Mathematical modelling of the environmental and ecological drivers of zoonotic disease with an application to Lassa fever

Lauren Anne Attfield

Imperial College London

School of Public Health

Department of Infectious Disease Epidemiology

2022

This thesis is submitted for the degree of

Doctor of Philosophy (PhD)

Statement of Originality

I, Lauren Attfield, confirm that the work presented in this thesis is my own. My supervisors supported me in planning this work and provided valuable discussion and comments, and I performed all analysis. Any other work mentioned or used in this document is referenced and credited as appropriate.

Copyright Declaration

Abstract

Due to the essential role of zoonotic hosts, zoonotic spillover results from a complex system of environmentally-driven ecological and epidemiological processes. Despite being a global public health concern, zoonotic disease epidemiology is rarely viewed through the lens of disease ecology, meaning that the ecological factors driving zoonotic disease risk are typically not quantified. In this thesis I develop mathematical models to understand the zoonotic disease system from a process-based perspective informed by ecology, dependent on environmental variables, and tested using human health data. I focus these methods on a case study of Lassa fever which has historically been a neglected zoonosis but now may have improved opportunities for disease mitigation and surveillance. I present an overview of the topic in Chapter 1, outlining the challenges in zoonotic disease modelling and management. In Chapter 2, I find evidence of a severity bias in Lassa fever case data and estimate that infection incidence is likely on a much greater scale than previously thought. To elucidate environmental and ecological drivers of the Lassa virus system, in Chapter 3 I quantify the climatic dependence of reservoir host demographic processes. Along with strong seasonality, I estimate that year-on-year changes in precipitation can lead to substantial changes in the reservoir host population. In Chapter 4, I extend this population model to include pathogen transmission dynamics. Applying this model to states in Nigeria and linking reservoir host virus dynamics to observed human cases, I find that patterns of Lassa fever are significantly and positively correlated with predicted prevalence of infectious reservoir hosts. Finally, in Chapter 5 I summarise the findings and discuss future directions for the management and mitigation of zoonotic disease, concluding that ecological process-based modelling – facilitated by increased integration of knowledge, methods, and data – is essential for understanding zoonotic disease systems.

Acknowledgements

Thank you, Christl, for your keen eye on my methodology and for sharing your epidemiology expertise. Conversations with you inspired me to focus on what I was best able to bring to the table, and to be confident doing so. The Donnelly group at Imperial College and at the University of Oxford has been a source of positivity and experience which I am thankful to have been a part of. Thank you to other students in the Department of Infectious Disease Epidemiology for being friendly faces with a font of knowledge, and for being so welcoming.

Thank you, Kate, for sharing your ecology expertise and writing insights throughout this process. You encouraged me to submerge myself in the literature in order to become an interdisciplinary scientist and to continuously improve my writing style. Thank you to all in the Kate Jones lab at ZSL and UCL, including Dave Redding, Rory Gibb, Ella Browning, Lydia Franklinos, David Simons, Mikael Maes, and Roi Maor. I have greatly appreciated being able to share tips, solve problems, commiserate, and celebrate with you all. This thesis was shaped by valuable conversations and brainstorming sessions with Dave, Rory, and David – thank you for your openness, curiosity, and generosity with your time. The wider Centre for Biodiversity and Environment Research (UCL CBER) has been an indispensable support network. I am thankful to everyone at CBER for their kindness, passion, and positivity, not least Jess Williams, Adrienne Etard, and Ben Burgess.

My friends and family have been incredibly supportive while I pursued this research. You are too numerous and your contributions too diverse to list, but I'm grateful for your interest in my work and for listening to me when it's been challenging. To my parents I would like to say thank you for supporting my dreams and encouraging me to be a mathematician and a scientist. To Andy, thank you for everything.

Thank you to the current and historic directors of the Quantitative Methods in Ecology and Evolution CDT, and to my fantastic peers on this programme who have been inspiring and supportive. I am grateful for the opportunity to be part of an interdisciplinary programme and a cohort with such a diversity of research interests, and for the support that came from this.

The analysis I carried out in Chapter 2 was enhanced by Rory Gibb's relative abundance model for *Mastomys natalensis* in West Africa and David Simons' collated data on rodent trapping studies in West Africa; thank you for sharing your work with me.

I would like to thank Herwig Leirs and members of the Evolutionary Ecology group (University of Antwerp, Belgium) and Sokoine University of Agriculture (Morogoro, Tanzania) for providing data from their long-term *M. natalensis* trapping study and for valuable comments and discussion around Chapter 3 and early drafts of Chapter 4.

I am grateful to the Nigeria Centre for Disease Control for making their Lassa fever situation reports publicly available. It is my hope that open science will be an integral part of future zoonotic disease research.

Table of contents

List of figures

List of tables

Terminology

The word "zoonotic" was formed from the classical Greek **ζῷον** (zōon), meaning animal, and **νόσος** (nosos), meaning disease, in the late 20th century. "Zoonoses" or "zoonotic diseases" are defined by the World Health Organization (WHO) as "diseases and infections that are naturally transmitted from vertebrate animals to humans" (WHO 2020d). Vector-borne disease – disease caused by a pathogen which can be transmitted to humans via invertebrates or "vectors" – is sometimes considered a subset of zoonotic disease; depending on whether the vector is considered an independent host or simply a transmission mechanism. In this thesis, I focus on diseases for which the principal mode of transmission to humans is from a vertebrate host, with the pathogen sustained within a vertebrate reservoir community – meaning that invertebrates do not play a key role in the transmission among humans, hosts, reservoirs, or between these groups. For ease of reading, I will use "zoonotic disease", "zoonosis" and similar terms to refer to such diseases, and I acknowledge here this excludes some disease systems which could by some interpretations be categorised under zoonotic disease. Vector-borne disease modelling is associated with a different set of challenges and opportunities due to, for instance, the common role of human-to-human transmission via vectors in sustaining disease (WHO 2020a), therefore many of the methods developed and topics discussed in this thesis are not applicable to disease for which vectors are an important part of the disease system.

Definitions

Basic reproduction number (R_0): The expected number of secondary infections or infections arising from a single infected individual or disease case during their entire infectious period in a fully susceptible population. If $R_0 > 1$ then the pathogen has the potential for epidemic spread in the population.

Disease: A condition which negatively impacts an organism, manifested by symptoms and/or signs.

Emerging infectious disease: An infectious disease that has recently appeared within a (human) population or one whose infection or disease incidence or geographic range is rapidly increasing or threatens to increase in the near future (UNEP & ILRI 2020).

Endemic disease: The constant presence and/or usual prevalence of a disease in a population within a geographic area (UNEP & ILRI 2020).

Epidemic: The rapid spread of an infection or disease through a population. A large outbreak could be considered an epidemic.

Fatality ratio: The proportion of a given infected group for which the case/disease/infection is fatal.

> **Case fatality ratio:** The proportion of recorded disease cases which are fatal.
>
> **Disease fatality ratio:** The proportion of all symptomatic infections (disease) which are fatal.
>
> **Infection fatality ratio:** The proportion of all infections (both symptomatic and asymptomatic) which are fatal.

Host: An organism infected with a pathogen.

> **Dead-end host:** A host which cannot or does not transmit a given pathogen to any other organisms.
>
> **Intermediate host:** A host which provides a "bridge" between a reservoir and humans. The intermediate host can be infected by the reservoir and can transmit the pathogen to humans, even if the reservoir does not typically infect humans.
>
> **Principal host:** Here used to describe the host which is the main source of human infection.
>
> **Reservoir host:** An organism, population or community which serves as both a maintenance community/population and a host.
>
> **Zoonotic host:** A non-human vertebrate host of a zoonotic disease.

Incidence of disease: The number of instances of a given disease, across a given population, unit of space, or time. This may be different to the number of observed cases of a disease, which is influenced also by surveillance processes.

Incidence of spillover: The number of spillover events of a given pathogen, across a given population, unit of space, or time.

Infection: The state of being infected by a pathogen, which may or may not result in symptoms or clinical signs of disease.

Maintenance community/population: A community or population within which a pathogen is sustained, without which the pathogen would not be able to persist.

Outbreak: An increase in the incidence of an infection or disease in a given population or geographic area.

Pathogen: An organism which can cause disease. Here this definition includes micro-organisms (such as viruses, bacteria, and protozoa) as well as small animals commonly referred to as parasites.

> **Zoonotic pathogen:** A pathogen which can cause zoonotic disease in a target species or population.

Phenology: The periodic recurrence of natural phenomena due to climate and season.

Population model: A model describing changes in one or more populations.

> **Mathematical population model:** A mathematical formulation of a population model, often through differential equations. For example, the Lotka-Volterra equations (Hofbauer & Sigmund 1998) describe changes in interacting populations through coupled ordinary differential equations. The Fisher-KPP equation (Grindrod 1996) describes how a single population grows in space and time using a spatiotemporal partial differential equation.
>
> **Compartmental population model:** A mathematical model containing "compartments", with each compartment representing a sub-population, describing movement between these compartments. For example, the susceptible-infected-recovered epidemic (Kermack & McKendrick 1927) model is a compartmental population model with population moving between compartments depending on parameter values and the size of each compartment. All of the implicit individuals within a compartment are homogeneous and considered as one entity.

Demographic population model: A mathematical model in which changes in the population are dependent on demographic trait(s). For example, an age-structured matrix population model is an example of a demographic population model, with the demographic trait being age.

Matrix population model: A mathematical population model formed by describing the population as a vector, with each element of the vector representing a sub-population, and a transition matrix describing how the population vector changes. To obtain the population vector at the next time step, the current population vector is multiplied by the transition matrix. The way the population is allocated to elements depends on a structure; a stage-structured matrix population model assigns each element of the population vector to a different life stage that individuals can be in.

Reservoir: Here used synonymously with **Maintenance population/community**.

Seasonality: The way an observed phenomenon changes periodically and generally regularly over a year. For example, temperature seasonality is a measure of how temperature typically varies over the course of a year.

Severity bias: The phenomenon in which more severe cases of disease are more likely to be reported or recorded under imperfect surveillance.

Spillover: [See **Zoonotic spillover**]

Vector-borne disease: Human disease caused by a pathogen which is transmitted by invertebrates (vectors).

Vulnerability: An individual's susceptibility to infection, disease, or adverse health outcomes upon exposure to or infection with a pathogen.

Wild meat: Meat of a wild-caught vertebrate animal other than fish.

Zoonotic disease / Zoonosis: "Diseases and infections that are naturally transmitted from vertebrate animals to humans" (WHO 2020d). In this thesis, these terms refer to diseases for which the *principal mode of transmission* to humans is from a vertebrate zoonotic host, with the pathogen sustained within vertebrate reservoir(s).

Zoonotic hazard: Defined here as the abundance of infectious zoonotic hosts which may vary in space and time.

Zoonotic spillover: The event of transmission of a zoonotic pathogen from a zoonotic host to a human. This is sometimes defined as transmission of a zoonotic pathogen from any one species to another, but here will not be used in this way.

Acronyms

AIC: Akaike information criterion

AIDS: Acquired immunodeficiency syndrome

CEPI: Coalition for Epidemic Preparedness Innovations [cepi.net]

CFR: Case fatality ratio

CI: Confidence interval

CrI: Credible interval

CMR: Capture-mark-recapture

DALY: Disability-adjusted life-years

zDALY: Zoonotic DALY

DFID: Department for International Development, UK Government. This Department has now been replaced by Foreign, Commonwealth and Development Office (FCDO) [gov.uk/government/organisations/foreign-commonwealth-development-office].

DFR: Disease fatality ratio

EDCPT: European and Developing Countries Clinical Trials Partnership [edctp.org]

ELISA: Enzyme-linked immunosorbent assay

EU: European Union [european-union.europa.eu/index]

FAO: Food and Agriculture Organization of the United Nations [fao.org]

GLM: Generalised linear model

HIV: Human immunodeficiency virus

IAVI: International AIDS Vaccine Initiative [iavi.org]

ICONZ: Integrated Control of Neglected Zoonosis [ed.ac.uk/global-health/research/research-programmes/iconz]

IDSR: Integrated Disease Surveillance and Response (WHO Regional Office for Africa 2019)

ILRI: International Livestock Research Institute [ilri.org]

IPBES: Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services [ipbes.net]

IPM: Integral projection model

IQR: Interquartile range

MCMC: Markov chain Monte Carlo

MFR: Model fatality ratio (defined for the severity bias model in Chapter 2)

MNA: Minimum number alive

NCDC: Nigeria Centre for Disease Control [ncdc.gov.ng]

NPC: National Population Commission (Nigeria) [nationalpopulation.gov.ng]

ODE: Ordinary differential equation

PCR: Polymerase chain reaction

- **RT-PCR**: Reverse transcription PCR

PRNT: Plaque reduction neutralisation test

SDM: Species distribution model

SI (model): Susceptible-Infected

SIR (model): Susceptible-Infected-Recovered

TDR: Special Programme for Research and Training in Tropical Diseases [tdr.who.int]

UN: United Nations [un.org]

- **UNDP**: UN Development Programme [undp.org]
- **UNDP HDR**: UNDP Human Development Reports [hdr.undp.org]
- **UNEP**: UN Environment Programme [unep.org]

USAID: United States Agency for International Development [usaid.gov]

WHO: World Health Organization [who.int]

- **WHO-AFRO**: World Health Organization Regional Office for Africa [afro.who.int]

zDALY: [See under **DALY**]

Chapter 1: Introduction

Zoonotic disease – disease caused by transmission of a pathogen from a non-human vertebrate host to a human – presents a global research challenge due to the diversity and combination of factors which drive risk of zoonotic spillover and disease (Plowright *et al.* 2017). These drivers can relate to wildlife ecology and behaviour, immunology and pathogen evolution, epidemiology, and socioeconomics, resulting in a complex system of interacting factors which would traditionally be studied in distinct fields. As a result it is difficult but critical to understand zoonotic disease under this interdisciplinary paradigm (Cunningham *et al.* 2017). To understand a given zoonotic disease requires the consideration of pathways for transmission, potential for emergence, opportunities for mitigation, healthcare outcomes, and possible impacts of environmental change, making every zoonotic disease system unique. Meanwhile, globally, zoonoses cause substantial disease burden (Grace *et al.* 2012; Welburn *et al.* 2015) and danger of epidemic emergence (Jones *et al.* 2008; Morse *et al.* 2012), highlighting the importance and value of furthering the study of these complex disease systems. In the following sections I will outline the distinct challenges of managing zoonotic disease, identify research opportunities, and introduce Lassa fever as a valuable case study.

1.1 Local and global implications of zoonotic disease

At an individual level, the implications of contracting a zoonotic disease can be the same as for any other disease; the individual suffers from acute or chronic symptoms or clinical signs which alter quality of life (morbidity) and can lead to death (mortality), and the individual may require medical treatment, may be advised to isolate from others, and may have to modify their lifestyle in response to a new health condition. The impact of a disease on morbidity and mortality is known as disease burden, which is commonly measured through a metric called "disability-adjusted life years" (DALYs) (Mathers *et al.* 2013), the sum of years lived with a disability due to the disease and the years of life lost due to disease-induced mortality. While it is challenging to assess the burden of zoonotic diseases since they are often underreported – especially in low-income countries (Schelling *et al.* 2007) – it has been estimated that zoonoses, vector-borne diseases, and previously zoonotic diseases which are now maintained by human-to-human transmission account for 10% of the total burden of all diseases and health conditions in low-income countries, and only 0.02% in high-income countries (Grace *et al.* 2012). While in high-income countries endemic diseases originating in

animals are typically of relatively low concern to individual health, in low-income countries they can contribute substantially to local morbidity and mortality. Since individual endemic zoonoses do not present an immediate threat to global health and their disease burden is typically underestimated, endemic zoonoses have generally been overlooked for funding (Maudlin *et al.* 2009; Halliday *et al.* 2015).

Endemic zoonoses can also create a financial disease burden not limited to direct impacts on human health, for instance due to the economic losses caused by zoonotic disease in livestock which can impact both the livelihoods of agricultural workers and food security (WHO *et al.* 2011). An adjusted disease burden indicator for zoonotic disease, zDALY, has been proposed, made up of the standard DALY approach plus an animal loss equivalent, which is the number of years it would take for the average individual to earn the equivalent monetary value as the livestock loss (Torgerson *et al.* 2018). Torgerson *et al.* found that calculated animal loss equivalents far outweigh DALYs for cystic echinococcosis, a zoonotic disease caused by infection with the parasite *Echinococcosis granulosus*, in several countries. Further research on the societal consequences of animal diseases may become available following the implementation of the Global Burden of Animal Diseases programme (Rushton *et al.* 2018). It is notable that the intervention costs associated with reducing zoonotic disease incidence can also be high. For example, farmers' willingness to pay for livestock vaccination can be low even when a vaccine is considered to provide a net economic benefit (Jemberu *et al.* 2020). While mass livestock vaccination is often considered a cost-effective strategy, the costs associated with such a programme can be too high to be met solely by one sector (Roth *et al.* 2003). Interventions against zoonotic disease can also lead to societal costs in indirect ways. Although pressure is placed on governments to ban hunting of wildlife for meat and markets where wild meat is sold to reduce risk of zoonotic spillover and biodiversity loss (Pooley *et al.* 2015), this industry can provide income which may alleviate poverty (Kümpel *et al.* 2010) and a protein source which can improve food security (Friant *et al.* 2020). In summary, the local implications of endemic zoonotic disease extend beyond morbidity and mortality to financial and social costs, with large-scale interventions incurring additional costs which could outweigh the societal benefit of disease alleviation.

Risks to global health from zoonotic diseases do not only lie in endemic disease burden. Every spillover event is an opportunity for onward transmission between humans, which in rare cases leads to onward transmission of a pathogen with epidemic potential (Karesh *et al.*

2012). A zoonotic disease with epidemic potential, which becomes sustainable by human-to-human transmission due to a reproduction number greater than one, would become an emerging infectious disease if not contained. Emergence of a zoonosis in the human population is a highly stochastic event resulting from an already complex socioecological zoonotic transmission pathway, with proposed risk factors including population biology and range of the zoonotic host (Woolhouse 2002), pathogen type (Taylor *et al.* 2001; Woolhouse & Gowtage-Sequeria 2005), host plasticity (Kreuder Johnson *et al.* 2015), and pathogen evolutionary potential (Bonneaud & Longdon 2020). It is evident that zoonotic diseases present an immense risk to global health, beyond that of the local community; for example, the H1N1 influenza A virus emerged, probably from swine or birds, leading to the so-called "Spanish" flu pandemic in 1918 resulting in at least 20 million and perhaps as many as 100 million deaths (Johnson & Mueller 2002). Human immunodeficiency viruses (HIV), first recognised in 1981, emerged from wild primates (Sharp & Hahn 2011) and are the cause of the ongoing HIV/AIDS pandemic, with an estimated 37.7 million people globally living with HIV in 2020 (WHO 2021b). The coronaviruses SARS-CoV-1, MERS-CoV, and SARS-CoV-2 all emerged in the 21st century (da Costa *et al.* 2020), with the latter's emergence in 2019 leading to the ongoing COVID-19 pandemic, currently estimated to have caused around 18 million deaths by the end of 2021 (Wang *et al.* 2022). Diseases with animal origins are estimated to make up over 60% of all known emerging infectious diseases (Jones *et al.* 2008), and it is well-recognised that control of zoonoses is an important component in reducing risk of future pandemics (UNEP & ILRI 2020).

