# 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

# 1 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, developing robust predictions is challenging. Recent efforts to overcome this obstacle have capitalized on machine learning to predict spillover risk. A weakness of these approaches has been their reliance on human infection data, which is known to suffer from strongly biased reporting. We develop a novel approach that combines sub-models for reservoir species distribution, pathogen distribution, and transmission into the human population. We apply our method to Lassa virus, a zoonotic pathogen with a high threat of emergence in West Africa. The resulting model predicts the distribution of Lassa virus spillover risk and allows us to revise existing estimates for the annual number of new human infections. Our model predicts that between 961,300 – 4,037,400 humans are infected by Lassa virus each year, an estimate that exceeds current conventional wisdom. Our model also predicts that Nigeria accounts for more than half of all new Lassa cases in humans, making it a high-risk area for Lassa virus to become an emergent pathogen.

# 2 Keywords

Lassa, Machine learning, zoonotic pathogen, emerging infectious disease, spillover, risk map

# 3 Introduction

Emerging infectious diseases (EIDs) pose a deadly threat to mankind. 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, can give an animal-borne pathogen a foothold for genetic mutations that allow 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 is occurring. 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], and Leishmaniases [8]. 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 classify areas as high or low risk [4, 5]. Though useful, these forecasts do not explicitly quantify the spillover rate of a pathogen into humans, and, as a consequence, do not reveal the relative risk of inter-species pathogen-host transmission. Furthermore, 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 [9–11].

Herein, we develop a multi-layer machine learning framework that accounts for the differences between how data involving a wildlife reservoir, and data

from human cases, inform spillover risk in people. Our approach uses machine learning algorithms that, when trained on data from the wildlife reservoir alone, estimate the likelihood of the reservoir and the zoonotic pathogen being present in an area. These predictions are combined into a composite estimate of spillover risk to humans. Estimates of human Lassa virus seroprevalence, as well as estimates of human population density, are then used to translate the composite estimate into a realized rate of zoonotic spillover into humans. 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 [10, 12]. Though LASV can transmit directly between humans and often does so in hospital settings [13], rodent-to-human transmission is believed to account for the majority of new LASV infections [10, 14]. 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 [15]. 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 use our model to update estimates of the annual rate of LASV spillover into humans. Data from longitudinal serosurveys has been used to estimate that between 100000 and 300000 LASV infections occur each year, and that between 74 – 94% of LASV infections result in sub-clinical febrile illness or are asymptomatic [16]. Though these estimates are often used to describe the magnitude of LASV spillover into humans [10, 17, 18], their generality is unclear because they are based on extrapolation from serosurveys conducted in the 1980's in Sierra Leone [16]. More recent estimates indicate that as many as 13 million LASV infections may occur each year [19]. Using our machine learning

framework to account for data from both rodents and humans, we endeavored to refine these estimates of total LASV spillover into humans.

# 4 Methods

We developed a model that predicts the rate of LASV infection in humans within individual 0.05°x0.05° areas (i.e., pixels) of a gridded region of 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* [20]. 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.

Outputs from the model are generated in two stages. The first stage uses environmental features to estimate different layers of LASV spillover risk (see Appendix for a complete list of environmental variables). 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). More details on the data-sets can be found in the Appendix. The second stage of our framework uses a generalized linear model to regress the estimates of human LASV seroprevalence onto a composite layer made from $D_M$ and $D_L$. Lastly, we used a susceptible-infected-recovered model to derive human incidence from the predictions of LASV seroprevalence.

## 4.1 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 [21]. 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 of the features [22].

In the $D_M$ layer, fitting the BCT 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 [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. Because our data-set contains locations in which LASV was present in sampled *M. natalensis* populations, as well as locations in which LASV was not found, background points were not used in the $D_L$ layer. For both layers, each pixel in the data-set was assigned a value of zero (for background or survey absences) or one (presences). To help the models reliably discriminate between locations of presence and absence, each model was fit with an equal number of absences

and presences [24].

For a given training set, we fit the BCT model using the gbm.step function of the "dismo" package in the statistical language R [27]. 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 [27]. 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 [21].

Each fitted model assigns a score between zero and one that indicates whether a given set of environmental features describes a presence or absence. To ensure that the relationships the model finds are robust, we bootstrapped the fitting procedure 25 times, sampling a different set of training points each time.

## 4.2 Connection to human seroprevalence and incidence

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.

To connect the new risk parameter $D_X$ to human LASV seroprevalence, we assume that the seroprevalence measures that were obtained from historical serosurveys describe LASV infection at steady-state (i.e. are unchanging in time). We then regressed counts of seropositive humans on the $D_X$ layer, as

well as an intercept, and included an offset term that accounts for the number of individuals tested. We used negative binomial regression because preliminary analyses indicated significant over-dispersion of the residuals under Poisson regression [28].

