. Boosted trees are commonly used in species and disease distribution models because they are simultaneously resistant to over-fitting in scenarios where many feature variables are implemented and are also able to model complex interactions among features [48].

Prior to inclusion in the model-fitting procedure, each feature variable was vetted for its ability to distinguish between presences and absences in each of the layers. Specifically, for each risk layer's binary response variable, we performed a Mann-Whitney U-test on each candidate feature. In doing so, we test the null hypothesis that the distribution of a feature is the same between pixels that are classified as a presence or (pseudo) absence. We only include predictors for which the null hypothesis is rejected at the $\alpha = 0.05$ level.

For a given training set, we fit the BCT model using the gbm.step function of the "dismo" package in the statistical language R [49]. This specific function uses 10-fold cross-validation to determine the number of successive trees that best model the relationship between response and features without over-fitting the data [49]. The learning rate parameter, which determines the weight given to each successive tree, was set to small values (D_M: 10^{-2}, D_L: 10^{-3}) that encourage a final model that is composed of many small incremental improvements. A smaller learning rate was used in the D_L layer because the corresponding data-set was smaller. The parameter that describes the maximum number of allowable trees was set to a large value (10^7) to ensure that the cross-validation fitting process was able to add trees until no further improvement occurred [47].

For the D_M layer, we trained 25 boosted classification trees to learn how

environmental predictors influence the suitability of a habitat for *M. natalensis*. Each model was fit by selecting 184 presence pixels and pairing these with 184 background pixels in which only non-*M. natalensis* murids were found. Both presences and background pixels were chosen with replacement. By choosing equal numbers of presences and background pixels for each training set, we encourage each model to learn patterns in features that allow presences to be discriminated from background pixels, rather than having the model learn the (likely biased) distribution of presences and background pixels that are available in the overall data-set [24].

For each model fit for the D_M layer, presence and pseudo-absence pixels that were not used to train the model (i.e., out-of-bag data) were used to test the model using the area-under-the-receiver-curve (AUC). The AUC measures a classifier's ability to assign a high classification score to presences, and a low score to background pixels. A score of one indicates a perfect classifier, and a score of 0.5 indicates a classifier that is no better than chance. Because some of the background pixels likely contain unreported *M. natalensis* (i.e., are false negatives), this is a conservative estimate of the model's performance. A pairwise-distance sampling scheme was used to pair an equal number of test-background pixels to the out-of-bag presences that together comprise the test set. Specifically, for each test presence point, the pairwise distance sampling method chooses a test background point so that the minimum distance between the training presences and test presence is similar to the minimum distance between the test background point and training presences [50]. Compared to random selection of test background points, pairwise distance sampling oftentimes results in a lower AUC score that more accurately measures the model's ability to generalize to new regions [50].

The D_L layer is generated by the averaged predictions of 25 boosted classification tree models, each of which is trained to discriminate between pixels that are Lassa positive or Lassa negative. The LASV rodent survey data-set contains 36 locations that were classified as Lassa negative, and 38 that were classified as Lassa positive. We trained each model on a data-set comprised of 36 absence locations and 36 presence locations, sampled from the full data-set with replacement. The estimation of error in the D_L layer is similar to that described in the D_M layer. Specifically, we calculate the AUC of the fitted model on an equal number of out-of-bag presences and absences.

Next, we combined the D_M and D_L layers into a composite feature, denoted by D_X,

that is indicative of whether a pixel simultaneously has environmental features that are suitable for *M. natalensis*, as well as LASV in *M. natalensis*. The combined feature is defined as $D_X = D_M \times D_L$ and summarizes the realized risk of LASV spillover to humans within the local environment.

Connection to human seroprevalence and incidence

To connect the new risk parameter D_X to human arenavirus seroprevalence, and to evaluate the ability of the D_X layer to explain historical LASV spillover in humans, we regressed seroprevalence from human arenavirus serosurveys on the D_X layer and an intercept. We used quasi-binomial regression to account for over-dispersion in seroprevalence measurements that could otherwise contaminate hypothesis tests on model coefficients [51]. In the regression, the log-odds of each seroprevalence estimate is weighted by the number of individuals tested in the serosurvey.

Next, we used a SIRS model that includes waning immunity to derive an equation that relates a given LASV spillover rate into humans and the resulting seroprevalence in a human population. Throughout, we assume that the seroprevalence measures that were obtained from historical serosurveys describe LASV infection at steady-state (i.e., are unchanging in time). This derivation, in turn, is used to translate the regression model's predictions of LASV seroprevalence into incidence (spillover cases per year) in humans. For the SIRS model, we employ several assumptions: 1) humans within each 0.05x0.05° pixel constitute a closed population with constant birth rate b and per-capita death rate d. Within each pixel, humans are compartmentalized into three non-overlapping classes: susceptible (S), infected with LASV (I), and recovered from LASV infection (R). The size of the human population is assumed to be large enough so that stochastic events (LASV extinction) do not occur. 2) All LASV infections in humans are caused by contact with infectious rodents. Though human-to-human transmission of LASV is common in nosocomial outbreaks, rodent-to-human transmission is believed to be the primary pathway by which the virus is spread outside of hospital environments [15]. 3) Susceptible humans become infected with LASV at a constant rate FS, where F denotes the rate of infectious contact between a human and infected *M. natalensis* (i.e., the force of infection). Any seasonal fluctuation in the