1.2 The role of zoonotic host ecology

Zoonotic spillover is fundamentally a process in a complex socioecological system involving zoonotic hosts, humans, and pathogens. Therefore, the unique traits, behaviour, and spatiotemporal distribution of zoonotic hosts play an essential role in the chain of transmission (Morse *et al.* 2012). The role of zoonotic host ecology in the pathway to zoonotic spillover can take diverse forms for different pathogens and even for the same pathogen in different locations, due to the variety of disease system structures (Figure 1.1) and different modes of transmission. As a result, any given zoonotic disease system has unique and nuanced socioecological processes underpinning spillover. The spatiotemporal distribution of zoonotic disease incidence can therefore appear to be cryptic and unpredictable (Plowright *et al.* 2017).

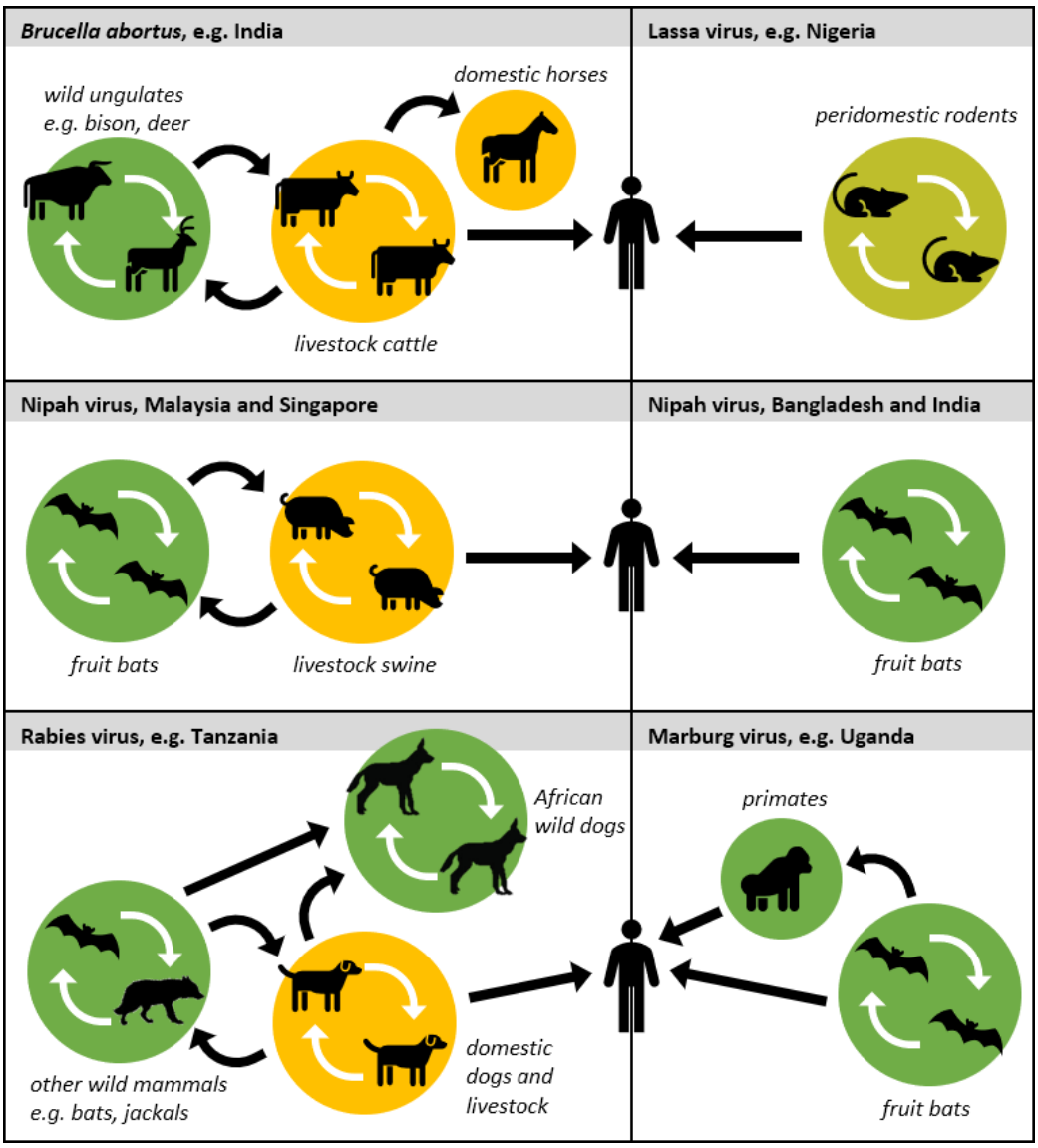


Figure 1.1: Example zoonotic disease dynamics. Green circles contain wild communities, while yellow circles contain domestic and livestock animals. Green-yellow circles contain peridomestic species which commonly live in and around human settlements. Arrows indicate common transmission routes. Summarised from Deka et al. (2019) (*Brucella abortus*), Olayemi & Fichet-Calvet (2020) (Lassa virus), Clayton et al. (2012) (Nipah virus), Lushasi et al. (2021) (rabies virus) and Feldmann et al. (2020) (Marburg virus). Domestic horses are believed to be dead-end hosts of *B. abortus* (Moreno 2014), but are relevant to disease burden due to the cultural and financial significance of horse ownership. African wild dogs do not contribute to the rabies virus transmission cycle, probably because they are critically endangered, but are shown on the transmission diagram due to the negative impact of rabies infection on conservation efforts (Lushasi et al. 2021). Similarly, endangered Ethiopian wolves are threatened by rabies virus in Ethiopia (Randall et al. 2004). Primates and some other mammals are believed to be either dead-end, intermediate, or reservoir hosts of Marburg virus (Feldmann et al. 2020), so transmission from this group to humans is uncertain. Bats are known to transmit rabies to humans but in the Tanzanian example given, this is very rare (Lushasi et al. 2021).

Since climate, landscape, and human pressures vary across space and time, so do animal populations, pathogen transmission dynamics, and resulting risk of zoonotic spillover and disease. Studying ecology, the way in which a species interacts with and depends on their environment, provides an opportunity to quantify how the spatiotemporally variable environment ultimately impacts incidence of zoonotic spillover and disease; a relationship termed "disease ecology" (Ostfeld 2015). Disease ecology also depends on qualities of the pathogen such as pathology, persistence in the environment, and host plasticity, and human density, behaviour, and vulnerability (Wilcox & Gubler 2005; Plowright *et al.* 2017). Zoonotic host and pathogen ecology provides a necessary foundation for zoonotic disease ecology since ecological processes such as population dynamics determine the spatial and temporal distribution of a pathogen and where it can be shed, here referred to as "zoonotic hazard" (Gibb *et al.* 2020a). It is now recognised that not only are insights into host-pathogen ecology valuable for modelling zoonotic disease, but that these insights are critical to accurately predicting and preventing zoonotic spillover (Hayman *et al.* 2013; Gibb *et al.* 2020a).

Exploring zoonotic host ecology through large-scale analyses has led to research into whether some species or communities are more likely to be hosts, with the aim of generating predictions which can be applied across systems. Broad principles have been identified through large studies or meta-analyses, such as that species with larger body sizes tend to host fewer pathogens (Stephens *et al.* 2016). Gibb *et al.* (2020b) found that animals in disturbed landscapes are significantly more likely to host human-shared pathogens than animals in nearby undisturbed habitats, particularly in rodents, bats, and passerine birds, suggesting that traits that make animals more resilient to human pressure may also increase their propensity for hosting zoonotic disease. This finding helped to explain inconsistent observations of the dilution effect, the idea that zoonotic spillover is less likely to occur in biodiverse areas due to the reduced frequency of effective hosts (Ostfeld & Keesing 2000; Schmidt & Ostfeld 2001; Keesing & Ostfeld 2021a). By understanding the relationship between human disturbance and zoonotic host diversity, it may be possible to target surveillance of disease systems to more susceptible landscapes, and to incorporate disease risk into land-use planning. Large-scale ecological modelscan therefore be a valuable tool through which to understand broad drivers of zoonotic disease.

One of the biggest research areas in the field of ecology is understanding the historic and future impacts of anthropogenic change on global species communities (Gimenez *et al.* 2014). Climate and land-use change have both had devastating impacts on global ecosystems which are only predicted to increase in magnitude (IPBES 2019). Emergence of zoonotic disease has for some time been linked with anthropogenic changes (Jones *et al.* 2008), perhaps due to the relative resilience of zoonotic hosts to human disturbance (Gibb *et al.* 2020b). Indeed, spillover prevention is seen as a potential opportunity to increase funding for conservation initiatives (Glidden *et al.* 2021). Due to the complex ecological dynamics underpinning zoonotic disease, it is impossible to make accurate predictions about future zoonotic spillover or disease incidence under novel environmental conditions. Instead, improving our understanding of zoonotic host ecology provides an opportunity to create more informed and nuanced predictive models which are more likely to accurately capture the nonlinear dynamics which ultimately lead to zoonotic spillover and disease. In this way, species ecology is essential to predicting how different scenarios of anthropogenic change will affect global health, which will feed into decision-making for sustainable development, biodiversity conservation, and climate change mitigation (Wilcox & Gubler 2005; Pongsiri *et al.* 2009).

1.3 Challenges in monitoring and control

Despite the importance of non-human hosts in the zoonotic disease system, monitoring zoonoses in non-humans is challenging due to the historically disparate policy areas of human health and animal health (Jerolmack 2013; Johnson *et al.* 2018). With animal health typically a lower priority, capacity to monitor livestock, domestic animals, and wildlife is particularly limited in low-income countries which are disproportionately impacted by zoonotic disease burden (Keusch *et al.* 2009). Monitoring of zoonotic hosts can provide insights into drivers of zoonotic spillover and disease incidence which can be used to create predictive models and inform intervention activities (Carroll *et al.* 2018), and screening of potential zoonotic hosts is also suggested for predicting disease emergence (Howard & Fletcher 2012). Such activities are made increasingly feasible by the identification of risky species or hotspots in space and time based on large-scale studies and historic discovery which can help target surveillance (Levinson *et al.* 2013; Allen *et al.* 2017; Olival *et al.* 2017; Gibb *et al.* 2020b). However, monitoring of animals – particularly wildlife – is costly and, if

considered to be for the benefit of human health, should be weighed against alternative activities which would benefit health (Holmes *et al.* 2018).

Monitoring the human incidence of disease is perhaps a more tractable way to explore and predict incidence of zoonotic disease than monitoring of livestock and wildlife. Disease incidence estimates in humans can inform strategies to research and mitigate disease (Grace *et al.* 2012; Mathers *et al.* 2013), including allocation of healthcare provisions, setting funding and research priorities, and evaluation of emergence and epidemic risk. Depending on the purpose of a monitoring effort, different measures of the effects of zoonotic disease are relevant; to allocate healthcare provision, a measure such as DALYs helps inform how healthcare could improve quality of life (Torgerson *et al.* 2018), while to assess likelihood of outbreak or emergence, number of spillover or infection events may be more useful. It is important to distinguish between infection and disease, since infection does not always lead to disease, and these quantities are measured and monitored in different ways; for instance, serosurveys which measure pathogen-specific antibodies in human sera cannot typically distinguish between symptomatic disease and asymptomatic infection (Turgeon 2020). Monitoring human health comes with its own challenges, especially in low-income countries, where disease is often underreported through passive sentinel systems such as in-country public health records (Burniston *et al.* 2015). Meanwhile, active surveillance activities, such as seroprevalence studies, when not coordinated at a high level can have biases (Townsend Peterson *et al.* 2014) which can in turn preclude accurate infection incidence estimates.

Due to challenges in monitoring and other factors, unique challenges are present in the development of strategies to manage and mitigate zoonotic disease. For non-zoonotic diseases whose most common source of infection is human-to-human transmission, interventions often involve reducing risky direct or indirect contact among humans (e.g., Mahase 2020; WHO 2021). When considering zoonotic transmission, the same strategy requires humans to avoid direct contact with the zoonotic host or contact with the shed pathogen, which is easier in some contexts than others. For instance, risk factors for exposure to ebolaviruses include consuming or hunting wild meat and visiting wild meat markets (Leroy *et al.* 2004; Dobschuetz *et al.* 2019; WHO 2021a). While it is physically possible to avoid wild meat and wild meat markets, it may be economically impractical for people to obtain food from elsewhere or to seek other income, and top-down interventions can result in unintended social and economic consequences (Bonwitt *et al.* 2018). Even identifying risk

factors to avoid is challenging since there are many potential hosts and transmission modes across zoonotic diseases. The context around avoiding zoonotic pathogens can be complex and involve socioeconomic factors including education (Saylors *et al.* 2021), culture, tradition, and religion (Burniston *et al.* 2015), and profession (Swai *et al.* 2010). It is therefore vital to engage local people and create effective and context-specific community-led interventions for the prevention and management of zoonotic disease (WHO *et al.* 2011).

It can be challenging to scale up local mitigation strategies and coordinate them at a national or international level (Bardosh *et al.* 2017), which may preclude efficient large-scale mitigation of disease through local strategies. Meanwhile, vaccination has historically been a highly effective large-scale intervention for eradicating and reducing disease, such as the major success of the global eradication of smallpox, preventing an estimated 40 million deaths out of 350 million potential new cases in the 20th century (Ehreth 2003). A study by Toor *et al.* (2021) estimated that between 2000 and 2030 between 80 and 120 million deaths will have been averted due to vaccination against 10 key pathogens (excluding SARS-CoV-2). While vaccination of humans does reduce zoonotic disease incidence and risk of human-to-human transmission, eradication cannot result from a human-only vaccination programme, since the pathogen persists in the maintenance community. Instead, to achieve disease elimination human populations would have to be periodically vaccinated. The limitations of human vaccination in controlling zoonotic disease may make the funding and development of such vaccines unappealing when comparing the resources required with the benefits. Additionally, coverage of human vaccination is limited by increasing levels of vaccine hesitancy (Gostin *et al.* 2020). Promisingly though, the development of vaccines against both Lassa virus and Nipah virus have now reached human clinical trials (CEPI 2020; CEPI *et al.* 2021).

In theory, eradication of zoonotic disease could be possible through vaccination targeted at the maintenance community, however this carries a variety of challenges, especially if this community is made up of wild species. When this community is composed of wildlife, this strategy can be hindered by imperfect immunity due to, for example, patchy vaccine coverage or variation in vaccine efficacy across the population (Barnett & Civitello 2020). Oral vaccination has been adopted in some cases to combat the logistical challenges of vaccine coverage (Cross *et al.* 2007) and self-disseminating vaccines (Nuismer & Bull 2020) and trait-based vaccination (Barnett & Civitello 2020) have been proposed as a new direction for

wildlife vaccine deployment. Vaccination of zoonotic hosts is more practical for domestic and livestock species, which is highly cost-effective for several zoonoses, with wide vaccination coverage reducing disease incidence in both animals and humans, including rabies (Layan *et al.* 2021) and brucellosis (Moreno 2014). Vaccination of domestic animals and livestock carries a secondary financial and social benefit of protecting high-value animals from disease, therefore vaccination of livestock can be strategically better than vaccinating humans – since humans are in turn protected against livestock-human transmission – especially if livestock are the principal zoonotic hosts or form the maintenance community.

It is now estimated that the costs of primary pandemic prevention activities such as surveillance are outweighed by the costs of not responding until a zoonosis emerges (Bernstein *et al.* 2022). Despite the challenges facing both human and animal monitoring of zoonotic disease and pathogens, recent capacity building – for instance, surveillance improvements through the Emerging Pandemic Threats programme (USAID 2021) and the Sentinel programme (Botti-Lodovico *et al.* 2021) – in combination with increased motivation following the realised risk of SARS-CoV-2 emergence presents opportunities for enhanced monitoring going forwards (FAO 2021). To make best use of this surveillance, models are needed which can predict incidence of spillover of endemic and novel zoonoses, and which can inform on the likely impacts of mitigation methods.

1.4 Modelling approaches and opportunities

1.4.1 Phenomenological and mechanistic modelling in ecology

There are many different approaches which can be taken when developing a model. Occam's razor posits that – all else being equal – of two competing explanations, the one with fewer entities is to be preferred (Duignan 2021). In a similar vein Albert Einstein is famously paraphrased as saying "Everything should be made as simple as possible, but no simpler" (Robinson 2018). On the face of it, these philosophical viewpoints suggest that simpler models are generally better models. Ironically, Einstein's exact words in a lecture aimed at reducing the complexity of theoretical physics were "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience" (Einstein 1933) . Perhaps the original, extended, quote demonstrates the subjectivity of the word "simple". While his phrasing seems verbose, Einstein may have felt that every word of his sentence was necessary to adequately represent his assertion (or

perhaps the phrasing was, in itself, a joke). Modelling is a form of theory which describes the irreducible basic elements which Einstein describes, and how they interact.

Phenomenological modelling has long been used to identify patterns in observations, an important part of the scientific process which has ultimately led to development of theory. When analysing patterns through phenomenological modelling, aside from accounting for control variables, the processes which give rise to these patterns are neglected. Process-based modelling (also "mechanistic" – see Box 1.1) explicitly describes the processes underlying a system. In doing so, current theory informs the model structure, and control or confounding variables are explicitly incorporated into the framework. However, this model structure runs the risk of inaccurately describing a system by incorporating incorrect theory. Taking Einstein's (extended) quote as fact, an advocate for process-based modelling would posit that explicit descriptions of processes are needed to adequately represent scientific experience, while an advocate for phenomenological modelling might counter that those complex processes muddy the identification of the essential, simple, basic elements. Which, then, is "as simple as possible, but no simpler"?

Phenomenological model: Describes a system by seeking patterns in empirical data.

Process-based model: Describes a system through recognition of its component processes and first principles. For example, mass-action mathematical population models are process-based models.

Mechanistic model: A subclass of process-based models. Describes a system through the ecology of its fundamental units. For instance, individual-based models could be classed as mechanistic models, while mass-action population models would not. Alternatively (not here), sometimes used synonymously with "process-based model".