Finally, a susceptible-infected-recovered (SIR) model was used to estimate the combined number of asymptomatic and symptomatic human LASV infections per year (i.e., incidence). This estimate was derived from the predicted human LASV seroprevalence described above (details in Appendix). For reference, we also calculated human LASV incidence from a simpler null model that assumes a spatially homogeneous distribution of seroprevalence. Though the epidemiological characteristics of LASV infection in humans are still being clarified, at least some longitudinal data indicates that loss of seropositivity and subsequent reinfection with LASV is possible [16]. In the SIR setting, this implies a nonzero rate of recovered individuals becoming susceptible. We used the SIR model to explore the implications of a range of possible seroreversion rates on our estimates of LASV incidence. Specifically, we compared the predicted incidence of LASV given no seroreversion, to the predicted incidence when seroreversion occurs at a rate $\lambda = 0.064$ per year, as estimated by McCormick, Webb, Krebs (1987).

# 5 Results

## 5.1 LASV risk layers

Figure 1a shows each of the fitted risk layers (top row), as well as the combined layer of realized risk, $D_X$. The risk layers were produced by averaging 25 bootstrapped predictions. As indicated by the IUCN range map for *M. natalensis* [20], all West Africa countries likely harbor this primary rodent reservoir

of LASV (Figure 1a). Figure 1b shows the predicted classification score for the occurrence of LASV in *M. natalensis*, averaged over 25 bootstrapped predictions. Similar to other Lassa risk maps [4, 5], our map indicates that the risk of LASV in rodents is primarily concentrated in the eastern and western extremes of West Africa. The combined risk, shown in Figure 1c, indicates that environmental features suitable for rodent-to-human LASV transmission are primarily located in Sierra Leone, Guinea, and Nigeria. Regions of central West Africa are also at moderate risk.

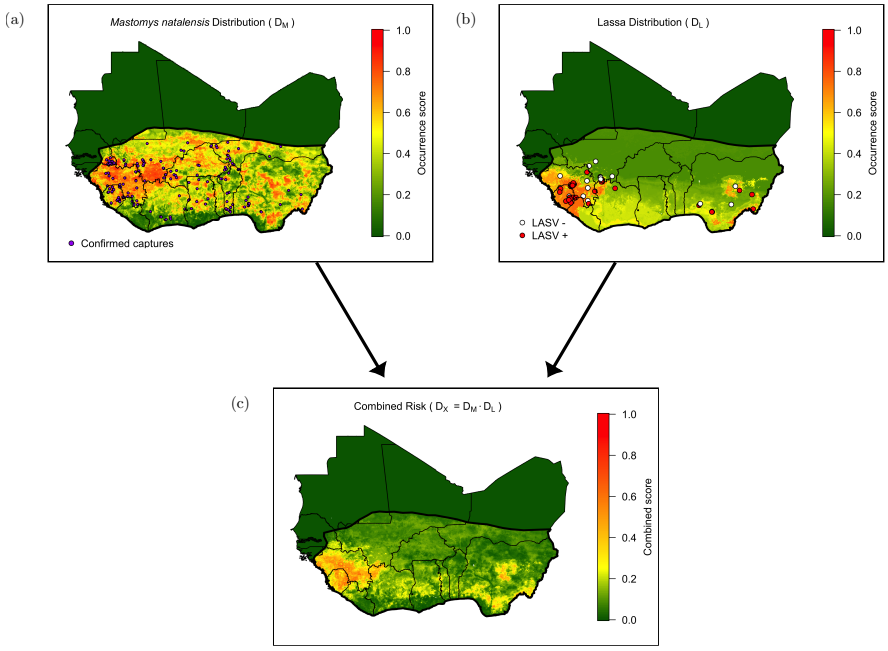


Figure 1: (a) map shows the likelihood that each 0.05° pixel in West Africa contains the primary reservoir of Lassa virus, *M. natalensis*. Purple dots indicate locations of captures that were confirmed using molecular techniques and were used to train the model. Black line indicates the IUCN *M. natalensis* 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.

## 5.2 Connection to human seroprevalence and incidence

We found a significant, positive association between the combined LASV risk predictor $D_X$, and the human LASV seroprevalence measured in serosurveys ($p = 0.0145$). The fitted model has a McFadden's pseudo r-squared value of 0.18, indicating a moderate ability to explain variation in human seroprevalence data. By applying the general linear model to the combined LASV risk layer, we extrapolate the human LASV seroprevalence across West Africa (Fig 2). 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.