contact rate between humans and rodents, as well as fluctuation in the prevalence of LASV infection in rodents, is assumed to average out over the decades-long timescales we consider. 4) We assume that LASV infection in humans is non-fatal, so that infected humans recover at rate γ. All recovered individuals gain temporary LASV immunity through antibodies. Though LASV infection causes mortality in approximately 2% of non-nosocomial cases, this rate is small enough to be negligible for our predictions [11]. As described by McCormick et al. (1987), we assume that recovered individuals lose immunity to LASV infection and transition back into the susceptible class at a rate λ. For each pixel across West Africa, the equations that describe the number of humans in each of the classes are:

$$\begin{aligned} \frac{dS}{dt} &= b - dS - FS + \lambda R, \\ \frac{dI}{dt} &= FS - dI - \gamma I, \\ \frac{dR}{dt} &= \gamma I - dR - \lambda R. \end{aligned} \tag{1}$$

We find the steady-state values of S, I, and R by setting the left-hand-side of equations (1) to zero, and solving the resulting algebraic equations for each state variable. This yields the steady-state values

$$S^* = \frac{b}{d} \cdot \frac{(d+\gamma)(d+\lambda)}{(d+F)(d+\gamma)+\lambda(d+F+\gamma)}, \tag{2}$$

$$I^* = \frac{b}{d} \cdot \frac{F(d+\lambda)}{(d+F)(d+\gamma)+\lambda(d+F+\gamma)}, \tag{3}$$

$$R^* = \frac{b}{d} \cdot \frac{F\gamma}{(d+F)(d+\gamma)+\lambda(d+F+\gamma)}. \tag{4}$$

By dividing R^* by the total population size at steady-state, $\frac{b}{d}$, we find that the long-term seroprevalence, denoted Ω^*:

$$\Omega^* = \frac{F\gamma}{(d+F)(d+\gamma)+\lambda(d+F+\gamma)}. \tag{5}$$

Next, we use Eq (5) to estimate the LASV spillover rate FS^*, given that the

steady-state LASV seroprevalence is Ω^*. Solving Eq (5) for F in terms of Ω^* yields:

$$F = \frac{\Omega^*(d+\gamma)(d+\lambda)}{-\gamma + \Omega^*(d+\gamma+\lambda)}. \tag{6}$$

The rate of new cases is given by

$$\eta := FS^* = \frac{b}{d}\,\Omega^*\,(d+\gamma)\,(d+\lambda)\,\gamma^{-1}. \tag{7}$$

These analyses were derived using Mathematica. The notebook file is available in the github repository [21].

By substituting our prediction of human LASV seroprevalence for Ω^*, we can estimate the total human infection rate using Eq (7). Calculating these estimates requires values for b, d, γ and λ. We choose parameters that are broadly in line with the epidemiology of LASV, and the demography of humans in West Africa. We use the unprocessed WorldPop 2020 population data (see Data section) as an estimate of the steady-state population size, $\frac{b}{d}$, within each pixel of the original 0.0083° resolution. We choose d = 0.02 yr^{-1} to describe a mean human lifespan of 50 years. Studies indicate that the duration of LASV infection is typically about one month, so that $\gamma = 12$ yr^{-1} [11].

The rate of seroreversion is difficult to estimate empirically. McCormick et al. (1987) estimated that $\lambda = 0.064$ yr^{-1} using a longitudinal study of immune markers in individuals. However, it is unclear whether their results indicated true seroreversion, or whether the reduction of LASV immune markers below detectable levels made it appear as though seroreversion occurred. Furthermore, LASV antibodies have been shown to exist decades after infection in at least some individuals [10]. Because of the uncertainty in the rate of reversion, we report the number of new cases estimated with values in accordance with McCormick et al. (1987) ($\lambda = 0.064$), and also in the scenario where seroreversion never occurs ($\lambda = 0$).

The general effect of seroreversion can be understood by comparing estimates of new cases, η, with that obtained by the same equation with $\lambda = 0$, denoted η_0. We find that with our parameter values,

$$\frac{\eta}{\eta_0} = \frac{d+\lambda}{d} \approx 4.2. \tag{8}$$

In words, the estimated number of LASV spillover cases per year is 4.2 times greater when seroreversion is included in the model, relative to estimates obtained when seroreversion does not occur.

We also derive a null estimate of the yearly number of LASV spillover cases in humans from Eq (7). This estimate assumes that the incidence of LASV in humans is the same everywhere in the West African study region. Specifically, we calculate Ω^* as the mean seroprevalence across all serosurveys, weighted by the number of individuals tested ($\Omega^* \approx 0.184$). The population size b/d is set equal to the population of humans in the West African study region (b/d $\approx$ 374 million).

Results

LASV risk layers

The D_M layer is constructed by averaging the predictions of 25 boosted classification tree models. Across all 25 bootstrap fits, the average out-of-bag AUC was 0.63, with a standard deviation of 0.05. This AUC indicates that the model has a modest ability to correctly discriminate pixels in which *M. natalensis* has been captured from background pixels, and is similar to out-of-bag AUC scores obtained in another study with a similar assessment criterion [5]. The algorithm assigned especially great importance to maximum precipitation, precipitation contingency (i.e. regularity of precipitation), elevation, and the coefficient of variation of precipitation (S1 Appendix). Across 25 fitted models that made up the D_L layer, the average AUC was 0.83, with a standard deviation of 0.08. This indicates a model that is good at discriminating between Lassa presences and absences. The algorithm primarily used precipitation contingency to determine whether or not a pixel is suitable for endemic LASV in *M. natalensis*(S1 Appendix).