Box 1.1: Phenomenological and process-based modelling in ecology.

The debate around the relative appropriacy and value of phenomenological (also termed "correlative" or "purely statistical") models versus process-based models has been present in statistical and mathematical ecology for decades (Aber 1997), as ecology has moved from hypothesis testing to parameter fitting and model comparisons (Connolly *et al.* 2017). Phenomenological and process-based modelling approaches both have value in modern ecology, with the former usually used for identifying broad patterns such as the impact of conservation strategies on biodiversity (Gray *et al.* 2016), and the latter for modelling systems such as ecological networks (Ings *et al.* 2009) and animal movement (Goossens *et al.*

2020) which are considered to be dynamic. In reality, many ecological models incorporate some aspects of both modelling approaches. Take, for instance, species distribution models (SDMs). Correlative SDMs relate observed species occurrence with environmental data to predict a species' distribution (Elith & Leathwick 2009). Meanwhile, mechanistic SDMs build this prediction from the bottom-up by explicitly considering the environmental constraints on a species' distribution through established ecology of the species such as functional trait analysis and physiology (Kearney & Porter 2009). It is challenging to create accurate mechanistic SDMs because this requires exhaustive insight into the species, and conversely correlative SDMs are unlikely to be robust to novel conditions which represent variables outside of the sample space (Fitzpatrick & Hargrove 2009). A combination of correlative and mechanistic SDMs, representing a blend of phenomenological and process-based approaches, can result in models which make best use of the available data and knowledge (top-down observations of occurrence and bottom-up mechanisms) while maintaining accuracy and robustness (Rougier *et al.* 2015; Pertierra *et al.* 2020).

A growing trend of process-based, bottom-up modelling in ecology has taken place alongside methodological improvements allowing such models to be informed by large datasets and take into account biases such as data collection methods and spatiotemporal autocorrelation (Gimenez *et al.* 2014). With an increased focus of the effects of anthropogenic change on ecology (McCallen *et al.* 2019), models incorporating causal mechanisms are considered more appropriate for understanding how ecological processes will be altered by the changing environment (Gimenez *et al.* 2014). The process-based approach therefore appears to be a useful tool for zoonotic disease ecology, since zoonotic disease systems are dynamic and are likely to continue to vary under anthropogenic changes (Estrada-Peña *et al.* 2014; Gibb *et al.* 2020a; Keesing & Ostfeld 2021b).

1.4.2 Modelling challenges in zoonotic disease ecology

The zoonotic disease system is a complex network of processes which interact with one another, producing nonlinear dynamics. However, it is helpful to simplify this network by conceptualising it as a series of layers which may result in successful infection events between individuals and populations (Plowright *et al.* 2017). I will demonstrate this conceptual layered system for a "simple" zoonosis with a single reservoir host and no human-to-human transmission – meaning that the pathogen is both solely maintained within and transmitted from one species or population (Figure 1.2). Each layer represents a

set of processes which drive and restrict all layers above it – although they can also influence layers below. I have viewed the zoonotic disease system from the perspective of the reservoir host ecology, however the same system from an alternative view would expand human processes (for example, human exposure is driven by human population density and behaviour) and could simplify others.

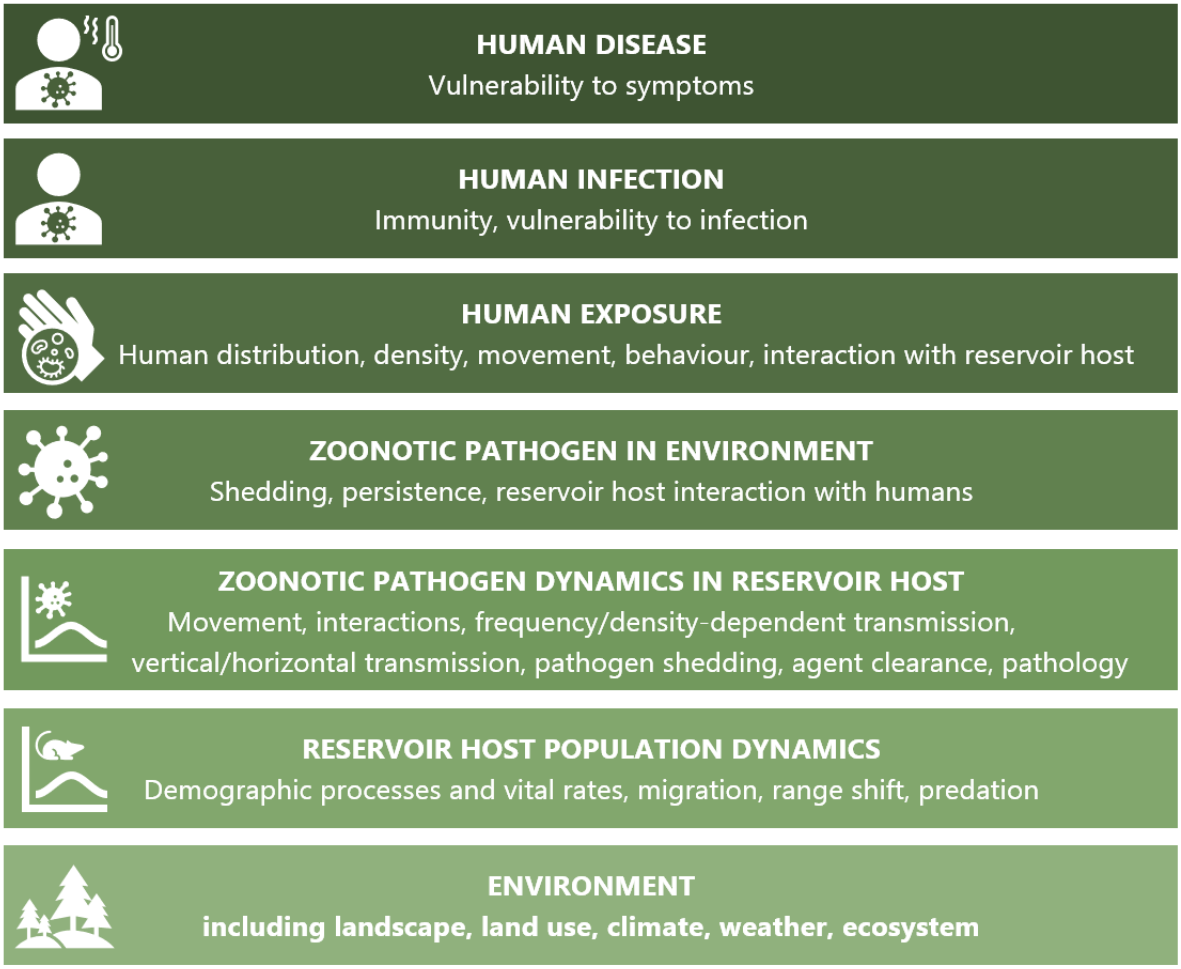


Figure 1.2: Conceptual layered framework for a straightforward zoonotic spillover pathway. Each layer is driven by processes described in its box and is fundamentally restricted by the layer below. For example, human disease is only possible where human infection has occurred, but also depends on vulnerability to symptoms. Layers may also be affected by other layers above and below. For example, reservoir host population dynamics are fundamentally driven by the environment due to resource availability, habitat quality and other environmental limitations that drive processes (e.g., survival, recruitment, migration) which change the distribution and density of the reservoir host. Meanwhile, if the zoonotic pathogen can induce disease in the reservoir host, pathogen dynamics – the layer above – will also impact population dynamics.

Any model relating to zoonotic disease includes or makes simplifying assumptions about these layers, either explicitly or implicitly, since fundamentally these processes feed into one

another to lead eventually to human infection and disease. However, it is rare for data or theory fully informing every layer in the zoonotic disease system to be available to modellers (Scoones *et al.* 2017); for instance, the factors determining quantity and timing of pathogen shedding in many animals are still poorly understood (Peel *et al.* 2018). Instead, simplifying assumptions must be made. The accuracy of a model in capturing the zoonotic disease system depends on which layers are informed by theory or data, and which are assumed constant. As a result, models informed only by human disease data neglect the underlying processes and implicitly assume that these dynamic processes are constant. Incorporating some types of observations and not others into a model can lead to different implicit assumptions about underlying processes and how they interact. The challenge in constructing a model of zoonotic disease is therefore determining which layers of the system can be simplified, and which are necessary to accurately answer the question which is being asked (Alexander *et al.* 2012).

Current modelling approaches and opportunities for development

A common integrated method for modelling zoonotic disease is identification of the "zoonotic niche". A zoonotic niche is found by correlating occurrence of a zoonotic host, reservoir, pathogen, or disease cases with environmental factors, and is then projected onto a map (e.g., Blackburn *et al.* 2007; Pigott *et al.* 2014, 2015; Chalghaf *et al.* 2016). An analogue to the SDM, zoonotic niche models can include varying degrees of detail. Models of predicted zoonotic disease risk from zoonotic niche models can be used to inform surveillance across broad spatial scales and aid differential diagnosis (Alexander *et al.* 2012; Pigott *et al.* 2015). However, these models usually present an indirect measure of zoonotic disease risk; for instance, a model of zoonotic hazard only has limited application to understanding or predicting zoonotic disease incidence. Additionally, these models are not temporal and, depending on which processes are modelled, may not be appropriate for fine-scale descriptions of disease risk (Plowright *et al.* 2017) or predicting nonlinear impacts of environmental changes (Ehrlén & Morris 2015). Zoonotic niche models suffer from the same limitations as correlative SDMs, such as reporting bias (Townsend Peterson *et al.* 2014) and the choice of presence-absence or presence-only formulation (Zeimes *et al.* 2012). Additionally, zoonotic niche models only predict the probability of presence of zoonotic hosts, pathogens, or cases of disease, not their relative abundance or density which will have important implications for total incidence of zoonotic spillover and disease (Alexander *et al.* 2012). As such, the zoonotic niche framework – while useful – does not present a complete

view of the zoonotic disease system; although building transmission mechanisms and socioeconomic variables upon this framework has been achieved to assess impacts of different scenarios on disease risk (Redding *et al.* 2019).

There is an increasing development of process-based models to understand the mechanics underlying specific zoonotic disease systems. In terms of the zoonotic pathway (Figure 1.2), these models describe one or more layers in isolation. For instance, the understanding of the epidemiology of zoonoses of bats is limited by knowledge of within-host and between-host pathogen dynamics since these processes impact the timing of "episodic" shedding and persistence of the virus through hibernation (George *et al.* 2011; Hayman *et al.* 2013). Process-based models of zoonotic disease dynamics in bat populations have therefore focussed on elucidating these dynamics, such as by modelling bat immune responses and host-to-host epidemiology (Gentles *et al.* 2020). The epidemiology of zoonoses of rodents has also been explored through process-based models; for example, exploring the effect of seasonality on spillover probability (Nandi & Allen 2021) and comparing the contribution of environmental pathogen persistence and zoonotic host abundance on human disease incidence (Carver *et al.* 2015). Models of individual processes, or layers, in zoonotic disease systems help to determine which processes are important and how they should be described, making them valuable for building accurate and robust bottom-up models of the full system.

It is rare to find models of zoonotic disease systems which capture most of the constituent dynamics through process-based approaches. This gap could be due to the enigmatic nature of different zoonotic disease systems in different locations which makes it inappropriate to apply one method across multiple contexts, requiring independent research into each individual system to construct a complete model. The lack of process-based models of zoonotic disease systems could also represent perceived research priorities; for influenza viruses, process-based models have usually focussed on human-to-human infection, with few modelling swine and bird host dynamics (Dorjee *et al.* 2013). Whilst being able to accurately describe a zoonotic disease system through a bottom-up model opens important avenues of exploration (Hassell *et al.* 2021), reaching this point appears to be highly research intensive and currently only possible for well-resourced zoonoses (Alexander *et al.* 2012).

While the ultimate goal of modelling zoonotic disease systems may be to accurately capture relevant dynamics in a single framework, at present few – if any – zoonoses are studied in

this complete manner. It is important to design data collection around the models which they will inform, and vice versa, to maximise resources and model usefulness (Restif *et al.* 2012), and this integrated approach will give the best chance of an accurate understanding of zoonotic disease systems and management recommendations. However, given the urgent implications of zoonotic disease on global health, there is also merit in developing methods which use data currently available, since data collection is a time-consuming activity; longitudinal zoonotic host studies are called for (Guth *et al.* 2022) which could take place over the scale of years or decades. There is therefore an opportunity in zoonotic disease modelling for novel, interdisciplinary thinking around types of data which can inform on system processes and work towards the delivery of timely modelling frameworks. The neglected zoonotic disease Lassa fever, outlined in detail in the next part of this chapter, presents a valuable case study around which to design such a framework.

1.5 Lassa fever as a global health priority

Lassa fever is a disease caused by infection with Lassa virus, an arenavirus principally transmitted to humans by the rodent *Mastomys natalensis* (McCormick *et al.* 1987). Lassa fever is endemic and restricted to West Africa, with occasional exportation to other countries garnering international attention (Amorosa *et al.* 2010; Ehlkes *et al.* 2017; Overbosch *et al.* 2020; UK Health Security Agency 2022). As a disease which disproportionately affects those living in poverty, surveillance of Lassa fever and healthcare for those affected is limited and heterogeneous despite substantial disease burden (Richmond & Baglole 2003; Gibb *et al.* 2017). In the last decades Lassa fever has been recognised as a priority disease in terms of protecting biosecurity (Schuler 2005), and reducing endemic disease burden and epidemic risk (WHO 2015; Røttingen *et al.* 2017). While historically the annual incidence of human Lassa virus infection has been quoted as 100,000 to 300,000 in West Africa, these figures were based on outdated and spatially-restricted serological surveys (Gibb *et al.* 2017), and a more recent analysis estimates at least 890,000 annual human Lassa virus infections (Basinski *et al.* 2021). To date, no formal studies have estimated the financial or health burden of Lassa fever.

The early symptoms of Lassa fever are nonspecific, making it difficult to distinguish from other febrile diseases such as malaria, which has high incidence across much of West Africa (WHO 2020c), without laboratory testing (Raabe & Koehler 2017). More advanced symptoms which often result in fatality can include bleeding, organ failure, shock, and haemorrhaging

(Yun & Walker 2012). The incubation period of Lassa fever is estimated as ranging from 6 to 21 days (WHO 2017) but there are no published data or analyses verifying this. Sensorineural deafness is estimated to occur in one third of Lassa fever survivors and may occur even when acute symptoms of Lassa fever were not present (Cummins *et al.* 1990). Infection with Lassa virus is thought to lead to symptomatic disease in 20% of cases (WHO 2017), although this proportion may be lower as limited surveillance through serosurveys make it challenging to assess the extent of asymptomatic infection (Gibb *et al.* 2017). Meanwhile, it is estimated that fatality occurs in 15% of hospitalised cases (WHO 2017). Pregnant women are at particularly high risk of complications of Lassa fever, with mortality estimated to be around three times higher for a pregnant woman compared with a non-pregnant woman (Kayem *et al.* 2020); during the third trimester of pregnancy, maternal death or foetal loss occurs in over 80% of cases (WHO 2017). There is no cure, and currently no approved vaccine, for Lassa fever, but the nonspecific antiviral drug ribavirin is advised as a prophylaxis and may reduce mortality by a factor of more than 88% if administered in the first six days of clinical disease (McCormick *et al.* 1986). However, it is difficult to diagnose Lassa fever this rapidly, ribavirin is contraindicated in pregnancy, and there are no published data of the effectiveness of ribavirin in humans since the 1986 study (Hallam *et al.* 2018).

The principal host of Lassa virus is a reservoir host, meaning that infected *M. natalensis* are the primary source of human infection and that populations of this species form maintenance communities for the virus. The most common mode of human infection is believed to be via the excreta of infected rodents which have entered homes seeking shelter or food (Bonwitt *et al.* 2017). It is unclear how long the pathogen persists in either the domestic or exterior environment (Smither *et al.* 2020), or where in particular the pathogen tends to be shed, with suggestions that infected rodents can contaminate food stores (Bonwitt *et al.* 2017). Another possible mode of infection is consumption of the meat of infected wild-caught rodents (Ter Meulen *et al.* 1996; Bonwitt *et al.* 2016) which is considered high in nutritional value and flavour (Bonwitt *et al.* 2016). Lassa virus can also be transmitted between humans, and there has been debate in recent years over the degree to which this occurs. It now appears that while human-to-human transmission is rare (Siddle *et al.* 2018), super-spreading events in hospitals have occurred (Lo Iacono *et al.* 2015), where the practice of good hygiene and use of personal protective equipment should minimise risk of human-to-human transmission (Helmick *et al.* 1986).

The Lassa fever disease system is simpler than some other zoonotic diseases since the roles of principal host and reservoir of Lassa virus are occupied by the same species. While some systems are difficult to visualise and describe due to many interacting reservoirs and hosts (Alexander *et al.* 2012), Lassa virus can be visualised as a simple system consisting only of interactions between *M. natalensis* and humans. However, the dynamics of Lassa fever appear complex and cryptic due to the striking seasonality of outbreaks (Gibb *et al.* 2017), the increase in diagnosed cases in recent years (Roberts 2018), and uncertainty around the degree to which human-to-human transmission contributes to outbreaks (Lo Iacono *et al.* 2015; Siddle *et al.* 2018). These complex dynamics probably arise due to variable environmental conditions impacting the reservoir host (Olayemi *et al.* 2018; Redding *et al.* 2021), increasing but heterogeneous surveillance (Ilori *et al.* 2019; Redding *et al.* 2021), and a paucity of genomic data from the endemic community (Siddle *et al.* 2018). Deciphering the nature of Lassa fever epidemiology requires the continued addressing of some of the many open research questions.

Mitigation and surveillance

In terms of preventative measures, modelling suggests that annual rodent control measures would likely be ineffective due to their high fecundity (Mariën *et al.* 2019). However, continuous rodent control or rodent vaccination could reduce disease (Mariën *et al.* 2019), and there are now a number of human vaccines for Lassa virus progressing through clinical trials (Hallam *et al.* 2018). Until these measures are available, the only mitigation strategy for Lassa fever will continue to be the reduction of human-rodent contact (Bonwitt *et al.* 2017). In Nigeria, one of the strategies for reducing risky contact is education via health workers, although this appears to have been limited in its effect (Oladeinde *et al.* 2014; Akinwumi *et al.* 2016). Other suggestions have included measures to reduce the number of rodents in and around houses such as improved housing or housing repairs, and rodent-proof food storage, though no major campaigns for these measures exist (Mari Saez *et al.* 2018).