Furthermore, by assuming that our predictions are representative of LASV infection at steady-state, we derived the number of LASV cases per year in humans (see Appendix for derivation). If the human LASV seroprevalence is assumed homogeneous in the study region, and equal to the average seroprevalence across all available serosurveys (18.4%), our model implies 1,380,400 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 5,797,600 cases per year. In contrast, if LASV seroprevalence in humans is spatially heterogeneous, and spatial heterogeneity is described by the LASV spillover risk layer $D_X$, our model estimates that 961,300 – 4,037,400 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. Our results indicate that more than half of new human LASV infections (531,500) in West Africa will occur in Nigeria (Fig 3). 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 (73,700 cases per year) and the Ivory Coast (64,400 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.

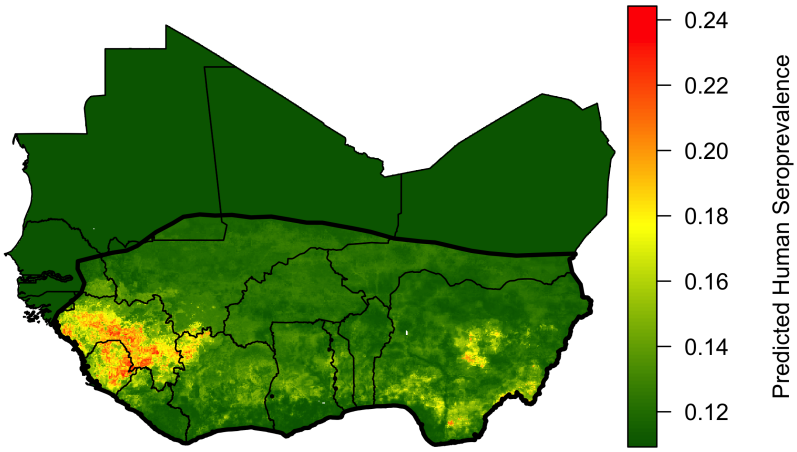


Figure 2: Predicted human seroprevalence of Lassa virus in West Africa, averaged over 25 bootstrap iterations.

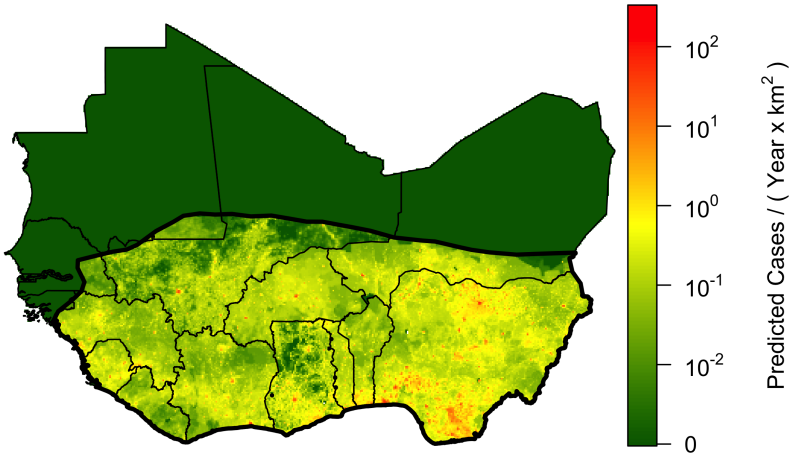


Figure 3: Predicted annual number of Lassa virus infections in humans, averaged over 25 bootstrap iterations. Red areas show regions with high population density that are also predicted to have high Lassa virus seroprevalence in humans.

| Country | 1000's of Cases | Rate |
|---|---|---|
| Nigeria | 531.5 | 2.6 |
| Ghana | 73.7 | 2.4 |
| Ivory Coast | 64.4 | 2.5 |
| Niger | 55.6 | 2.4 |
| Burkina Faso | 51.5 | 2.5 |
| Mali | 49.3 | 2.5 |
| Guinea | 46.4 | 3.3 |
| Benin | 30.4 | 2.5 |
| Sierra Leone | 22.8 | 3.2 |
| Togo | 20.5 | 2.5 |
| Liberia | 12.8 | 2.5 |
| Mauritania | 1.3 | 2.4 |
| Senegal | 1.1 | 2.5 |

Table 1: Predicted annual number of asymptomatic and symptomatic Lassa virus cases in the study region, as well as infection rate (No. cases per year per 1000 people). Estimates in the table are derived assuming seroreversion and reinfection do not occur.