Figure 3 shows maps of each of the fitted risk layers (top row), as well as the combined layer of realized risk, D_X. As indicated by the IUCN range map for *M. natalensis* [45], all West Africa countries likely harbor this primary rodent reservoir of LASV (Fig 3a). Similar to other Lassa risk maps [4,5], our D_L layer predictions indicate that the risk of LASV in rodents is primarily concentrated in the eastern and

western extremes of West Africa (Fig 3b). The combined risk, shown in Fig 3c, indicates that environmental features suitable for rodent-to-human LASV transmission are primarily located in Sierra Leone, Guinea, and Nigeria.

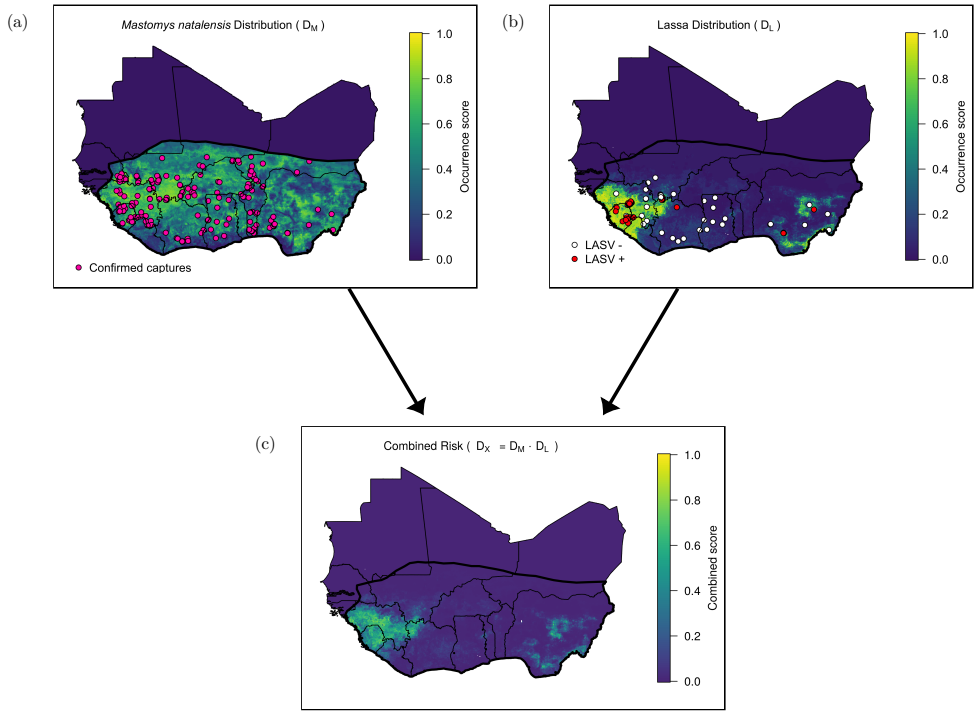

Fig 3. (a) Map shows the likelihood that each 0.05° pixel in West Africa contains the primary reservoir of Lassa virus, *M. natalensis*. Pink dots indicate locations of captures that were confirmed using molecular techniques and were used to train the model. Black line indicates the IUCN *M. natalensis* range map. (b) Predicted distribution of Lassa virus in *M. natalensis*. Dots indicate locations in which *M. natalensis* were surveyed for the virus. (c) Combined risk, defined as the product of the above two layers.

Connection to human seroprevalence and incidence

A quasi-binomial regression indicated a significant, positive association between the combined LASV risk predictor D_X, and the human arenavirus seroprevalence measured in serosurveys ($p = 0.0045$, Fig 4). The Pearson's correlation coefficient between the fitted model's predictions and actual human seroprevalence is 0.23 when all seroprevalence observations are equally weighted and 0.29 when weighted by the number of individuals tested in each survey. By applying the general linear model to the combined LASV risk layer, we extrapolate the human LASV seroprevalence across West Africa (Fig 5). Our results indicate that human LASV seroprevalence is greatest in the

eastern and western regions of West Africa, with especially high seroprevalence in central Guinea, Sierra Leone, and Nigeria.