Passive sentinel systems provide a foundation for Lassa fever reporting across the recognised endemic area, for instance, in Nigeria suspected cases are reported to local and public health authorities (NCDC 2018). WHO, the Ministries of Health of Guinea, Liberia, and Sierra Leone, and other partners established the Mano River Union Lassa Fever Network to improve research, surveillance, management and control of Lassa fever, especially diagnostic facilities (Khan *et al.* 2008). Lassa fever is notifiable to Integrated Disease Surveillance and Response

(IDSR), a strategy implemented by the WHO African region to collate, report, and respond to disease outbreaks (WHO Regional Office for Africa 2019). Based on passive reporting processes, Lassa fever appears to be endemic to Nigeria and the Mano River Union (Guinea, Sierra Leone, and Liberia) with only occasional infection in other parts of West Africa (Figure 1.3). However, active surveillance in the form of serological surveys suggest that infection with Lassa virus, or a cross-reactive arenavirus, also occurs in Mali, Côte d'Ivoire, Ghana, and Burkina Faso (Richmond & Baglole 2003). Genomic analysis of recent infections originating in Benin and Togo provide more substantive evidence that Lassa virus is endemic in these countries (Whitmer *et al.* 2018; Yadouleton *et al.* 2020), while a number of cases have been exported from outside of Nigeria and the Mano River Union (Sogoba *et al.* 2012). This evidence suggests under-surveillance of Lassa fever in the central portion of West Africa rather than an absence of disease. Additional active surveillance of Lassa fever takes place in the form of a contact tracing programme carried out by the Nigeria Centre for Disease Control (NCDC 2018).

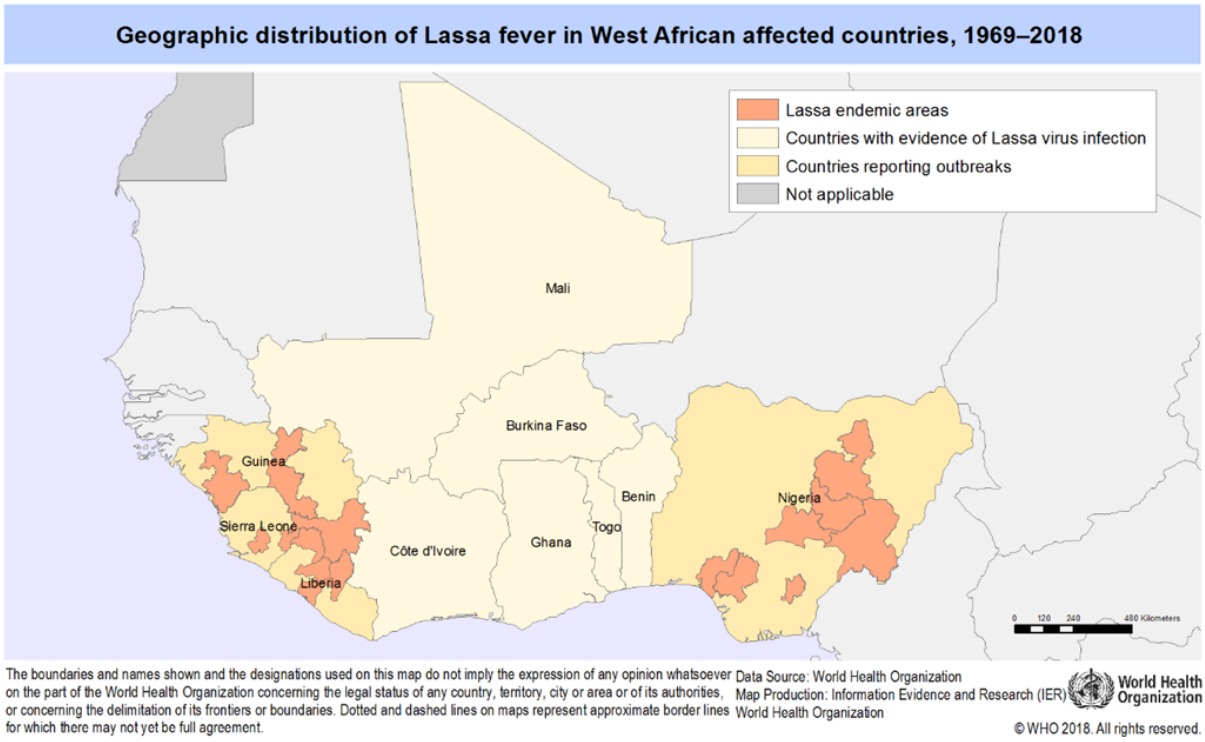


Figure 1.3: Map of reported distribution of Lassa fever in West Africa. © World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.

1.6 *Mastomys natalensis,* the principal host and reservoir of Lassa virus

Essential to the context of spatiotemporal distribution of Lassa virus infection and disease incidence is *M. natalensis*, the principal host and reservoir of Lassa virus. *M. natalensis*, or the

natal multimammate mouse, is one of the most widespread rodents in sub-Saharan Africa (Coetzee 1975). Its ecological range includes natural savannahs as well as landscapes which have been altered by humans (agricultural fields and human habitations) (Fichet-Calvet *et al.* 2007). They are omnivorous and in the wild tend to live for less than two years. The species has an extremely high propagation rate (Figure 1.4), with sexual maturation of females occurring after only 94 days, a gestation period of 23 days, and an oestrous occurring only 25 days after giving birth (Coetzee 1975). *M. natalensis* is an agricultural pest, motivating studies of their populations across sub-Saharan Africa in terms of potential for mitigation (Sluydts *et al.* 2009).

Figure 1.4: Photograph of wild M. natalensis, *the principal host and reservoir of Lassa virus*. A pregnant female in a group nest, with juveniles of different ages, from multiple litters. © Joe Blossom, Alamy Stock Photo; 2009. Reproduced under licence.

Wild *M. natalensis* population dynamics are highly seasonal, with high recruitment rates in the rainy season and subsequent population booms in the dry season (Christensen 1993; Makundi *et al.* 2007; Mari Saez *et al.* 2018). It is not clear whether these population dynamics are purely phenological, with one theory being that vegetation growth signals ideal reproductive timing (Firquet *et al.* 1996), or opportunistic, with increased resources and habitat leading to improved chances of pregnancy. As a result, it is uncertain whether higher-

or lower-than average precipitation leads to differences in population dynamics; in other words, whether inter-annual variability in seasonality causes timing and size of population booms to differ between years. When considering transmission of Lassa virus to humans, an additional layer of complexity is added by rodent behaviour, movement, and migration. *M. natalensis* is a peridomestic animal, meaning it is often found in human-habited areas, not through human-seeking behaviour but due to resource-seeking behaviour which leads it to human settlements (Bonwitt *et al.* 2017). It is thought that, in agricultural landscapes in West Africa, *M. natalensis* migrate from crop fields to houses during the dry season when the crops are harvested, perhaps following the movement of food (Fichet-Calvet *et al.* 2007). Alternatively, this observation could be attributed to an increased population size (following the population boom) and reduction in natural resources (due to adverse climatic conditions or depletion of resources), therefore causing rodents to seek alternative food sources in rural settlements. The drivers of the timing of the movement and population dynamics of *M. natalensis* are the subject of ongoing research and are important for elucidating the dynamics of Lassa virus.

Lassa virus sits within the genus mammarenavirus, of which six others are known to infect *M. natalensis* in different parts of its range (Figure 1.5). Of these seven mammarenaviruses, Lassa virus is the only known to be capable of infecting humans. In different parts of its range, *M. natalensis* is divided into different clades or lineages which can be distinguished by differences in mitochondrial DNA (De Bellocq *et al.* 2020). Excepting Luna virus and Mopeia virus which both infect the same mitochondrial lineage of *M. natalensis* (albeit in different parts of its range), each discovered mammarenavirus is unique to one rodent lineage. Even where different lineages are in contact, no evidence of arenavirus sharing has been found (Gryseels *et al.* 2017), suggesting that there are barriers to infection specific to each lineage which may have arisen from co-evolution of arenavirus and rodent (Mariën *et al.* 2017). These host-specific barriers are proposed as a reason for the geographic constriction of Lassa virus to West Africa (Gryseels *et al.* 2017). However, evidence suggests that another group of mammarenaviruses in rodents is capable of opportunistic host-switching (Irwin *et al.* 2012), meaning that the theory of lineage-restricted arenaviruses is still uncertain. As a result, it is unclear whether, if initially transported, Lassa virus has the potential to spread among *M. natalensis* in other parts of sub-Saharan Africa. This discussion is further complicated by the uncertain roles of other zoonotic host species in some parts of the Lassa virus endemic range (Olayemi *et al.* 2016b).

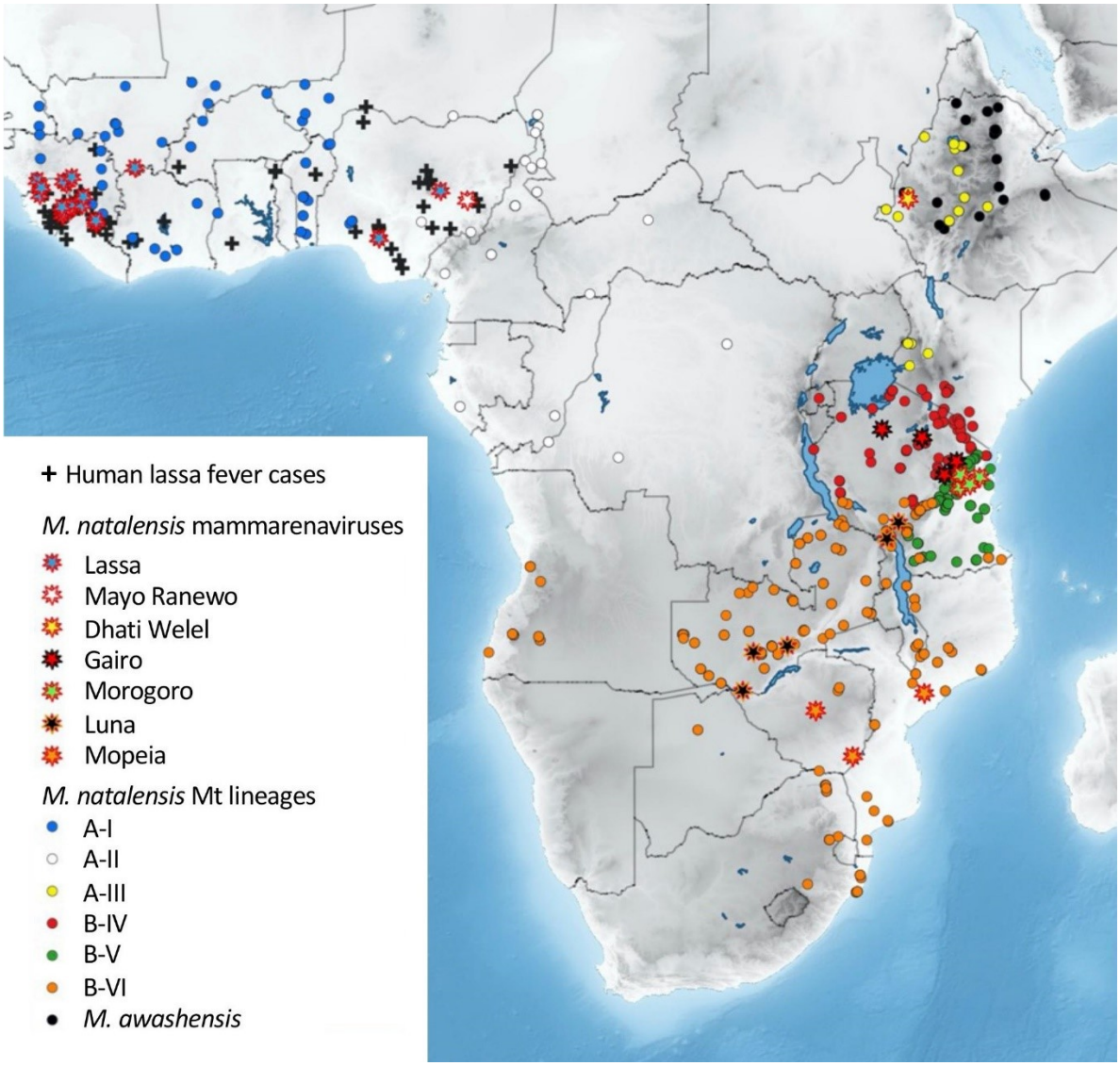


Figure 1.5: The geographic distribution of M. natalensis *lineages and mammarenaviruses which infect them.* Adapted from De Bellocq *et al.* (2020). Each mammarenavirus has been found in exactly one mitochondrial lineage of *M. natalensis*, and all but one lineage is associated with just one mammarenavirus, supporting hypotheses that mammarenaviruses are highly specialised to their rodent host. Lassa virus is the only mammarenavirus found to be capable of infecting humans, with Lassa fever cases shown with black plus signs. Captures of *M. awashensis* are also included on the map since evidence of transmission of the newly-discovered Dhati Welel virus between this species and *M. natalensis* was found in the study. © Journal of Vertebrate Ecology; 2020. Licence: CC-BY.

Longitudinal individual-based studies of species communities are useful for understanding demographic processes and population changes, especially for temporally variable populations like *M. natalensis* (Clutton-Brock & Sheldon 2010). Such studies are often carried out through capture-mark-recapture (CMR) methods in which an animal is captured, marked, and later captured again (Seber & Schofield 2019). As well as enabling abundance estimates due to repeat observations of the same individuals (Bright Ross *et al.* 2022), CMR studies also

provide an opportunity to collect data on individuals at the capture stage(s) which can provide insights into population and demographic dynamics. However, CMR studies in Lassa virus endemic countries are considered unethical, since the capture of a potentially infected rodent is an opportunity to reduce a potential source of infection. Instead, longitudinal studies in West Africa focus on rodent serology through capture-kill studies (Fichet-Calvet *et al.* 2007). One major CMR study of *M. natalensis* has continued since 1994 in Tanzania (Borremans *et al.* 2014; Mariën *et al.* 2020), providing potential opportunities to study rodent demography, albeit outside the range of the West Africa *M. natalensis* lineages. In this CMR study, Morogoro virus provides an opportunity to observe rodent-to-rodent transmission dynamics (Mariën *et al.* 2020). Horizontal transmission results in acute infection with transient shedding, and appears to be the main mode of transmission in the wild, mostly among the same generation (Borremans *et al.* 2015). Meanwhile, vertical transmission from mother to offspring is believed to lead to chronic infection with no recovery and persistent pathogen shedding (Borremans *et al.* 2015). Chronic Morogoro virus infections are thought to sustain the virus within a population with horizontal transmission increasing with abundance following population growth (Mariën *et al.* 2020). Neither Lassa virus nor Morogoro virus appear to have pathogenicity in *M. natalensis,* consistent with theories of co-evolution (Mariën *et al.* 2017). These dynamics have important implications for the structure of a host-to-host transmission model for Lassa virus.

1.7 Recent developments and remaining gaps in Lassa fever modelling

Spatial models of Lassa fever and Lassa virus have typically taken the zoonotic niche approach. Fichet-Calvet & Rogers (2009) produced the first "risk maps" of Lassa fever by predicting presence of Lassa fever from temperature, rainfall, and vegetation, using records of human disease collated from the literature. These models identified rainfall and temperature as the main spatial predictors of the presence of Lassa fever, with the Mano River Union and Nigeria having the most suitable conditions for Lassa virus endemicity (Figure 1.6A). This modelling framework incorporated areas of both presence and absence of human cases, however, it is possible that absence reflected lack of surveillance rather than true absence of disease. Mylne *et al.* (2015) (Figure 1.6B) took a similar approach but incorporated a separately-fitted SDM for *M. natalensis* using data from the Global Biodiversity Information Facility (GBIF). Interestingly, this model suggested that environmental suitability for Lassa virus is greater in Côte d'Ivoire than Nigeria. Data used in

models of Fichet-Calvet & Rogers (2009) and Mylne *et al.* (2015) likely suffer from spatial reporting bias due to spatial heterogeneity in Lassa fever awareness which, when not taken into account, can lead to inaccurate distributions (Townsend Peterson *et al.* 2014)

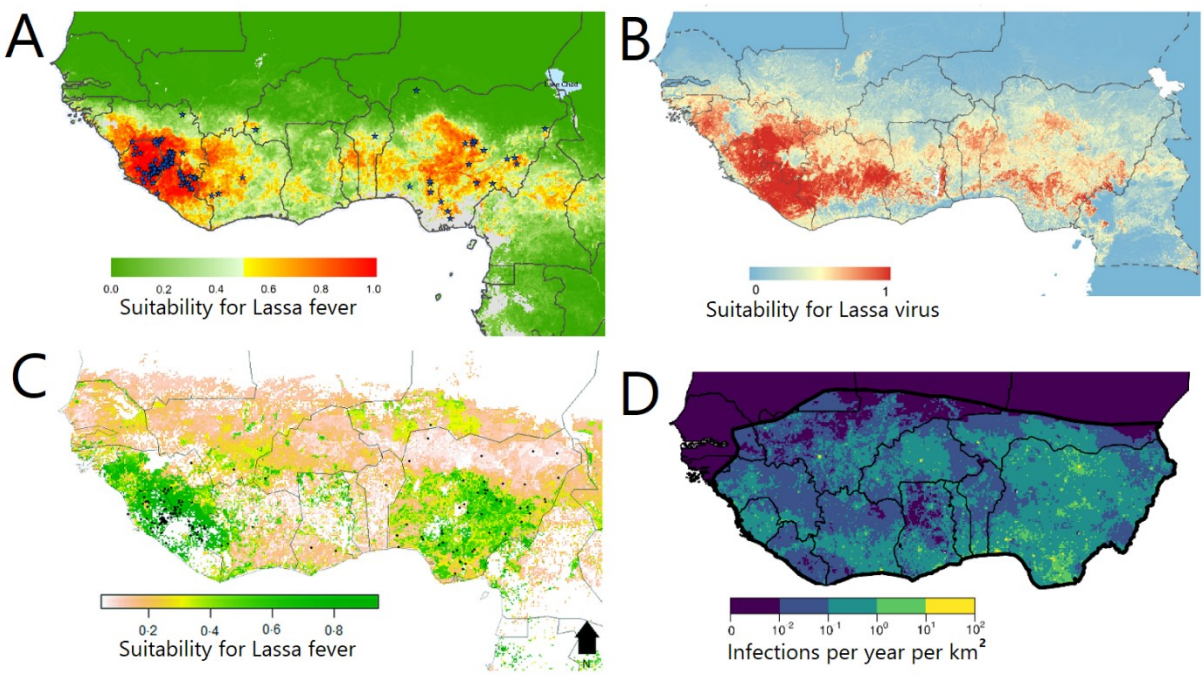


Figure 1.6: Spatial models of Lassa fever cases, Lassa virus presence, and Lassa virus infection. A: Predicted probability of Lassa fever presence based on a model correlating spatial environmental data with human disease records. Stars show human disease records. Adapted from Fichet-Calvet & Rogers (2009). © Fichet-Calvet, Rogers; 2009. Licence: CC BY. B: Predicted probability of Lassa virus presence based on a model correlating spatial environmental data with reported Lassa virus infections in humans and rodents. Adapted from Mylne *et al.* (2015). © Mylne *et al.*; 2015. Licence: CC BY. C: Predicted suitability for Lassa fever across space, based on a model correlating environmental and socioeconomic data with human disease records. Black points show human disease records. Adapted from Redding *et al.* (2016). © Redding et al.; 2016. Licence: CC BY. D: Predicted number of human Lassa virus infections per year, per km^2, based on a model correlating the predicted zoonotic niche of Lassa virus with human serosurvey data to predict spatial human seroprevalence, then using a compartmental disease model to estimate force of zoonotic infection from modelled human seroprevalence. Adapted from Basinski *et al.* (2021). © Basinski *et al.*; 2021. Licence: CC BY.