# 6 Discussion

Machine learning approaches that forecast the risk of emerging infectious disease have shown promise for revealing geographical ranges of emerging pathogens. Our forecasting framework ties together data from different aspects of spillover risk posed by the primary rodent reservoir of LASV, to the seroprevalence measured in human serosurveys across West Africa. Using this approach, 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 [29].

In addition to identifying the countries most at risk for viral emergence, our model 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 [16]. Using seroprevalence data from studies across West Africa, our model predicts between 961,300 – 4,037,400 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 val-

ues exceed the reported number of confirmed LF cases in Nigeria [30, 31].

Several factors contribute to the discrepancy between previous estimates of LASV spillover, and our revised estimates. McCormick, Webb, and Krebs (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 [32]. 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 [19]. 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 35 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.

By integrating spatial heterogeneity in Lassa risk and human density across West Africa our model allows us to predict which countries have the highest per-capita risk of LASV infection (e.g., Guinea, Sierra Leone) 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 plan and 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.

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

structured insight into how other attributes of LASV risk 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 [33]. Temporal fluctuations in population density, due to seasonal rainfall, would provide another important insight into the seasonal burden of human LF cases [10]. 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 [34, 35]. Incorporating these temporal layers will become more feasible as more time-series data on population density in the focal reservoir species becomes 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 [36]. 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 reliable data from human populations across West Africa. The sparseness of 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.

# 7 Acknowledgements

# 8 Funding

Funding was provided by DARPA grant no. D18AC00028 and NIH grant no. R01GM122079.

# References

[1] Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L. & Daszak, P., 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–993. (doi:10.1038/nature06536).

[2] Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L. & Lloyd-Smith, J. O., 2017 Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502—-510. (doi:10.1038/nrmicro.2017.45).

[3] Hughes, J. M., Wilson, M. E., Pike, B. L., Saylors, K. E., Fair, J. N., LeBreton, M., Tamoufe, U., Djoko, C. F., Rimoin, A. W. & Wolfe, N. D., 2010 The origin and prevention of pandemics. *Clin. Infect. Dis.* **50**, 1636–1640. (doi:10.1086/652860).

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

[5] Mylne, A. Q., Pigott, D. M., Longbottom, J., Shearer, F., Duda, K. A., Messina, J. P., Weiss, D. J., Moyes, C. L., Golding, N. & Hay, S. I., 2015 Mapping the zoonotic niche of Lassa fever in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109**, 483–492. (doi:10.1093/trstmh/trv047).

[6] Pigott, D. M., Deshpande, A., Letourneau, I., Morozoff, C., Reiner Jr, R. C., Kraemer, M. U., Brent, S. E., Bogoch, I. I., Khan, K., Biehl, M. H. *et al.*, 2017 Local, national, and regional viral haemorrhagic fever pandemic potential in Africa: a multistage analysis. *Lancet* **390**, 2662–2672. (doi: 10.1016/S0140-6736(17)32092-5).

[7] Pigott, D. M., Golding, N., Mylne, A., Huang, Z., Henry, A. J., Weiss, D. J., Brady, O. J., Kraemer, M. U., Smith, D. L., Moyes, C. L. *et al.*, 2014 Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* **3**, e04395. (doi:10.7554/eLife.04395.001).

[8] Pigott, D. M., Bhatt, S., Golding, N., Duda, K. A., Battle, K. E., Brady, O. J., Messina, J. P., Balard, Y., Bastien, P., Pratlong, F. *et al.*, 2014 Global distribution maps of the leishmaniases. *Elife* **3**, e02851. (doi:10.7554/eLife.02851.001).

[9] Bond, N., Schieffelin, J. S., Moses, L. M., Bennett, A. J. & Bausch, D. G., 2013 A historical look at the first reported cases of Lassa fever: IgG antibodies 40 years after acute infection. *Am. J. Trop. Med. Hyg.* **88**, 241–244. (doi:10.4269/ajtmh.12-0466).

[10] Gibb, R., Moses, L. M., Redding, D. W. & Jones, K. E., 2017 Understanding the cryptic nature of Lassa fever in West Africa. *Pathog. Glob. Health.* **111**, 276–288. (doi:10.1080/20477724.2017.1369643).

[11] Peterson, T. A., Moses, L. M. & Bausch, D. G., 2014 Mapping transmission risk of Lassa fever in West Africa: the importance of quality control, sampling bias, and error weighting. *PLoS One* **9**, e100711. (doi: 10.1371/journal.pntd.0000388).

[12] Maes, P., Alkhovsky, S. V., Bào, Y., Beer, M., Birkhead, M., Briese, T., Buchmeier, M. J., Calisher, C. H., Charrel, R. N., Choi, I. R. *et al.*, 2018 Taxonomy of the family Arenaviridae and the order Bunyavirales: update 2018. *Arch. Virol.* **163**, 2295–2310. (doi:10.1007/s00705-018-3843-5).