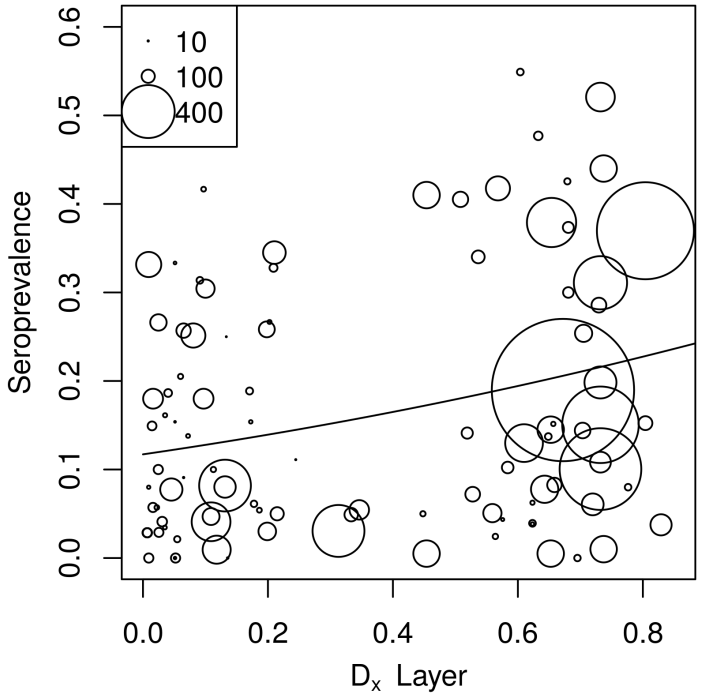

Fig 4. Human LASV seroprevalence (circles) and predictions of the quasi-binomial model (line) vs the D_x combined risk layer. Each dot represents a different serosurvey. The size of the dot indicates the number of humans that were tested.

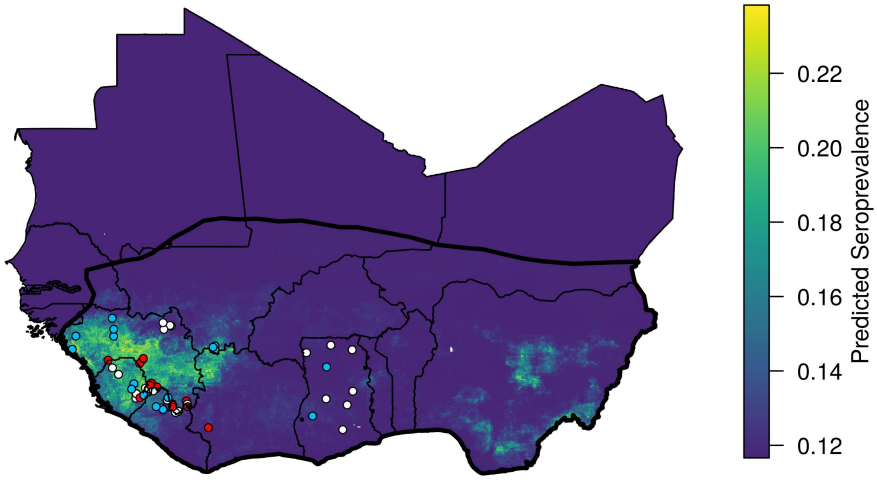

Fig 5. Predicted human seroprevalence of Lassa virus in West Africa. Dots show locations of human serosurveys that sampled at least 50 individuals, and dot color indicates the residual of the predicted seroprevalence. White dots indicate locations for which measured seroprevalence fell within 0.1 of the prediction. Measured seroprevalence at red dots was 0.1 or more greater than that predicted, and seroprevalence at blue dots was 0.1 or more below the prediction.

Furthermore, by assuming that our predictions are representative of LASV infection at steady-state, we can derive the number of LASV cases per year in humans. The simplest path to estimating human infections is to assume spatial homogeneity of LASV infection across West Africa. In this case we do not use the LASV risk layer D_X. Instead, we assume that human LASV seroprevalence is uniformly equal to the average seroprevalence across all available serosurveys (18.4%). This spatially uniform model implies 1,342,000 LASV infections occur in humans each year. When LASV reinfection (i.e., LASV infection following seroreversion) is included in the model, the estimate increases to 5,636,500 cases per year.

We can develop more spatially refined estimates using the spatially heterogeneous LASV risk that is predicted by the D_X layer. If LASV seroprevalence in humans is spatially heterogeneous, and spatial heterogeneity is described by the D_X layer, the model estimates that 935,200 – 3,928,000 new human infections occur each year. Table 1 shows the number of LASV infections per year by country, ordered by number of cases, when reinfection is assumed not to occur. Inclusion of reinfection does not change the ranking of countries. These spatially heterogeneous predictions indicate that more than half of new human LASV infections (513,200) in West Africa will occur in Nigeria (Fig 6). This distribution of LASV infection is largely due to the greater population size within Nigeria, as the per person incidence rates do not differ dramatically between countries (Table 1). After Nigeria, Ghana (72,500 cases per year) and the Ivory Coast (62,400 cases per year), respectively, are predicted to have the highest incidence of human LASV infections. Guinea and Sierra Leone are predicted to have the highest per-capita rates of LASV infection (Table 1), but because of their relatively small population sizes, these countries are predicted to have relatively few total cases.