Redding *et al.* (2016) modelled the drivers of spillover risk, incorporating socioeconomic factors (such as ethnic groupings, poverty, and crop yields) in addition to environmental variables into a zoonotic niche model to control for more spatial determinants of Lassa fever risk (Figure 1.6C). Redding *et al.* also fit a SDM for *M. natalensis* by fitting occurrence records to environment and habitat variables (such as land-use). This SDM also used GBIF records

but, unlike Mylne *et al.* (2015), restricted the GBIF records used to the Western clade of *M. natalensis* to reduce phylogenetic differences in environmental and habitat requirements. By including a simple transmission model to predict relative number of spillover events from reservoir host occurrence and human population size, Redding *et al.* used their model to predict the impacts of future change (such as changes in land-use and human population size) on spillover risk. This model predicted that by 2070 there would be approximately twice as many (estimated range 2.0 to 2.1) Lassa virus spillover events as 2016, motivating further research into the effects of anthropogenic change on Lassa fever risk.

Basinski *et al.* (2021) (Figure 1.6D) took the zoonotic niche framework a step further by using a zoonotic niche approach to estimate spatial human seroprevalence based on human serosurveys. By incorporating a compartmental model of infection to quantify the number of infections based on modelled spatial serology, they estimated an annual 898,000 million (symptomatic and asymptomatic) Lassa virus infections across West Africa. The analysis of Basinski *et al.* represents a major step forward in Lassa fever modelling since it estimates absolute number of infections rather than relative risk, which is important for understanding disease burden and appropriately allocating resources. However, any estimates based on serology are limited by our knowledge of duration of seropositivity before seroreversion and limitations in the serosurveys themselves. Regarding the first point, Basinski *et al.* estimated that introducing a period of 15.6 years before seroreversion (rather than their base assumption that Lassa virus infected individuals remain seropositive for life) would increase the estimated number of infections by a factor of five. However, Basinski *et al.* did not incorporate the changing population size of West African countries into their model, which may cause an underestimate of infection incidence. The model of Basinski *et al.* also did not produce uncertainty in incidence estimates, making it unclear to what extent uncertainty in their model of spatial human seroprevalence could impact their results.

Several temporal mathematical transmission models have recently been fit to Lassa fever case data (Musa *et al.* 2020; Barua *et al.* 2021; Ibrahim & Dénes 2021; Abdulhamid *et al.* 2022; Abidemi *et al.* 2022), however these have limited predictive utility since they are non-spatial, and sometimes neglect the structure of the disease transmission system. None of these mathematical transmission models have explicitly incorporated variable rodent population dynamics informed by ecological studies. Theoretical rodent population dynamics based on climatic variables into a mathematical transmission model have shown how precipitation

could be linked with Lassa fever cases via rodent population dynamics or migration (Akhmetzhanov *et al.* 2019), however these modelled population dynamics were not fit to observed data. The absence of modelling approaches centred on reservoir host ecology and processes has led to a lack of understanding around the relationship between disease risk and the environment, hindering the development of process-based forecasting models especially with respect to environmental changes.

1.8 Thesis overview

Spatiotemporal modelling of the Lassa virus system has, as highlighted above, been historically limited. Therefore, in this thesis I seek to understand the ways in which environment – mediated by reservoir host ecology – underpins human disease risk, through methodological developments designed to make use of available data sources. Even though seasonality is central to the reservoir host's population dynamics, the link between inter-annual patterns in disease cases and the environment was previously unknown. Led by Dave Redding and Rory Gibb (Redding *et al.* 2021), I and colleagues showed that the timing of Lassa fever cases in Nigeria reported by the NCDC is linked with precipitation and vegetation (Supplementary Figure S1.1) using a spatiotemporal framework fitted to lagged environmental variables. Our findings were consistent with the assertion that the population ecology of *M. natalensis* drives Lassa virus spillover.

At the conception stages of Redding *et al.* (2021) I created a theoretical model of seasonal reservoir host population dynamics and resulting pathogen and spillover dynamics. This motivated us to include lag times in the spatiotemporal framework, representing the time taken for the reservoir host population size to increase following ideal climatic conditions. Acknowledging potential spatial biases in surveillance, I explored whether severity bias was present in these data. This analysis informed the decision to control for year-on-year expansion of surveillance effort in the spatiotemporal model in Redding *et al.* (2021) and to allow for spatially heterogeneous surveillance.

Leading on from this work, in Chapter 2, I investigate the severity bias present in Lassa fever case data in Nigeria for 2018 to 2021, an extended form of the analysis which informed consideration of surveillance regimes in Redding *et al.* (2021). Finding a severity bias is detectable throughout these years, I conclude that any disease incidence estimates of Lassa fever based on case data alone would represent a substantial underestimation. I therefore explore alternative sources of data to inform on disease and infection incidence, choosing to

use human seroprevalence and contact tracing data in Nigeria. I then design models to assess both the infection incidence (by updating the method of Basinski *et al.* (2021)) and disease incidence (using contact tracing data in Nigeria) and to understand how different assumptions of spatial zoonotic hazard and exposure impacted assessments of overall disease incidence across West Africa. My findings demonstrate how uncertainty in spatial disease risk structure and period before seroreversion substantially impacts incidence estimates, but nonetheless indicates that Lassa virus disease incidence and infection incidence are likely much greater than previously estimated. This result provides renewed motivation to better understand the processes underpinning zoonotic disease risk and to quantify these processes in such a way that forecasting and prediction of the effects of environmental change can be carried out.

In Chapter 3, motivated by my theoretical models of *M. natalensis* population dynamics and time lags, I quantify the climatic drivers of the reservoir host by analysing the way seasonal climate patterns and inter-annual variability in these patterns impact the observable demographic processes in a population of *M. natalensis* in Morogoro, Tanzania. Since this population is outside of the Lassa virus endemic area, a CMR study is feasible, enabling observations such as whether an individual becomes pregnant between two captures. Informed by the theoretical population models I developed in the early stages of Redding *et al.* (2021), I fit climate lags to allow for ecological delays between climate conditions and population responses. I find that the demographic processes of survival, recruitment, and body growth are highly seasonal, but that there is also a significant effect of inter-annual variability in precipitation on survival and recruitment. I then construct a demographic population model based on these processes, outputting a distribution of weights in the population at each given time based on lagged seasonal temperature and precipitation and inter-annual variability in precipitation. Finding that this model is a good fit for population size estimated directly from capture data, I conclude that the climatic-demographic model approach is a useful tool through which to predict population changes based on climate patterns.

Seeking to extend the work of Chapter 3 to include more aspects of disease ecology, in Chapter 4 I incorporate a transmission model within the reservoir host population into the climatic-demographic model framework. This transmission model is theoretical, but its structure and parameters are informed by observations of arenaviruses in *M. natalensis*

populations. Since, in Chapter 3, I modelled the population in terms of demographic processes – thereby capturing nonlinearities in the population dynamics – rather than simply fitting population size to environmental variables, the climatic-demographic model provided an opportunity to explore population changes under differing climate patterns. I therefore apply the full model including transmission to the five states in Nigeria with the highest number of cases during the period 2018 to 2021 to assess the resulting model's ability to predict zoonotic hazard. Taking into account the delay between infection and symptom onset by estimating an incubation period for Lassa fever using nosocomial outbreak data, I infer Lassa virus infection dynamics from the observed cases of Lassa fever. I find that predicted zoonotic hazard is significantly positively correlated with inferred disease incidence, indicating that climatically-driven reservoir host ecology is consistent with apparent variability in Lassa fever cases and that both seasonality and inter-annual variability in precipitation patterns could drive temporal Lassa fever risk. The resulting temporal reservoir host-virus model, which can be applied to differing climate patterns, presents an opportunity to forecast temporal changes in Lassa virus zoonotic hazard and to understand how future changes to climate patterns may affect disease risk. However, my observation that the model of zoonotic hazard is less well aligned with timing of Lassa virus infection in southern states of Nigeria may indicate a role of additional factors such as agricultural practices or behaviour leading rodents to come into increased contact with humans.

In Chapter 5 I synthesise the findings across these three pieces of work and put them into the wider context of Lassa fever research. I explore the implications of the thesis, including how temporal patterns in Lassa fever risk may vary across space, which is of increasing relevance given the anticipated changes in weather patterns across West Africa under climate change. Finally, I discuss open research questions, my perspectives on the future directions of zoonotic disease research and how this complex and interdisciplinary topic can continue to evolve in the future.

Chapter 2: Mathematical modelling enables new estimates of Lassa virus infection and disease incidence

2.1 Abstract

Understanding incidence of zoonotic infection and disease is key to allocating resources to improve public health. However, for most zoonoses, infection and disease incidences are not well-quantified, with a major challenge being the underreporting of human cases in public health data. Here, I address this challenge by developing incidence models based around two different data types, thereby increasing the robustness of incidence estimation. I use a case study of the neglected zoonotic disease Lassa fever, where previous Lassa virus infection incidence estimates are uncertain and range from 100,000 to 900,000. First, I test whether publicly available records of Lassa fever cases in Nigeria suffer from a severity bias. If a bias is present, this implies that underreporting in some states is substantial and that alternative approaches are needed. Finding evidence of a severity bias throughout the years 2018 to 2021, I then identify two alternative sources of data to inform incidence: serosurveys, and contact tracing. Building on an existing model informed by human seroprevalence but importantly accounting for the increasing population size of West Africa, I estimate an annual 3 million infections across West Africa if Lassa virus antibodies persist and protect for life. This number of infections increases to 7 million annually if a period before seroreversion of 15.6 years is assumed as has previously been indicated. I then develop a model for Lassa fever disease incidence based upon contact tracing data in Nigeria by accounting for the data collection procedure and different models of spatial zoonotic hazard and exposure, and estimate an upper bound for annual disease incidence across West Africa of between 4.5 and 32 million. The wide range in estimates results from sensitivity to the modelled spatial distribution of zoonotic hazard and exposure, with uncertainty arising from differing plausible models, thereby highlighting the limited degree to which the spatial determinants of zoonotic hazard and exposure are understood. Overall, my analysis implies that incidence of Lassa virus infection is much greater than previously thought, and that exploration of knowledge gaps may provide additional clarity around both infection incidence and disease incidence. The modelling frameworks developed here can be applied to similar data types for other zoonotic diseases for which incidence is unknown or uncertain.

2.2 Introduction

Zoonotic diseases represent the greatest global risk of epidemics (Morse et al., 2012), both through new diseases emerging from wildlife and through existing endemic zoonoses which may evolve to increase their potential for human-to-human transmission. This epidemic potential is driven by, among other factors, the number of times the pathogen is transmitted, either from the zoonotic host to humans – known as spillover – or between humans, with each transmission event providing evolutionary pressure for adaptation to humans. Zoonotic diseases also cause substantial disease burden (Grace *et al.* 2012; Torgerson *et al.* 2018) which is fundamentally driven by incidence. Understanding disease and infection incidence is therefore key to allocating resources to mitigate risks posed by a zoonotic disease, such as healthcare resources to reduce disease burden and interventions to reduce transmission opportunities, yet the quantity of transmission events is unknown or substantially underestimated for most zoonotic diseases (Maudlin *et al.* 2009).

Neglected zoonotic diseases, such as rabies (Wunner & Briggs 2010), bovine tuberculosis (Mableson *et al.* 2014), and cystic echinococcosis (Larrieu *et al.* 2019) are allocated few resources to manage and mitigate human disease burden and epidemic risk (WHO & DFID-AHP 2005; Welburn *et al.* 2015), making accurate estimates of spatiotemporal zoonotic hazard and exposure especially valuable. Such models can highlight hotspots or predict future changes for particular diseases (Redding *et al.* 2016, 2019) and potentially compare risks from different diseases. However, existing models for the incidence of infection and disease for neglected zoonoses typically relate to relative, not absolute, risk of these events occurring. Limited resources for diagnostic testing, public health surveillance, and research result in substantially underreported human case numbers (WHO & DFID-AHP 2005; Grace *et al.* 2012) and limited insight into disease incidence and transmission dynamics. With neglected zoonoses being most prevalent in areas in which the greatest anthropogenic change is predicted to occur (Popp *et al.* 2017) – change that will likely increase zoonotic spillover rates (Glidden *et al.* 2021) and thereby increase global risk of zoonotic emergence (Jones *et al.* 2008; Allen *et al.* 2017) – it is vital to address this challenge now.

As a high-incidence and high-burden neglected zoonosis, Lassa fever is an example of a disease for which better incidence estimates are needed but insufficient data are available to construct an accurate and complete spillover pathway (Scoones *et al.* 2017). Little is known about how often people are in contact with the principal host *Mastomys natalensis* or shed

Lassa virus, what proportion of zoonotic hosts are infected at any given time, or the likelihood of zoonotic spillover following human exposure to the pathogen (Gibb *et al.* 2017). Additionally, public health case records are believed to suffer from under-ascertainment, precluding top-down incidence estimation (Arruda *et al.* 2021). While a recent approach enabled the incorporation of surveillance bias into a model of Lassa fever cases (Redding *et al.* 2021), this was nevertheless unable to assess the scale of overall case under-ascertainment. However, advances in reservoir host abundance estimation (Gibb 2020), increased spatial representation in human serosurveys (e.g., Sogoba *et al.* 2016; Nimo-Paintsil *et al.* 2019), and new initiatives for contact tracing (NCDC 2018) create an opportunity for an approach blending public health data with host-pathogen ecology.

Until recently, Lassa virus infection incidence was estimated at between 100,000 and 300,000 infections annually across West Africa (McCormick *et al.* 1987). However, these figures were based on spatially limited serosurveys. These serosurveys could also be considered outdated since the population size of West Africa has more than doubled since the study. An updated analysis (Basinski *et al.* 2021) combined *M. natalensis* occurrence data and Lassa virus occurrence in *M. natalensis* using a niche modelling approach to create a spatial model of zoonotic hazard. Spatial zoonotic hazard was then correlated with observations from human serosurveys to produce a spatiotemporal model of human seroprevalence which, in turn, was used to predict per-capita force of infection for humans across West Africa, estimating at least 897,000 human infections per year. This approach represents an important step in updating the discourse around Lassa fever which is centred on estimates which predate a dramatic expansion in the population of West Africa. However, the estimate of Basinski *et al.* has some limitations. Firstly, this estimate assumed that the population size of countries in West Africa was temporally constant, which could cause underestimates of incidence – since an increasing population size, with more births than deaths, will present lower seroprevalence than one with equal births and deaths. Secondly, the duration of seropositivity is unknown. Basinski *et al.* estimated that allowing for seroreversion could greatly increase the incidence estimate, but it would be valuable to perform analysis on sources of data relating to human infection other than seroprevalence. Finally, the zoonotic niche approach to estimating spatial zoonotic hazard is subject to uncertainty which is challenging to measure. Since there are a variety of approaches to take when estimating

zoonotic hazard and exposure (Chapter 1), use of different models could provide a sense for the uncertainty in estimated incidence.

Here, I develop a variety of models to estimate Lassa virus infection and disease incidence, leveraging multiple approaches to assess critical biases and knowledge gaps. Firstly, I examine the observed fatality ratio in Lassa fever case records kept by the Nigeria Centre for Disease Control (NCDC), which has advanced Lassa fever detection through the implementation of improved testing capacity (NCDC 2018; Redding *et al.* 2021). By testing for a severity bias, I explore whether these case records are nevertheless spatially heterogeneous with different states achieving different levels of surveillance. Secondly, I build upon the work of Basinski *et al.* (2021) by introducing variable annual birth and death rates into a human seroprevalence model to estimate force of infection and infection incidence across West Africa. Thirdly and finally, I estimate annual probability of disease for individuals in contact tracing in Nigeria and project this estimated annual disease probability across West Africa, using spatial models of zoonotic hazard and exposure to account for the increased risk of disease experienced by the contact tracing population. To do so, I utilise a variety of plausible models of spatial zoonotic hazard and exposure, each based on different assumptions of spatial determinants of human exposure to Lassa virus, resulting in a representation of the uncertainty in disease incidence estimates. Through these approaches, I aim to assess how current knowledge gaps limit estimation of total incidence and spatial distribution of Lassa fever, and to provide an updated and nuanced assessment of the total incidence of Lassa fever.

2.3 Severity bias in case records in Nigeria

2.3.1 Methods

Case reports

I obtained case reports of Lassa fever in Nigeria from the NCDC situation reports (NCDC 2022). From these I extracted annual confirmed cases and deaths in confirmed cases for 2018–2021 inclusive by state. I restricted these data to states which recorded at least one confirmed case in the given year (Supplementary Data S2.1). These data were then restructured to record, for each individual confirmed case, (1) the total number of confirmed cases in that state in that year, and (2) whether the case was fatal or not as a binary outcome (1 for death and 0 otherwise).