[13] Fisher-Hoch, S., Tomori, O., Nasidi, A., Perez-Oronoz, G., Fakile, Y., Hutwagner, L. & McCormick, J., 1995 Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. *BMJ* **311**, 857–859. (doi:10.1136/bmj.311.7009.857).

[14] Iacono, G. L., Cunningham, A. A., Fichet-Calvet, E., Garry, R. F., Grant, D. S., Khan, S. H., Leach, M., Moses, L. M., Schieffelin, J. S., Shaffer, J. 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).

[15] Ter Meulen, J., Lukashevich, I., Sidibe, K., Inapogui, A., Marx, M., Dorlemann, A., Yansane, M., Koulemou, K., Chang-Claude, J. & Schmitz,

H., 1996 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.* **55**, 661–666. (doi: 10.4269/ajtmh.1996.55.661).

[16] McCormick, J. B., Webb, P. A., Krebs, J. W., Johnson, K. M. & Smith, E. S., 1987 A prospective study of the epidemiology and ecology of Lassa fever. *J. Infect. Dis.* **155**, 437–444. (doi:10.1093/infdis/155.3.437).

[17] Ogbu, O., Ajuluchukwu, E., Uneke, C. *et al.*, 2007 Lassa fever in West African sub-region: an overview. *J. Vector Borne Dis.* **44**, 1–11.

[18] Brosh-Nissimov, T., 2016 Lassa fever: another threat from West Africa. *Disaster and Mil. Med.* **2**, 8. (doi:10.1186/s40696-016-0018-3).

[19] Richmond, J. K. & Baglole, D. J., 2003 Lassa fever: epidemiology, clinical features, and social consequences. *BMJ* **327**, 1271–1275. (doi:10.1136/ bmj.327.7426.1271).

[20] Granjon, L., 2016 The IUCN red list of threatened species 2016: e.T12868A115107375 (doi:10.2305/IUCN.UK.2016-3.RLTS. T12868A22425266.en). Downloaded on 10 December 2019.

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

[22] Elith, J., Leathwick, J. R. & Hastie, T., 2008 A working guide to boosted regression trees. *J Anim. Ecol.* **77**, 802–813. (doi:10.1111/j.1365-2656.2008. 01390.x).

[23] Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A. *et al.*,

2006 Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151. (doi:10.1111/j.2006.0906-7590.04596.x).

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

[25] Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S., 2009 Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197. (doi:10.1890/07-2153.1).

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

[27] Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J., 2017 *dismo: Species Distribution Modeling*. R package version 1.1-4.

[28] McCullagh, P. & Nelder, J. A., 1989 *Generalized linear models*. CRC Press, 2 edition.

[29] Antia, R., Regoes, R. R., Koella, J. C. & Bergstrom, C. T., 2003 The role of evolution in the emergence of infectious diseases. *Nature* **426**, 658–661. (doi:10.1038/nature02104).

[30] CDC, 2020. NCDC Lassa cases. `https://ncdc.gov.ng/data`. Accessed: 2019-02-11.

[31] Yun, N. E. & Walker, D. H., 2012 Pathogenesis of Lassa fever. *Viruses* **4**, 2031–2048. (doi:10.3390/v4102031).

[32] Worldometers.info. Dover, Delaware, U.S.A.. Retrieved 03-03-20 from https://www.worldometers.info/world-population/western-africa-population/.

[33] Fichet-Calvet, E., Lecompte, E., Koivogui, L., Soropogui, B., Doré, A., Kourouma, F., Sylla, O., Daffis, S., Koulémou, K. & Meulen, J. T., 2007 Fluctuation of abundance and Lassa virus prevalence in *Mastomys natalensis* in Guinea, West Africa. *Vector Borne Zoonotic Dis.* **7**, 119–128. (doi:10.1089/vbz.2006.0520).

[34] Schreiner, C. L., Nuismer, S. L. & Basinski, A. J., 2020 When to vaccinate a fluctuating wildlife population: is timing everything? *J. Appl. Ecol.* (doi:10.1111/1365-2664.13539).

[35] Nuismer, S. L., Remien, C. H., Basinski, A. J., Varrelman, T., Layman, N., Rosenke, K., Bird, B., Jarvis, M., Barry, P. & Fichet-Calvet, E., 2019 Bayesian estimation of Lassa virus epidemiological parameters: implications for spillover prevention using wildlife vaccination. *bioRxiv* p. 841403. (doi:10.1101/841403).

[36] 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).