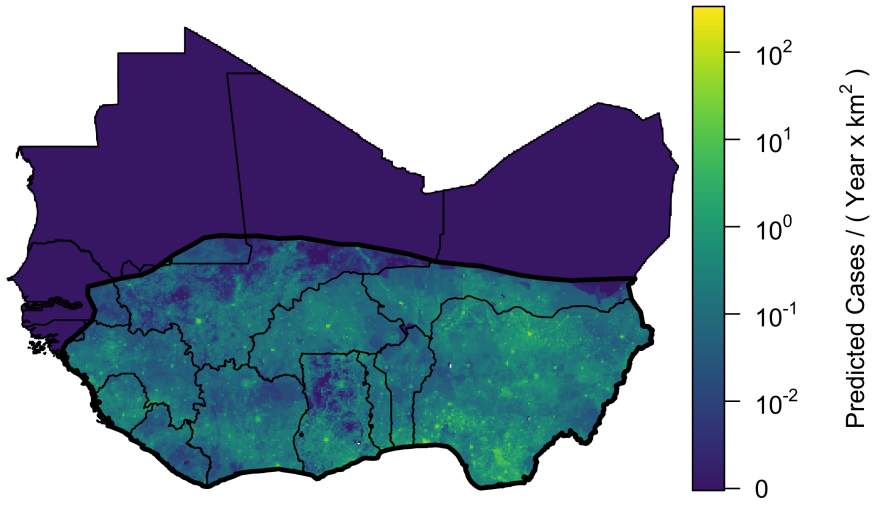

Fig 6. Predicted annual number of Lassa virus infections in humans, averaged over 25 bootstrap iterations. Yellow areas show regions with high population density that are also predicted to have high Lassa virus seroprevalence in humans.

Country	1000's of Cases	Rate
Nigeria	513.2	2.5
Ghana	72.5	2.4
Ivory Coast	62.4	2.4
Niger	54.4	2.4
Burkina Faso	49.9	2.4
Mali	47.8	2.4
Guinea	47.0	3.3
Benin	29.3	2.4
Sierra Leone	23.2	3.3
Togo	19.8	2.4
Liberia	13.2	2.6
Mauritania	1.3	2.4
Senegal	1.0	2.4

Table 1. Predicted annual number of Lassa virus cases in the study region, as well as infection rate (number of cases per year per 1000 people). Estimates in the table are derived assuming seroreversion and reinfection do not occur.

Discussion

Machine learning approaches that forecast the risk of emerging infectious diseases such as LF are often not assessed on their ability to predict proxies of pathogen spillover into human populations [5, 27]. Our forecasting framework advances these approaches by generating predictions of spillover risk based only on data from the primary rodent

reservoir of LASV, and directly assessing those predictions on data from human arenavirus serosurveys across West Africa. As indicated by a generalized linear regression, our reservoir-based model of spillover risk is able to explain a statistically significant amount of the variability in human seroprevalence. Furthermore, a generalized linear regression of human seroprevalence on the spillover risk yields predictions that are moderately correlated with the results of serosurveys (unweighted: 0.22; weighted: 0.29).

By distinguishing between the pathogen risk posed by a reservoir, and the realized seroprevalence in humans, our framework could allow for a more complete understanding of the factors that influence pathogen spillover into humans. In the case of LF, human factors such as the use of rodent-proof housing materials (concrete vs mud) and hunting habits can affect the extent to which LASV is able to transmit between rodents and humans [16, 52]. The residuals of seroprevalence predictions from our model could help guide understanding of where such human factors are mitigating or facilitating spillover into humans. If such human factors like housing type can be readily identified for regions within West Africa, they can be incorporated in the human stage of the model that connects spillover risk to human seroprevalence.

Because our framework traces the spillover risk into humans back to the spatial heterogeneity in Lassa risk and human density across West Africa, our approach allows us to predict which countries have the highest per-capita risk of LASV infection (e.g., Guinea, Sierra Leone) due to attributes of the reservoir and those that have the highest number of human cases because of their large human population size (e.g., Nigeria). Clarifying and distinguishing these two different types of risk helps to manage risk-reduction and behavior-change communication campaigns, countermeasures such as rodent population management or vaccination of rodent reservoir hosts, and travel advisories to high risk areas. In addition to intervention strategies such as vaccination or management of rodent populations, both of these areas of West Africa should be prioritized for surveillance of LASV emergence in rodents and at-risk human populations.

Using this framework, we are able to generate predictions of the number of new cases of LASV infection within different regions of West Africa. Our results indicate that Nigeria contributes the greatest number of new human cases each year, and that the

magnitude of new cases in Nigeria is driven primarily by its greater human population density, rather than an increased per-capita risk. If these predictions are correct, Nigeria is likely to represent the greatest risk of LASV emergence because the large number of annual spillover events allows for extensive sampling of viral strain diversity and repeated opportunities for viral adaptation to the human populations [53].

In addition to identifying the countries most at risk for viral emergence, our model framework provides updated estimates for the rate of LASV spillover across West Africa. Previous estimates of 100,000 – 300,000 cases per year were based on longitudinal studies from communities in Sierra Leone conducted in the 1980's [17]. Using seroprevalence data from studies across West Africa, our model predicts between 935,200 – 3,928,000 LASV infections in humans occur each year. Where the true value lies within this range depends on whether or not seroreversion and subsequent LASV reinfection are regular features of human LASV epidemiology, and reinforces the need to better understand the scope for LASV reinfection. It is important to realize that our predictions include both symptomatic and asymptomatic cases. Thus, because many human LASV infections result in mild flu-like symptoms or are asymptomatic, it is unsurprising that our predicted values exceed the reported number of confirmed LF cases in Nigeria [54, 55].