Severity bias model

I modelled fatality F_i (whether or not confirmed case i was recorded as a death) as a random variable dependent on the number of confirmed cases, C_i, in the same state in the same year (equation 2.1). Observed case fatality probability $p_F(C_i)$ minus a modelled fatality ratio MFR then depended on C_i (equation 2.2).

$$F_i \sim \text{Bern}\big(p_F(C_i)\big) \tag{2.1}$$

$$p_F(C) = MFR + b\left(1 - \left(\frac{C-1}{C_{\max}-1}\right)^a\right) \tag{2.2}$$

The parameters b (the bias parameter) and a (the strength-of-bias parameter) determine how number of observed cases C deviates p_F from the model fatality ratio, $MFR \in (0,1)$ (Figure 2.1). If $b = 0$ then $p_F \equiv MFR$ (no bias) while if $b = (1 - MFR)$ then $p_F(1) = 1$ (if only one case is reported then the probability of fatality is guaranteed; extreme severity bias). Any values of $b > 0$ indicate severity bias since the derivative of p_F with respect to C is negative ($p_F'(C) < 0$) (fewer cases mean higher fatality probability), while values $b < 0$ indicate bias in the opposite direction with $p_F'(C) > 0$ (fewer cases mean lower fatality probability) (Figure 2.1A). The bias parameter was restricted to $b \in [-MFR, (1 - MFR)]$ to ensure that p_F was constrained in $[0, 1]$. Meanwhile, a moderates the strength of the relationship, with small values of a causing observed case fatality probability $p_F(C)$ to rapidly decrease as C increases, and large values of a causing $p_F(C)$ to decrease more gradually with increasing C (Figure 2.1B). Very small or very large values of a result in a model which, across much of the domain C, could fit well to unbiased or slightly biased data (dashed lines in Figure 2.1B) and would make it impossible to identify b. Therefore, the strength-of-bias parameter was restricted to $a \in [0.1, 10]$ to ensure identifiability of the bias parameter b. The robustness of this restriction was ensured by verifying that the values close to the bounds ($a = 0.11, 9.9$) did not appear within the 95% confidence region, and that changes to the restricted range of a did not alter estimated values or confidence intervals for a, b, or MFR. The MFR was designed to reflect the case fatality ratio for an observation in the state-year record with the highest number of confirmed cases ($C_{\max}$), since $p_F(C_{\max}) = MFR$. If a severity bias exists ($b > 0$), the MFR provides an upper bound on true disease fatality ratio, as true disease fatality ratio (DFR) may be even lower than the MFR; for instance, if the state-year record with the highest number of confirmed cases also suffers from underreporting and severity bias.

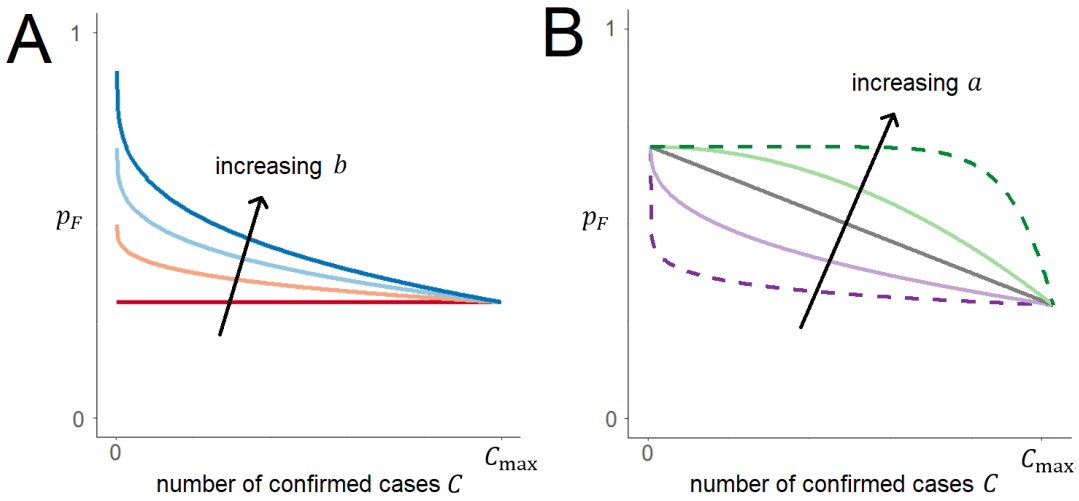


Figure 2.1: Structure of the severity bias model. A: Effect of varying the bias parameter b on the modelled relationship between CFR and number of confirmed cases. For all lines $a = 0.3$ and $MFR = 0.3$ are held constant while b takes the values 0 (red), 0.2, 0.4, and 0.6 (blue). B: Effect of varying the strength-of-bias parameter a on the modelled relationship between CFR and number of confirmed cases. $b = 0.4$ and $MFR = 0.3$ are held constant while a takes the values 0.1 (purple), 0.4, 1, 2, and 10 (green). Dashed lines ($a = 0.1$ and $a = 10$) show the permitted boundaries of a, since values outside of this interval would produce curves which – across much of the domain C – fit well to unbiased data, thereby making the bias parameter b non-identifiable.

Model fitting

This model was fit across all years of the data, and also separately by year to observe any increase or reduction in severity bias. Models were fit using maximum likelihood estimation, with likelihood ratio tests used to approximate distributions for the bias parameter b, strength-of-bias parameter a, and MFR. I tested for a severity bias ($b > 0$) by using likelihood ratio testing to perform a one-tailed hypothesis test (significance level of $\alpha = 0.05$ for all analyses). I then tested for differences in severity bias between years based on the estimated distributions of b using a two-tailed test; if b became smaller ($\Delta b < 0$) then severity bias decreases between years indicating improved surveillance, while $\Delta b > 0$ would indicate worsening surveillance.

I also compared observed CFR to MFR since, if a severity bias is present, this comparison gives an indication of how severity bias can deviate CFR from the true DFR; the presence of a severity bias alone does not necessarily imply that MFR is substantially different from CFR. I fit a simple binomial model wherein $p_F(C_i) = CFR$, where CFR is assumed constant across states and years. I used maximum likelihood and likelihood ratio testing to estimate CFR and

its 95% confidence interval, and to test whether this was significantly greater than the model fatality ratio, MFR, using a one-tailed hypothesis test. If $CFR > MFR$ then this implies that CFR is also greater than the true DFR, and that underreporting and severity bias are causing an inflation in apparent fatality of Lassa fever.

2.3.2 Results

I fit the model across all years (2018–21) and to each year individually (Figure 2.2). I identified a significant severity bias ($b > 0$) across the full data set containing fatality outcome for all Lassa fever cases in Nigeria between 2018 and 2021 ($n = 3165$, $p < 0.0001$) and for each individual year, 2018 ($n = 633$), 2019 ($n = 833$), 2020 ($n = 1189$), and 2021 ($n = 510$, $p < 0.0001$ for all) (Figure 2.3A, Supplementary Table S2.4). The estimate for the bias parameter across all years was 0.395 (95% CI [0.337, 0.455]). There was no significant increase ($\Delta b > 0$) or decrease ($\Delta b < 0$) in severity bias between any consecutive years (Figure 2.3B). The only pairs of years between which severity bias was observed to have changed was a significant decrease ($\Delta b < 0$) between years 2018 and 2020 ($p = 0.0093$) and 2018 and 2021 ($p = 0.0050$), indicating that surveillance may have improved in the last four years (Supplementary Table S2.5). Due to severity bias, the MFR was estimated to be lower than the CFR for all years both individually and combined ($p < 0.0001$ for all) (Figure 2.3C-D, Supplementary Table S2.6). While the CFR for the full case data was 21.8% $[20.6\%, 23.1\%]$, the MFR was 13.2% $[11.6\%, 14.8\%]$. The MFR for the year 2021 was estimated to be very close to 0 because the maximum number of observed cases in a state in 2021 was only 212, and only two states reported more than 50 confirmed cases of Lassa fever, thereby skewing the analysis. This may reflect a decrease in Lassa fever surveillance in 2021 due to a COVID-19 wave affecting Nigeria (Supplementary Figure S2.7) coinciding with the Lassa fever outbreak season.

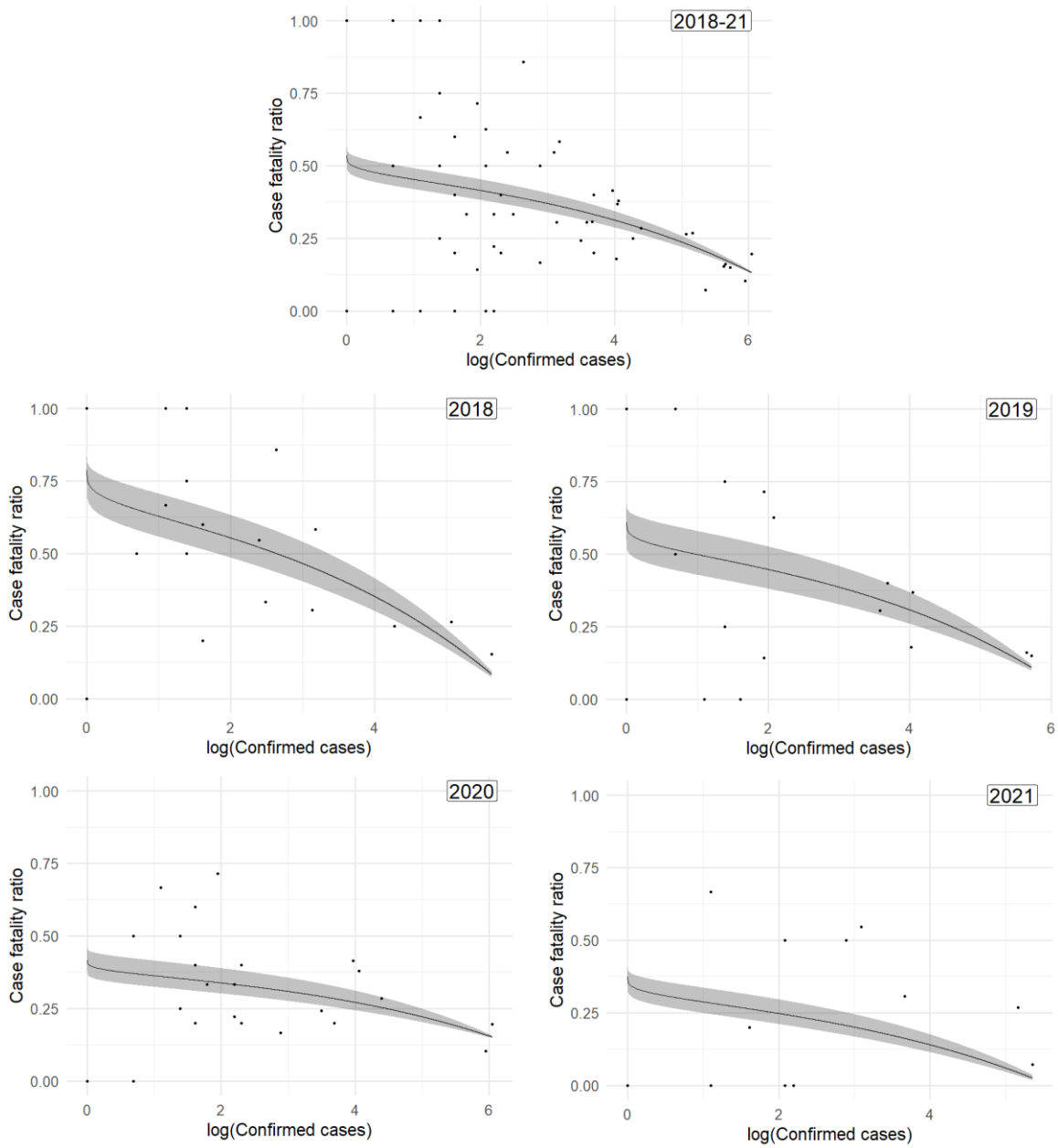


Figure 2.2: Severity bias model fit for the full data set (2018–21) and for each year individually. For each subset of data, the graph shows predicted CFR from the severity bias model and observed CFR from the case records. For ease of visualisation, the logarithm of confirmed cases C was used, since data are clustered around small values of C. Points show observed CFR against the number of confirmed cases for each state:year record. Lines and 95% confidence regions show the predicted relationship between CFR and the number of confirmed cases for the given data subset. The year(s) included are shown in the upper right corners of the graphs. Note that CFR is drawn from a binomial distribution with number of trials equal to the number of confirmed cases; therefore, observed CFR is expected to be more variable (and therefore deviated further from the predicted values) for smaller numbers of confirmed cases. For example, predicted CFR for a state:year record with only 1 confirmed case is between 0 and 1, while the observed CFR must be equal to either 0 or 1.

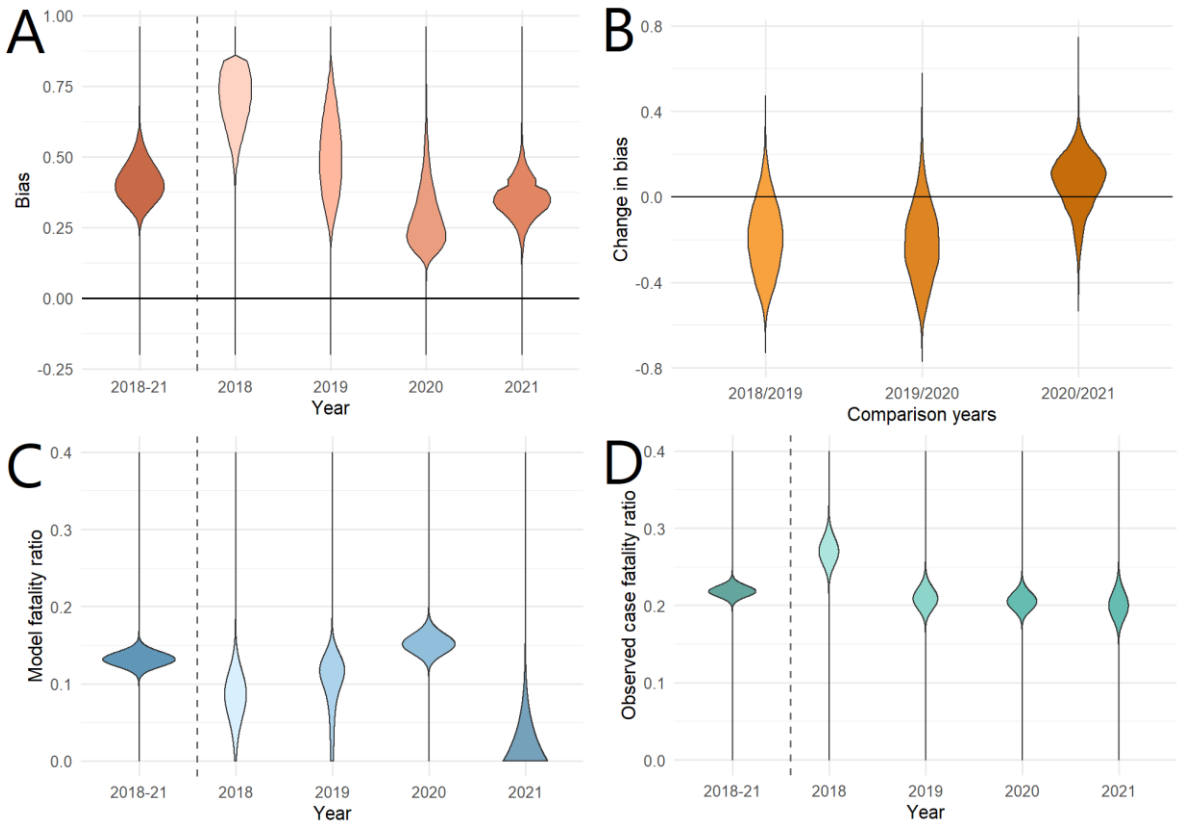


Figure 2.3: Bias in Lassa fever case fatality, changes in bias, and how bias affects observations of fatality ratio. A: Model distribution of bias parameter, b, for the model of all years and for the years 2018–2021 individually. Values less than 0 indicate that a severity bias is present in which observed case fatality ratio is greater in state:year records with fewer total confirmed cases. B: Model distribution of the difference between bias in consecutive years. Values of Δb less than 0 indicate that bias is less severe in the later comparison year. Changes between consecutive years were too small to detect with significance. C: Model distributions for model fatality ratio MFR after accounting for bias. D: Model distributions for estimated constant observed case fatality ratio CFR.

2.4 Severity bias in case records in Nigeria

2.4.1 Methods

Predicted spatial human seroprevalence

I used the spatial model of human Lassa virus seroprevalence created by Basinski *et al.* (2021), which was generated by correlating human serosurvey data with reservoir host and virus occurrence using a linear model. I treated this as a model of human seroprevalence in 2015, since this was the year of the latest serosurvey included in the model. In each 3 arcminute ($1/20$ degrees; $\approx$ 6km at the equator) grid cell, this model predicts the human seroprevalence of Lassa virus. To create this model, Basinski *et al.* used seroprevalence studies from 94 sites across different years spanning 1970–2015. However, the model is non-

temporal and predicts spatial human Lassa virus seroprevalence at a fixed point in time. Although it is possible that seroprevalence could have been variable across this large period, these temporal variations are ignored here for simplicity. Instead, I assume that this is a model of spatial human Lassa virus seroprevalence in 2015, the year of the latest serosurvey included in the model.

To estimate spatial patterns of zoonotic exposure in 2019–20, I obtained estimated global gridded population density in 2019 with a resolution of 30 arcseconds ($1/120$ degrees; $\approx$ 1km at the equator) from World Pop (WorldPop & CIESIN 2018). I aggregated the spatial global human population density model to the same scale as the spatial human seroprevalence model by summing the estimated population size in grid cells.

I restricted analysis to the West African countries Benin, Burkina Faso, Côte d'Ivoire, Ghana, Guinea, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, and Togo, based on believed spatial limitation of Lassa virus in the reservoir host, following Basinski *et al.* (2021). For each of these countries, to parameterise a per-country temporal model relating seroprevalence to infection, I obtained annual estimations of per capita birth rate, death rate, and population size for the years 1960 to 2019 inclusive from The World Bank world development indicators (The World Bank 2022a, b, c).

Modelling seroprevalence based on force of infection

I constructed a linear system of ordinary differential equations (ODEs) to describe movement of individuals between susceptible (seronegative) S, Lassa virus infected (seropositive) I, and recovered (seropositive) R compartments (equations 2.3–2.8, definitions in Table 2.1). Unlike an epidemic model, movement from the susceptible compartment to infected compartment does not depend on the infected population size since the majority of infection is assumed to be caused by spillover from the reservoir host. For simplicity, it was assumed that seropositivity indicated protective immunity. If this is not the case, then infection incidence estimates produced here are lower bounds. I also assumed that force of infection F was temporally (but not spatially) constant. Although force of infection likely varies within years due to seasonality of the zoonotic host, and may also vary between years, this assumption was made on the basis that the human seroprevalence model used is non-temporal, and that the parameter F reflects average force of infection across a year (averaged across seasonal reservoir host dynamics).

$$\frac{\mathrm{d}S}{\mathrm{d}t} = b_i(t)P - D(t)S - FS + \alpha R \quad (2.3)$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = -D(t)I + FS - \gamma I \quad (2.4)$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = -D(t)R + \gamma(1-\mu)I - \alpha R \quad (2.5)$$

$$P = S + I + R \quad (2.6)$$

$$D = d_i(t) - \frac{\gamma\mu I}{P} \quad (2.7)$$

$$Q = \frac{I + R}{P} \quad (2.8)$$

This ODE system takes the same form as the model of Basinski *et al.* (2021), but relaxing the assumption that birth and death rates are constant and equal (meaning a constant population size) and that the system was at steady state. As the population size of all countries in West Africa has been increasing throughout at least the last 60 years (Supplementary Figure S2.2), assumption of steady state would provide an underestimate of force of infection F. I set per-capita birth rate $b_i(t)$ and per-capita death rate $d_i(t)$ for each country i as the annual values from 1960 onwards (Table 2.1). I used per-country instead of per-grid cell estimates of birth and death rates and population size because the ODE model (equations 2.3–2.8) would need to be simulated for each spatial unit, therefore computing this across a grid would be computationally intensive.