Several factors contribute to the discrepancy between previous estimates of LASV spillover, and our revised estimates. McCormick et al. (1987) used seroconversion data from a 15 month period to infer a rate of LASV infection across West Africa. However, the population of West Africa has increased by a factor of 2.4 since that time, making these estimates outdated [56]. Later estimates that were partially based on the same longitudinal serosurveys derived an upper bound of 13 million LASV infections, but only considered the number of cases in Nigeria, Guinea, and Sierra Leone [20]. Furthermore, these later estimates are derived from the maximum observed human LASV seroconversion rate in the Sierra Leone study, which likely does not apply across West Africa. In contrast, our estimates are based on human seroprevalence data that comes from six countries in West Africa and spans a 47 year time period. Because our data-set was obtained from a broader spatial and temporal range, our estimates are less likely to be biased by sporadic extremes in LASV spillover.

Our modeling framework has the benefit of being extendable, thereby giving

structured insight into how other attributes of the reservoir and pathogen translate into observed human seroprevalence. Future iterations of this framework could include the contributions of 1) more detailed life history of *M. natalensis*; 2) additional LASV animal reservoirs; and 3) genomic variability in LASV strains. For example, the first stage of these advanced models could include the temporal probability of a rodent being inside a domestic dwelling. The incidence of LF is generally believed to peak in the dry season, when *M. natalensis* migrate into domestic settings [44, 57]. Temporal fluctuations in population density, due to seasonal rainfall, would provide another important insight into the seasonal burden of human LF cases [11]. Understanding this ecological connection is important because distributing vaccines at seasonal population lows in wildlife demographic cycles can, in theory, substantially increase the probability of pathogen elimination [58, 59]. Incorporating these temporal layers will become more feasible as more time-series data on population density in the focal reservoir species become available.

Other potentially important risk layers that could be added are geographic distributions for other known reservoirs of LASV. Specifically, several species of rodents are known to be capable of harboring the virus [27]. Though *M. natalensis* is believed to be the primary reservoir that contributes to human infection, it is unknown whether this holds across all regions of West Africa. Understanding the relationship between the habitat suitability of different rodent reservoirs and human LF burden may also help determine whether *M. natalensis* is the host at which intervention strategies should always be directed. Finally, additional virus sequence data could be used to train a risk layer that forecasts the presence or absence of specific genomic variants that are more likely to cause either severe disease or more efficient human-to-human transmission cycles.

Although the methods we have used here make efficient use of available data, the accuracy of our risk forecasts remains difficult to rigorously evaluate due to the limited availability of current data from human populations across West Africa. The sparseness of modern human data arises for two reasons: 1) the lack of robust surveillance and testing across much of the region where LASV is endemic and 2) the absence of publicly available databases reporting human cases in those countries that do have sophisticated surveillance in place. Improving surveillance for LASV across West Africa and

developing publicly available resources for sharing the resulting data would allow more robust risk predictions to be developed and facilitate targeting effective risk reducing interventions. Despite these limitations of existing data, the structured machine-learning models we develop here provide insight into what aspects of environment, reservoir, and virus, contribute to spillover, and the potential risk of subsequent emergence into the human population. By understanding these connections, we can design and deploy more effective intervention and surveillance strategies that work in tandem to reduce disease burden and enhance global health security.

Supporting information captions

S1 Appendix. Details on the predictors used in the model and model fits.

Acknowledgements

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–993. doi:10.1038/nature06536.

2. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to zoonotic spillover. Nat Rev Microbiol. 2017;15(8):502—-510. doi:10.1038/nrmicro.2017.45.

3. Hughes JM, Wilson ME, Pike BL, Saylors KE, Fair JN, LeBreton M, et al. The origin and prevention of pandemics. Clin Infect Dis. 2010;50(12):1636–1640. doi:10.1086/652860.

4. Fichet-Calvet E, Rogers DJ. Risk maps of Lassa fever in West Africa. PLoS Negl Trop Dis. 2009;3(3):e388. doi:10.1371/journal.pntd.0000388.

5. Mylne AQ, Pigott DM, Longbottom J, Shearer F, Duda KA, Messina JP, et al. Mapping the zoonotic niche of Lassa fever in Africa. Trans R Soc Trop Med Hyg. 2015;109(8):483–492. doi:10.1093/trstmh/trv047.

6. Pigott DM, Deshpande A, Letourneau I, Morozoff C, Reiner Jr RC, Kraemer MU, et al. Local, national, and regional viral haemorrhagic fever pandemic potential in Africa: a multistage analysis. Lancet. 2017;390(10113):2662–2672. doi:10.1016/S0140-6736(17)32092-5.

7. Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, et al. Mapping the zoonotic niche of Ebola virus disease in Africa. Elife. 2014;3:e04395. doi:10.7554/eLife.04395.001.

8. Pigott DM, Millear AI, Earl L, Morozoff C, Han BA, Shearer FM, et al. Updates to the zoonotic niche map of Ebola virus disease in Africa. Elife. 2016;5:e16412. doi:10.7554/eLife.16412.

9. Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, et al. Global distribution maps of the leishmaniases. Elife. 2014;3:e02851. doi:10.7554/eLife.02851.001.

10. Bond N, Schieffelin JS, Moses LM, Bennett AJ, Bausch DG. A historical look at the first reported cases of Lassa fever: IgG antibodies 40 years after acute infection. Am J Trop Med Hyg. 2013;88(2):241–244. doi:10.4269/ajtmh.12-0466.

11. Gibb R, Moses LM, Redding DW, Jones KE. Understanding the cryptic nature of Lassa fever in West Africa. Pathog Glob Health. 2017;111(6):276–288. doi:10.1080/20477724.2017.1369643.

12. Peterson TA, Moses LM, Bausch DG. Mapping transmission risk of Lassa fever in West Africa: the importance of quality control, sampling bias, and error weighting. PLoS One. 2014;9(8):e100711. doi:10.1371/journal.pntd.0000388.

13. Maes P, Alkhovsky SV, Bào Y, Beer M, Birkhead M, Briese T, et al. Taxonomy of the family Arenaviridae and the order Bunyavirales: update 2018. Arch Virol. 2018;163(8):2295–2310. doi:10.1007/s00705-018-3843-5.

14. Fisher-Hoch S, Tomori O, Nasidi A, Perez-Oronoz G, Fakile Y, Hutwagner L, et al. Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. BMJ. 1995;311(7009):857–859. doi:10.1136/bmj.311.7009.857.

15. Iacono GL, Cunningham AA, Fichet-Calvet E, Garry RF, Grant DS, Khan SH, et al. Using modelling to disentangle the relative contributions of zoonotic and anthroponotic transmission: the case of Lassa fever. PLoS Negl Trop Dis. 2015;9(1):e3398. doi:10.1371/journal.pntd.0003398.

16. Ter Meulen J, Lukashevich I, Sidibe K, Inapogui A, Marx M, Dorlemann A, et al. Hunting of peridomestic rodents and consumption of their meat as possible risk factors for rodent-to-human transmission of Lassa virus in the Republic of Guinea. Am J Trop Med Hyg. 1996;55(6):661–666. doi:10.4269/ajtmh.1996.55.661.

17. McCormick JB, Webb PA, Krebs JW, Johnson KM, Smith ES. A prospective study of the epidemiology and ecology of Lassa fever. J Infect Dis. 1987;155(3):437–444. doi:10.1093/infdis/155.3.437.

18. Ogbu O, Ajuluchukwu E, Uneke C, et al. Lassa fever in West African sub-region: an overview. J Vector Borne Dis. 2007;44(1):1–11.

19. Brosh-Nissimov T. Lassa fever: another threat from West Africa. Disaster and Mil Med. 2016;2(1):8. doi:10.1186/s40696-016-0018-3.

20. Richmond JK, Baglole DJ. Lassa fever: epidemiology, clinical features, and social consequences. BMJ. 2003;327(7426):1271–1275. doi:10.1136/bmj.327.7426.1271.

21. Basinski AJ. Pathogen Spillover Forecast; 2020. Github repository `https://github.com/54481andrew/pathogen-spillover-forecast.git`.

22. Van de Perre F, Adriaensen F, Terryn L, Pauwels O, Leirs H, Gilissen E, et al.. African mammalia; 2019. `http://projects.biodiversity.be/africanmammalia`.

23. Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 2006;29(2):129–151. doi:10.1111/j.2006.0906-7590.04596.x.

24. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol. 2012;3(2):327–338. doi:10.1111/j.2041-210X.2011.00172.x.

25. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl. 2009;19(1):181–197. doi:10.1890/07-2153.1.

26. GBIF occurrence download; 2020-01-22. `https://doi.org/10.15468/dl.pqbckb`.

27. Fichet-Calvet E, Becker-Ziaja B, Koivogui L, Günther S. Lassa serology in natural populations of rodents and horizontal transmission. Vector Borne Zoonotic Dis. 2014;14(9):665–674. doi:10.1089/vbz.2013.1484.

28. Monath TP, Newhouse VF, Kemp GE, Setzer HW, Cacciapuoti A. Lassa virus isolation from *Mastomys natalensis* rodents during an epidemic in Sierra Leone. Science. 1974;185(4147):263–265. doi:10.1126/science.185.4147.263.

29. Olayemi A, Oyeyiola A, Obadare A, Igbokwe J, Adesina AS, Onwe F, et al. Widespread arenavirus occurrence and seroprevalence in small mammals, Nigeria. Parasites & vectors. 2018;11(1):416. doi:10.1186/s13071-018-2991-5.

30. Safronetz D, Lopez JE, Sogoba N, Traore SF, Raffel SJ, Fischer ER, et al. Detection of Lassa virus, Mali. Emerg Infect Dis. 2010;16(7):1123. doi:10.3201/eid1607.100146.

31. Safronetz D, Sogoba N, Lopez JE, Maiga O, Dahlstrom E, Zivcec M, et al. Geographic distribution and genetic characterization of Lassa virus in sub-Saharan Mali. PLoS Negl Trop Dis. 2013;7(12). doi:10.1371/journal.pntd.0002582.

32. Wulff H, Fabiyi A, Monath T. Recent isolations of Lassa virus from Nigerian rodents. Bull World Health Organ. 1975;52(4-6):609.

33. Coulibaly-N'Golo D, Allali B, Kouassi SK, Fichet-Calvet E, Becker-Ziaja B, Rieger T, et al. Novel arenavirus sequences in *Hylomyscus* sp. and *Mus (Nannomys) setulosus* from Côte d'Ivoire: implications for evolution of arenaviruses in Africa. PloS One. 2011;6(6):e20893. doi:10.1371/journal.pone.0020893.