Table 2.1: Parameters and variables in the seroprevalence model.

Parameter	Description	Value	Source
$b_i(t)$	Per-capita birth rate for country i	Per-capita birth rate (annual) by country	The World Bank (2022a)
$d_i(t)$	Per-capita (overall) death rate for country i	Per-capita death rate (annual) by country	The World Bank (2022b)
P	Total population size	Modelled (equation 2.6) and equal to annual population size by country	The World Bank (2022c)
$D(t)$	Per-capita non-infection-induced death rate	Modelled (equation 2.7)	
F	Force of zoonotic infection	Estimated (equation 2.10)	
γ	Rate of recovery	$(3 \text{ weeks})^{-1}$	WHO (2017)
α	Rate of seroreversion	0 (no seroreversion) and $(15.7 \text{ years})^{-1}$	Bond *et al.* (2013) and McCormick *et al.* (1987)
μ	Probability of infection-induced mortality	0.01	WHO (2017)
Q	Seroprevalence (proportion of the population which is seropositive)	Calculated (equation 2.8)	
Q^*	Seroprevalence in 2015 inferred from ODE system	Calculated (equation 2.8)	
q	Seroprevalence in 2015 inferred from spatial model	Spatially modelled	Basinski *et al.* (2021)
$G_k(F)$	Difference between ODE-predicted seroprevalence in 2015 and spatially-modelled seroprevalence in 2015	Calculated (equation 2.9)	

Estimating infection incidence from predicted seroprevalence

I created a numerical solver for equations 2.3–2.8 for each country for the period 1960 to 2019, using the forward Euler method (Butcher 2016) to estimate compartment sizes with a time step of 1/13 years (4 weeks). This time step was chosen as a balance between computational speed and accuracy, and I found that reducing the time step did not substantially affect the resulting numerical solutions. Letting the initial seroprevalence be Q_{1960}, the initial compartment sizes were $S = (1 - Q_{1960})P_{1960}$, $I = 0$, $R = Q_{1960}P_{1960}$, where P_{1960} is the initial population size (in 1960) for the country in question. This resulted in a numerically-generated function $Q^*(F \mid i, Q_{1960})$ for seroprevalence at the end of 2015, Q^*, depending on force of infection F, for a given country i (which determines b_i, d_i, and P_{1960}) and initial seroprevalence Q_{1960}.

Since I assumed that, in a given grid cell, the model of Basinski *et al.* (2021) predicted seroprevalence in 2015, this can be linked to seroprevalence Q^* obtained from the ODE model. This leads to the relationship $Q^*(F \mid i_k, Q_{1960}(k)) = q_k$, where k is the grid cell, i_k is the country grid cell k is inside (taken to be the country containing the centroid of the cell), and $Q_{1960}(k)$ is the initial seroprevalence in the grid cell. The final seroprevalence Q^* obtained from the ODE system was not sensitive to initial seroprevalence due to the long duration of simulation (Supplementary Figure S2.3) therefore I assumed that $Q_{1960}(k) = q_k$ so that initial seroprevalence would be in the same order of magnitude as anticipated final seroprevalence. This meant that the force of infection F_k in grid cell k satisfies

$$G_k(F) = Q^*(F \mid i_k, q_k) - q_k = 0. \tag{2.9}$$

Since Q^* was a numerical, not analytical, solution, it was not possible to obtain an exact value of F_k. Instead, I aimed to minimise $|G_k|$ (equations 2.9–2.10).

$$\begin{aligned} F_k &= F: \left\{ G_k(F) = \min_x G_k(x) \right\} \\ \text{i.e. } F_k &= \operatorname*{argmin}_x G_k(x) \end{aligned} \tag{2.10}$$

For each grid cell k for which current seroprevalence q_k was spatially modelled, I minimised $|G_k|$ using the one-dimensional optimisation function *optimize* in base *R* (R Core Team 2020). For each k, $G_k(F)$ only crossed 0 once across the domain $F \in (0,1)$ and the computed F_k resulted in a target function $|G_k(F_k)|$ within a tolerance of 10^{-6} of 0, meaning that the problem was well-posed and the optimisation was successful. As a result, I obtained the estimated annual force of infection F_k for every grid cell k.

Finally, I multiplied the spatial model of per-capita force of infection F by population size to estimate the total number of annual infections by country and across West Africa. For each grid cell k, I simulated the system of equations 2.3–2.8 using the forward Euler method, this time starting from the end of 2015 to the end of 2019. The initial conditions were $S = (1 - q_k)P_{k,2015}$, $I = 0$, and $R = q_k P_{k,2015}$, where $P_{k,2015}$ is the population size in 2015 in grid cell k. From the end of 2019 I then continued the simulation, but also predicting the number of infections over the subsequent year (the year 2020), with the addition of a quantity N, initially 0, counting the number of new infections (equation 2.11).

$$\frac{\mathrm{d}N}{\mathrm{d}t} = FS \tag{2.11}$$

The value of N after one year is then the annual incidence of Lassa virus infection in that grid cell in 2020, $I_I(\boldsymbol{x}_k)$.

For the purpose of comparison with existing approaches, I also computed the infection incidence under both scenarios of seroreversion with the assumption of constant population size (birth and death rates equal to birth rate in 2015 for estimating force of infection and 2019 for estimating incidence in 2020).

2.4.2 Results

My compartmental population model applied to Basinski *et al.*'s prediction of spatial human seroprevalence produced an estimate of 3.03 million infections in the year 2019 assuming no seroreversion takes place, and 6.97 million annual infections assuming seroreversion occurs after an average of 15.6 years (McCormick *et al.* 1987) (Figure 2.4). In both cases, most predicted infections were in Nigeria (1.75 million infections with no seroreversion and 4.07 million infections with seroreversion), and the highest per-capita incidence was predicted to be in Sierra Leone (0.0346 infections per-capita with no seroreversion and 0.0742 infections per-capita with seroreversion) (Table 2.2). These estimations do not include uncertainty since the spatial human seroprevalence model used did not include uncertainty.

Under the assumption that population sizes are stationary, total annual infection incidence was estimated as 712,000 with no seroreversion, and 4.80 million with seroreversion. Therefore, assuming a steady state leads to a substantial underestimate in infection incidence; of around 75% if it is assumed that seroreversion does not occur, or around 30% if seroreversion is assumed to occur.

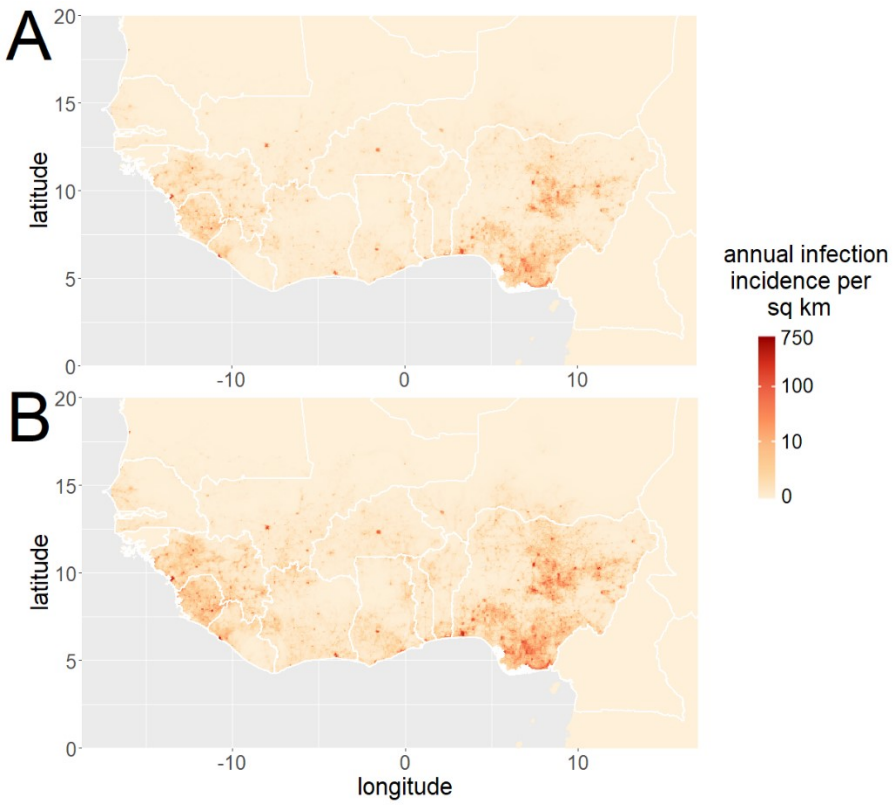


Figure 2.4: Spatial distribution of estimated Lassa virus infection incidence. Spatial density of infection incidence (number of infections per year per square kilometre) is shown on a scale from pale yellow (low incidence) to red (high incidence). The different spatial models represent different scenarios of seroreversion. A: Infection incidence under no seroreversion (Lassa virus antibodies remain detectable and protective for life). B: Infection incidence under seroreversion following an average period of 15.6 years, after which Lassa virus antibodies are no longer detectable or protective.

Table 2.2: Annual estimated infection incidence by country. Countries are listed in order of infection incidence (total and per-1000 people) given no seroreversion. Note that the total incidence ranking of Liberia and Ghana, and Niger and Benin, swap when seroreversion is included. Similarly, the per-capita incidence ranking of Togo and Ghana swap when seroreversion is included. All values are given to 3 significant figures. These estimates do not include uncertainty since the spatial human seroprevalence model used (Basinski *et al.* 2021) did not include uncertainty.

	Estimated annual total incidence of infection (1000s of infections)	
Country	No seroreversion	Seroreversion (≈15.6 years)
Nigeria	1,750	4,070
Guinea	335	742
Sierra Leone	222	475
Côte d'Ivoire	156	358
Liberia	114	244
Ghana	111	280
Mali	87.0	205
Burkina Faso	70.2	165
Niger	61.3	131
Benin	56.6	136
Senegal	30.8	73.5
Togo	29.7	72.0
Mauritania	7.93	19.8

	Estimated annual per-1000 person incidence of infection	
Country	No seroreversion	Seroreversion (≈15.6 years)
Sierra Leone	34.6	74.2
Guinea	28.3	62.7
Liberia	27.1	58.2
Nigeria	8.34	19.4
Côte d'Ivoire	6.04	13.9
Benin	4.53	10.8
Mali	3.93	9.25
Togo	3.52	8.54
Ghana	3.46	8.73
Burkina Faso	3.20	7.50
Senegal	2.66	6.34
Niger	2.66	5.70
Mauritania	2.03	5.07

2.5 Using contact tracing data in Nigeria to inform disease incidence across West Africa

2.5.1 Methods

Spatial zoonotic hazard and exposure models

Next, I explored how contact tracing data could be used to inform estimates of total Lassa fever disease incidence. However, an individual in contact tracing is likely to have a heightened risk of Lassa fever. Since such an individual has been identified as a contact of a Lassa fever case, they are at risk of Lassa virus infection from the same source as their contact, or from the contact themselves, and as a result are more likely to have Lassa fever than the average individual in West Africa. To account for the increased risk of zoonotic disease in contact tracing, I developed spatial models which would enable an estimate of the risk of disease in contact tracing relative to the rest of West Africa. Firstly, I modelled

zoonotic hazard. It is not clear whether zoonotic hazard in a particular place should relate to the probability of the reservoir host occurring there, or to the abundance of the reservoir host in that place. I therefore used two models of the reservoir host; one which modelled reservoir host occurrence, termed RH1 (Basinski *et al.* 2021), and one which modelled reservoir host abundance, termed RH2 (Gibb 2020). To model zoonotic hazard, I then combined each of these reservoir host models with a model of the spatial distribution of Lassa virus within the reservoir host. The first spatial model of Lassa virus, V1, was an occurrence model in which presence of the virus in the host was predicted from environmental correlates (Basinski *et al.* 2021), while the second spatial model of Lassa virus, V2, took a non-environmental approach, assuming instead that the distribution of the virus would depend on the spatial abundance of the reservoir host.

To estimate the spatial seroprevalence of Lassa virus in the rodent host (V2), I treated each 30 arc-second grid square as a node on a network and estimated the long-term connectivity of rodent populations in nearby and distant grid squares using network analysis. I obtained rodent serosurvey data from the dataset collated by Simons (2022). Grid squares which contained rodent serosurveys were "sampled nodes" and had a seroprevalence value from the serosurvey data. Nodes were connected to the grid squares adjacent to them by undirected weighted vertices. The weights of vertices were the assumed long-term contact rates between populations of rodents in the corresponding grid squares, with contact related to the abundance A (from RH2) in both squares. Therefore, grid squares i and j which were horizontally or vertically adjacent were connected by a path with weight $w_{i,j} = A(\boldsymbol{x}_i)A(\boldsymbol{x}_j)$, while if i and j were diagonally adjacent, the weight was $w_{i,j} = \left(A(\boldsymbol{x}_i)A(\boldsymbol{x}_j)\right)^{\sqrt{2}}$, to account for the distance along the diagonal. I then took the natural logarithm of these values to create transformed weights $y_{i,j} = -\log(w_{i,j})$ so that smaller values of y indicated more contact, and summing weights y along a path would be the equivalent of multiplying contact rates w. Using Dijkstra's algorithm (Dijkstra 1959) implemented in the *igraph* package (Nepusz 2022) for the programming language *R* (R Core Team 2020), I obtained the shortest path (minimising the sum of weights y) between every unsampled node i and sampled node j. Denoting the shortest distance between nodes i and j as $Y_{i,j}$, I estimated the connectedness of the two nodes as the exponent of this shortest distance. Finally, I

computed the estimated viral prevalence V in unsampled nodes by computing the average seroprevalence of sampled nodes weighted by connectedness:

$$V_i = \frac{\sum_{\text{sampled sites } j} V_j \exp\left(Y_{i,j}\right)}{\sum_{\text{sampled sites } j} \exp\left(Y_{i,j}\right)} \quad (2.12)$$

From the two spatial models of the reservoir host (occurrence, RH1; and abundance, RH2) and the two spatial models of Lassa virus (occurrence, V1; and prevalence, V2), I created three models of relative spatial zoonotic hazard (RH1V1, RH2V1, RH2V2) by multiplying the component reservoir host and virus models together. While other spatial models of zoonotic hazard similar to RH1V1 exist (Redding *et al.* 2016), the models of Basinski *et al.* (2021) were chosen so as to increase the comparability of the disease incidence estimated here with the infection incidence estimate described earlier. These infection incidence estimates (Section 2.3.2) were informed by the human seroprevalence model which was fit by Basinski *et al.* (2021) as a linear function of the zoonotic hazard model RH1V1. The relative zoonotic hazard in grid cell $\boldsymbol{x}_i$ estimated by a given model was denoted $H(\boldsymbol{x}_i)$.

Next, I created three spatial models for the scaling of human-rodent contact with human population size. The first contact scaling model (CS1) assumed that zoonotic exposure is directly proportional to human population size, therefore CS1 was equal to the population density P from the gridded human population density estimates (WorldPop & CIESIN 2018; described in Section 2.3.2). The second contact scaling model, following Gibb (2020), assumed that zoonotic exposure in a given location is proportional to the natural logarithm of human population size. In other words, as the number of people in a given grid square is increased there is a saturation effect in zoonotic exposure, assuming that zoonotic hazard is held constant. This logarithmic model is based on observations that contact between humans and rodents does not appear to increase in densely populated areas (Suwannarong & Chapman 2015). Finally, I created a third contact scaling model (CS3) which assumed that zoonotic exposure is proportional to human population size, but only in rural areas, since *M. natalensis* may not be prevalent in urban areas (Demby *et al.* 2001) – although this observation may be location-specific. Rather than describing human behaviour, this contact scaling model instead uses human population density to infer urban areas which are unsuitable for *M. natalensis*. Due to the spatial sparsity of *M. natalensis* data, this is a fine-scale factor which may affect zoonotic exposure but which has not been well-quantified; therefore, my inclusion of a rural exposure model enables an assessment of how this

assumption affects total disease incidence estimates. To distinguish between urban and rural areas, I obtained the total urban population size and total population size (urban and rural) for each country in West Africa in 2019 from The World Bank (The World Bank 2022c, d), estimating that 47.0% of the population lived in urban areas by dividing total urban population size by total population size in 2019 (≈180 million people in urban areas out of a total ≈390 million people). I then removed 47.0% of the population from the spatial population density P in order of decreasing population density, to produce a new model of rural population density. Nine spatial models of zoonotic exposure were created by multiplying each of the three contact scaling models with each of the three zoonotic hazard models (Table 2.3). For example, RH1V1-CS2 was generated by multiplying reservoir host occurrence (RH1) by Lassa virus occurrence (V1) and natural logarithm of population density, $\log(P+1)$ (CS2). For a given zoonotic exposure model, exposure was denoted $E(\boldsymbol{x}_i)$, where $\boldsymbol{x}_i$ is the grid cell.

Table 2.3: List of spatial models of zoonotic hazard and rodent-human contact used to estimate zoonotic exposure.

Name	Description	Source or method
RH1	Reservoir host occurrence	Basinski *et al.* (2021)
RH2	Reservoir host relative abundance	Gibb (2020)
V1	Virus occurrence	Basinski *et al.* (2021)
V2	Virus prevalence in reservoir host	Computed (described above) from synthesised trapping studies (Simons 2022) and RH2
RH1V1	Spatial model of zoonotic hazard, H	RH1×V1
RH2V1		RH2×V1
RH2V2		RH2×V2
CS1	Linear scaling of contact with human population size	Human population density P (WorldPop & CIESIN 2018)
CS2	Logarithmic scaling of contact with human population size	$\log(P+1)$
CS3	Linear scaling of contact with human population size in rural areas; no contact in urban areas	Computed (described above) by removing inferred urban population from P
RH1V1-CS1	Spatial model of zoonotic exposure, E	RH1V1×CS1
RH1V1-CS2		RH1V1×CS2
RH1V1-CS3		RH1V1×CS2
etc.		etc.