34. Lecompte E, Fichet-Calvet E, Daffis S, Koulémou K, Sylla O, Kourouma F, et al. Mastomys natalensis and lassa fever, West Africa. Emerg Infect Dis. 2006;12(12):1971. doi:10.3201/eid1212.060812.

35. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2017;45(D1):D37–D42. doi:10.1093/nar/gkw1070.

36. Akoua-Koffi C, Ter JM, Legros D, Akran V, Aidara M, Nahounou N, et al. Detection of anti-Lassa antibodies in the western forest area of the Ivory Coast. Med Trop. 2006;66(5):465–468.

37. Kernéis S, Koivogui L, Magassouba N, Koulemou K, Lewis R, Aplogan A, et al. Prevalence and risk factors of Lassa seropositivity in inhabitants of the forest region of Guinea: a cross-sectional study. PLoS Negl Trop Dis. 2009;3(11). doi:10.1371/journal.pntd.0000548.

38. Lukashevich I, Clegg J, Sidibe K. Lassa virus activity in Guinea: Distribution of human antiviral antibody defined using enzyme-linked immunosorbent assay with recombinant antigen. J Med Virol. 1993;40(3):210–217. doi:10.1002/jmv.1890400308.

39. Nimo-Paintsil SC, Fichet-Calvet E, Borremans B, Letizia AG, Mohareb E, Bonney JH, et al. Rodent-borne infections in rural Ghanaian farming communities. PloS One. 2019;14(4). doi:10.1371/journal. pone.0215224.

40. Sogoba N, Rosenke K, Adjemian J, Diawara SI, Maiga O, Keita M, et al. Lassa virus seroprevalence in sibirilia commune, Bougouni District, Southern Mali. Emerg Infect Dis. 2016;22(4):657. doi:10.3201/eid2204.151814.

41. Yalley-Ogunro J, Frame J, Hanson A. Endemic Lassa fever in Liberia. VI. Village serological surveys for evidence of Lassa virus activity in Lofa County, Liberia. Trans R Soc Trop Med Hyg. 1984;78(6):764–770. doi:10.1016/0035-9203(84)90013-0.

42. Monath TP, Mertens P, Patton R, Moser C, Baum J, Pinneo L, et al. A hospital epidemic of Lassa fever in Zorzor, Liberia, March-April 1972. Am J Trop Med Hyg. 1973;22(6):773–779. doi:10.4269/ajtmh.1973.22.773.

43. Fraser DW, Campbell CC, Monath TP, Goff PA, Gregg MB. Lassa fever in the eastern province of Sierra Leone, 1970–1972. Am J Trop Med Hyg. 1974;23(6):1131–1139. doi:10.4269/ajtmh.1974.23.1131.

44. Fichet-Calvet E, Lecompte E, Koivogui L, Soropogui B, Doré A, Kourouma F, et al. Fluctuation of abundance and Lassa virus prevalence in *Mastomys natalensis* in Guinea, West Africa. Vector Borne Zoonotic Dis. 2007;7(2):119–128. doi:10.1089/vbz.2006.0520.

45. Granjon L. The IUCN red list of threatened species 2016: e.T12868A115107375. 2016;doi:10.2305/IUCN.UK.2016-3.RLTS.T12868A22425266.en.

46. Leirs H, Verhagen R, Verheyen W. Productivity of different generations in a population of *Mastomys natalensis* rats in Tanzania. Oikos. 1993; p. 53–60. doi:10.2307/3545308.

47. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media; 2009.

48. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77(4):802–813. doi:10.1111/j.1365-2656.2008.01390.x.

49. Hijmans RJ, Phillips S, Leathwick J, Elith J. dismo: Species Distribution Modeling; 2017. Available from: `https://CRAN.R-project.org/package=dismo`.

50. Hijmans RJ. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology. 2012;93(3):679–688. doi:10.1890/11-0826.1.

51. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. CRC Press; 1989.

52. Bonner PC, Schmidt WP, Belmain SR, Oshin B, Baglole D, Borchert M. Poor housing quality increases risk of rodent infestation and Lassa fever in refugee camps of Sierra Leone. Am J Trop Med Hyg. 2007;77(1):169–175.

53. Antia R, Regoes RR, Koella JC, Bergstrom CT. The role of evolution in the emergence of infectious diseases. Nature. 2003;426(6967):658–661. doi:10.1038/nature02104.

54. CDC. NCDC Lassa cases; 2020. `https://ncdc.gov.ng/data`.

55. Yun NE, Walker DH. Pathogenesis of Lassa fever. Viruses. 2012;4(10):2031–2048. doi:10.3390/v4102031.

56. Worldometers info;.

57. Bausch DG, Demby AH, Coulibaly M, Kanu J, Goba A, Bah A, et al. Lassa fever in Guinea: I. Epidemiology of human disease and clinical observations. Vector Borne Zoonotic Dis. 2001;1(4):269–281. doi:10.1089/15303660160025903.

58. Schreiner CL, Nuismer SL, Basinski AJ. When to vaccinate a fluctuating wildlife population: is timing everything? J Appl Ecol. 2020;doi:10.1111/1365-2664.13539.

59. Nuismer SL, Remien CH, Basinski AJ, Varrelman T, Layman N, Rosenke K, et al. Bayesian estimation of Lassa virus epidemiological parameters: implications for spillover prevention using wildlife vaccination. bioRxiv. 2019; p. 841403. doi:10.1101/841403.