In this approach to estimating incidence of Lassa fever, I restricted analysis to the spatial extent from 21°W to 22°E and 2°N to 17°N to include the believed Lassa virus endemic region of West Africa. The resulting disease incidence estimates should be comparable to the Lassa virus infection incidence estimates since in non-overlapping areas between the two approaches, all models of zoonotic hazard and exposure were very low.

Contact tracing data

From the NCDC situation reports for Lassa fever in 2018 to 2021 (NCDC 2022), I obtained weekly contact tracing reports, detailing cumulative numbers of individuals since the start of the calendar year who had (1) entered contact tracing; (2) completed contact tracing; (3) been lost from contact tracing; (4) reported being symptomatic; and (5) reported being symptomatic and were then subsequently tested for and found to be positive with Lassa virus. Lassa virus infection in contact tracing is confirmed using polymerase chain reaction (PCR) testing. A contact is defined by the NCDC as "a person who has been exposed to an infected person, or to an infected person's secretions, excretions, or tissues within three weeks of last contact with a confirmed or probable case of Lassa fever" (NCDC 2018). A confirmed case refers to a laboratory-confirmed case and a probable case refers to someone who died from suspected Lassa fever but for whom no sample was obtained. The purpose of contact tracing is to improve the diagnosis of disease by performing laboratory testing if a fever is noted within the 21-day period, thus creating an enhanced surveillance system for these individuals. The contact tracing procedure also enables the administering of prophylaxis to reduce severity of potential disease in high-risk contacts.

Estimating probability of disease in the contact tracing population

I treated each individual completing contact tracing as an independent trial, so that the number of individuals who completed contact tracing having had Lassa fever symptoms and testing positive for Lassa virus followed a binomial distribution with "success" probability p_{CT} (equation 2.13). I denoted the number of individuals who completed contact tracing as $N_{\text{completed}}$, with N_{positive} of these individuals having developed Lassa fever symptoms and subsequently testing positive. I estimated a posterior distribution for p_{CT}, the per-capita probability of contracting Lassa fever during contact tracing, using the Metropolis-Hastings algorithm to implement the Markov chain Monte Carlo (MCMC) method with completely uninformative (uniform) prior distributions.

$$N_{\text{positive}} \sim \text{Bin}(N_{\text{completed}}, p_{CT}) \quad (2.13)$$

To obtain the annual per-capita probability of disease, p_A, I scaled up the number of infections as

$$p_A = 1 - (1 - p_{CT})^{\frac{365}{21}}, \quad (2.14)$$

since the contact tracing period was 21 days.

Modelling increased disease incidence in the contact tracing population

The probability of infection for an individual in contact tracing is expected to be different than for the average person in the population of West Africa since they are a contact of a Lassa fever case. The individual presumably lives in a Lassa virus endemic area and may be exposed to the same source of infection as their infected contact, so we would expect their zoonotic exposure to be higher than average. Also, they are at a greater risk of human-to-human infection than the general population because they are a known contact of an infected person, and their other contacts also presumably live in a Lassa virus endemic area. It is challenging to capture the elevated infection risk from clustering of contacts (e.g., an individual in contact tracing may live in the same home as their infected contact, who may have been infected in the home) or from human-to-human transmission. Therefore, I aimed to estimate an upper bound of disease incidence by taking into account elevated zoonotic hazard and exposure for an individual known to be in a Lassa virus endemic area, but not case clustering or human-to-human transmission from the infected contact. Since Lassa virus rarely leads to long human-to-human transmission chains, I will not distinguish between human and zoonotic sources of infection, and will assume that the likelihood of infection by a human is proportional to zoonotic exposure.

Let us assume a given model of zoonotic exposure E and population density P, and assume that infection incidence and disease incidence are proportional to this model of exposure E. Supposing that a Lassa virus infection has occurred in Nigeria (i.e., an individual has already been infected and now their contact is under surveillance in contact tracing), even with no information about location of infection we can compute the expected per-capita zoonotic exposure (zoonotic exposure E divided by population size P) in that location. The probability that an infection occurred in any given location $\boldsymbol{x}_i$ in Nigeria is $E(\boldsymbol{x}_i)$ divided by the total exposure across Nigeria. As a result, it is possible to calculate the expected per-capita zoonotic exposure experienced by an individual in the contact tracing population, by

assuming that the individual in contact tracing experiences the same per-capita zoonotic exposure as their infected contact. Using the assumption that disease incidence is proportional to zoonotic exposure, this expected per-capita zoonotic exposure in contact tracing (equations 2.15–2.16), denoted D^*, is related to per-capita zoonotic disease risk in the contact tracing population. By comparing per-capita zoonotic exposure across West Africa with this estimate of per-capita zoonotic exposure in the contact tracing population, it will therefore be possible to account for the elevated disease risk in contact tracing.

$$\begin{aligned} D^* &= \mathbb{E}\left(E/P \,\middle|\, \text{individual in contact tracing}\right) \\ &= \mathbb{E}\left(E/P \,\middle|\, \text{an infection already occurred}\right) \\ &= \sum_{i \text{ in Nigeria}} \frac{E(\boldsymbol{x}_i)}{P(\boldsymbol{x}_i)} \mathbb{P}(\text{infection occurred in } \boldsymbol{x}_i) \\ &= \sum_{i \text{ in Nigeria}} \frac{E(\boldsymbol{x}_i)}{P(\boldsymbol{x}_i)} \frac{E(\boldsymbol{x}_i)}{\sum_{j \text{ in Nigeria}} E(\boldsymbol{x}_j)} \qquad (2.15) \\ \Rightarrow D^* &= \sum_{i \text{ in Nigeria}} \frac{E(\boldsymbol{x}_i)^2}{P(\boldsymbol{x}_i)} \cdot \frac{1}{\sum_{j \text{ in Nigeria}} E(\boldsymbol{x}_j)} \qquad (2.16) \end{aligned}$$

Predicting Lassa fever incidence across West Africa

Consider a given grid cell $\boldsymbol{x}_i$ and let the per-capita zoonotic exposure be denoted $D(\boldsymbol{x}_i)$ (computed by dividing $E(\boldsymbol{x}_i)$ by $P(\boldsymbol{x}_i)$). We want to estimate the per-capita annual disease probability in this grid cell. Given that the estimated annual per-capita disease probability in contact tracing is p_A with a corresponding per-capita zoonotic exposure D^*, assuming that disease incidence is proportional to zoonotic exposure, it follows that the per-capita annual disease probability in grid cell $\boldsymbol{x}_i$ $p(\boldsymbol{x}_i)$ can be calculated as

$$p(\boldsymbol{x}_i) = \frac{p_A D(\boldsymbol{x}_i)}{D^*} = \frac{p_A E(\boldsymbol{x}_i)}{D^* P(\boldsymbol{x}_i)}. \qquad (2.17)$$

By multiplying this by the population size $P(\boldsymbol{x}_i)$ we can thereby obtain an annual number of symptomatic infections in grid cell $\boldsymbol{x}_i$ as

$$I_D(\boldsymbol{x}_i) = \frac{p_A E(\boldsymbol{x}_i)}{D^*}. \qquad (2.18)$$

Using equations 2.15–2.18 and the posterior distribution of p_A, I obtained a posterior distribution for the total disease incidence I_D given a zoonotic exposure model E. For each of the nine spatial zoonotic exposure models (Table 2.2) I therefore produced a spatial estimate of annual disease incidence $I_D(\boldsymbol{x}_i)$.

2.5.2 Results

The per-capita disease incidence probability in contact tracing (per 21 days) was $p_{CT} = 0.0059$ (95% CrI $[0.0052, 0.0067]$), which translates into an annual symptomatic infection probability of $p_A = 0.098$ $[0.087, 0.11]$ per-capita. The smallest estimate of disease incidence, 5.14 $[4.55, 5.79]$ million, was under the assumption that human risk of disease was proportional to reservoir host occurrence and Lassa virus occurrence, with only rural populations exposed (Figure 2.5–2.6, RH1V1-CS3). Meanwhile, the largest estimate of 28.5 $[25.3, 32.2]$ million was obtained when disease incidence was assumed proportional to abundance of the reservoir host, prevalence of Lassa virus within the reservoir host – estimated using weighted interpolation, and population size (Figure 2.5–2.6, RH2V2-CS1). Spatial and total estimates varied substantially among models (Figure 2.5–2.6) (Supplementary Table 2.8).

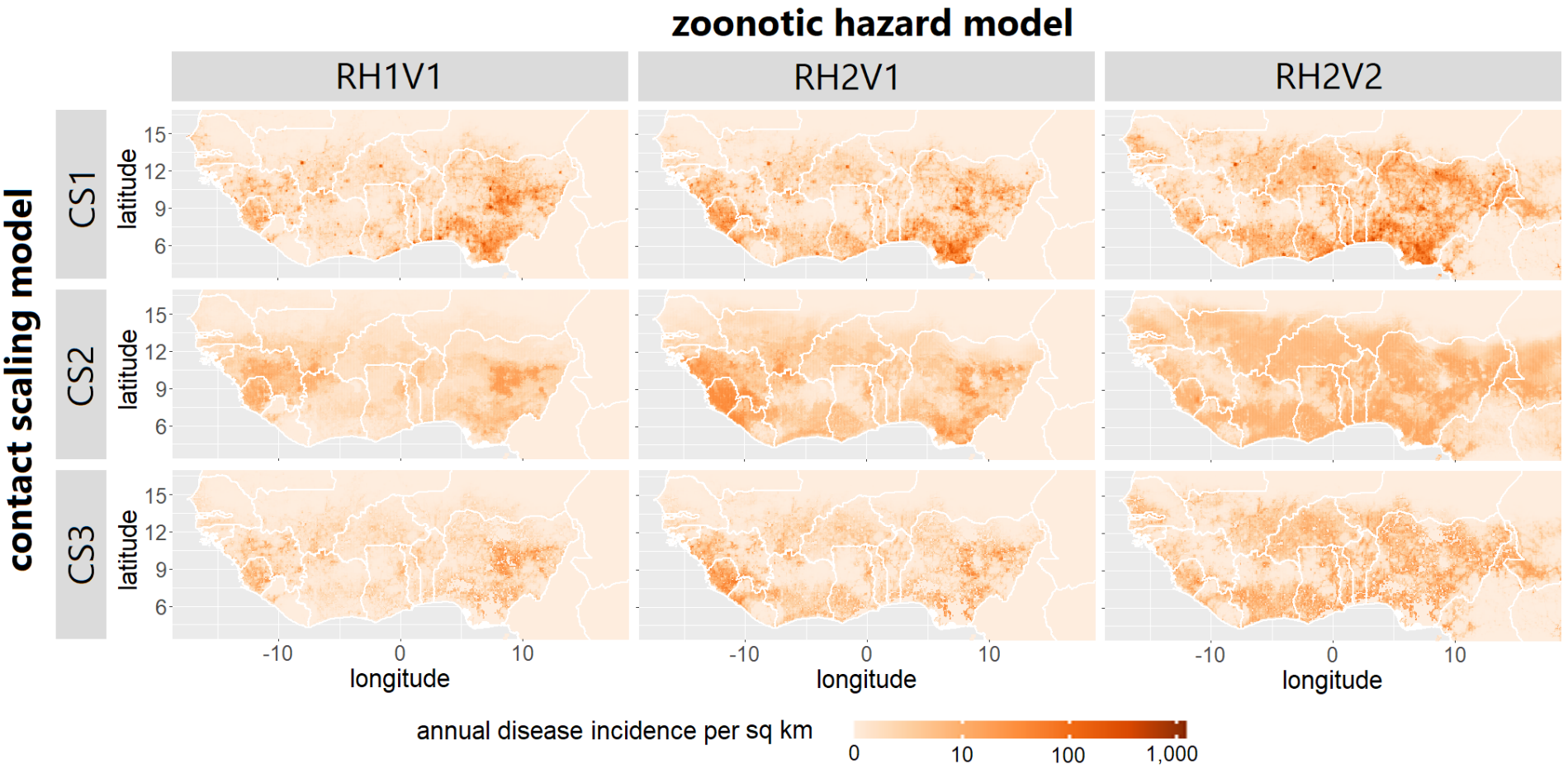


Figure 2.5: Spatial disease incidence estimates under different zoonotic hazard and exposure models. Left column (RH1V1): Zoonotic hazard is proportional to reservoir host occurrence and Lassa virus occurrence. Middle column (RH2V1): Zoonotic hazard is proportional to reservoir host abundance and Lassa virus occurrence. Right column (RH2V2): Zoonotic hazard is proportional to reservoir host abundance and Lassa virus prevalence in the reservoir host. Top row (CS1): Human-rodent contact scales linearly with human population size. Middle row (CS2): Human-rodent contact scales logarithmically with human population size. Bottom row (CS3): Human-rodent contact scales linearly with human population size but only occurs in rural populations.

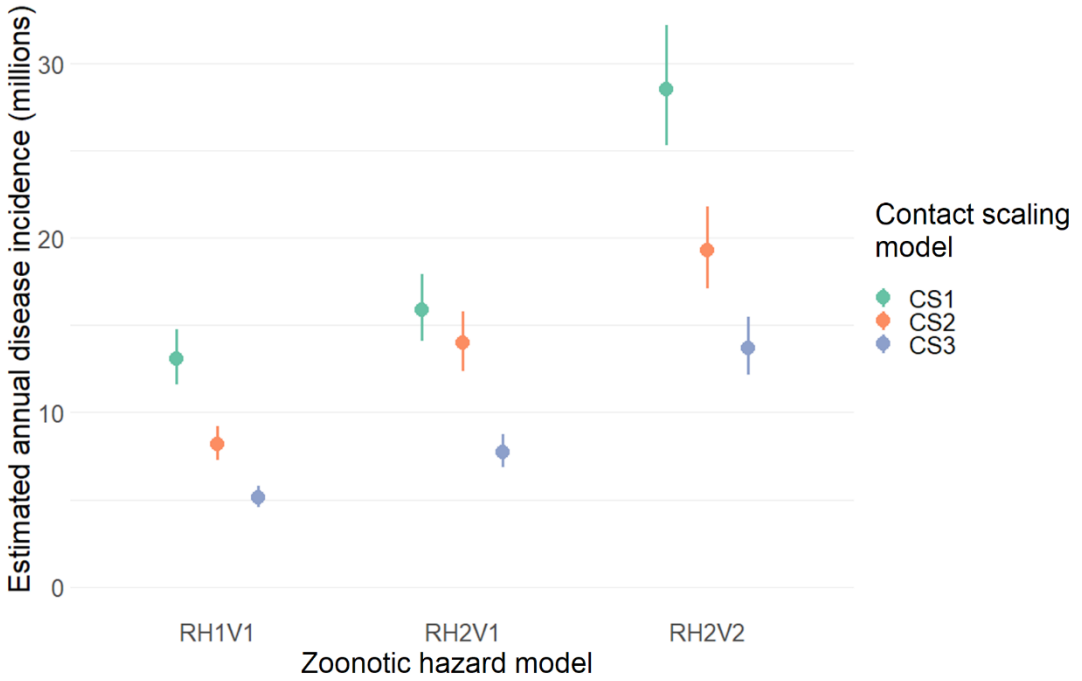


Figure 2.6: Disease incidence estimates under different zoonotic hazard and reservoir host-human contact models. Estimates of disease incidence are highly sensitive to different models of the reservoir host-virus distribution and how contact scales with human population size. Line ranges show the 95% credible intervals based on the posterior distribution for p_{CT}, with each line range showing estimated disease incidence under a different combination of models. Zoonotic hazard model RH1V1 (zoonotic hazard is proportional to occurrence of the reservoir host and occurrence of the virus) and contact scaling model CS3 (only rural human populations are affected) produce the smallest estimates, while models RH2V2 (zoonotic hazard is proportional to abundance of the reservoir host and prevalence of Lassa virus in the reservoir host) and CS1 (contact scales linearly with human population size) produce the largest estimates. RH2V1 (zoonotic hazard is proportional to abundance of the reservoir host and occurrence of Lassa virus) and CS2 (contact scales logarithmically with human population size) produce intermediate estimates.

2.6 Discussion

My analysis explored severity bias in passive surveillance data for Lassa fever and explored alternative routes by which to estimate absolute incidence of Lassa virus infection or disease, in doing so identifying sources of uncertainty and gaps in data and knowledge which prevent precise incidence estimates. I found that Lassa fever case records in Nigeria suffer from a severity bias which inflates the fatality ratio among observed cases in states which report fewer cases. While this severity bias is present throughout the period 2018 to 2021, there is evidence that the severity bias has decreased over this period, which may be

symptomatic of improved surveillance. I identified that the severity bias causes overall CFR in Nigeria to be almost twice the model fatality ratio which was designed to be more reflective of true disease fatality ratio. I then identified serosurveys and contact tracing as alternative sources of data which can inform estimates of infection incidence and disease incidence respectively. My estimate of annual infection incidence was 3.03 million given no seroconversion, or 6.97 million given a seropositive period of 15.6 years on average. This estimate is around twice that of the only other recent infection incidence estimate, due to accounting for the expanding population size across West Africa. Meanwhile, my estimates of annual disease incidence varied from 5.14 (95% CrI [4.55, 5.79]) million to 28.5 [25.3, 32.2] million, sensitive to both the scaling of rodent-human contact with human population density and the assumed spatial distribution of Lassa virus in the reservoir host. These disparate estimates highlight the limitations to scaling up Lassa fever risk across West Africa due to uncertainty in the spatial distribution of zoonotic hazard and exposure, and the lack of spatially representative data against which to validate estimates.

Taking the most conservative view of these new results, I estimate that there are between 600,000 and 4.5 million symptomatic infections per year, suggesting 30,000–225,000 Lassa fever fatalities per year assuming a total (asymptomatic and symptomatic) infection fatality ratio of 1% (WHO 2017). This compares with the 5,000 deaths per year in the only incidence estimate obtained prior to 2021 (McCormick *et al.* 1987). However, heterogeneity in surveillance – both within countries and more broadly across West Africa – make it impossible to be certain that incidence is not much greater than this range. Given these much greater estimates of disease incidence, implying a larger disease burden than previously thought, the development of forecasting models and the elucidation of ecological processes underpinning zoonotic disease risk can now be considered even more important.

My identification of a severity bias lends additional support to the argument that surveillance of Lassa fever is heterogeneous in Nigeria, which may be due to socioeconomic factors such as access to healthcare. Compared with urban areas, access to healthcare is poorer in rural populations (Appendix A), which are also more likely to be exposed to Lassa virus (Bonwitt *et al.* 2017). Additionally, individual likelihood of pregnancy is also greater, and access to prenatal care poorer, in rural locations (Appendix A), with pregnancy a risk factor which increases risk of fatality in Lassa fever substantially (Kayem *et al.* 2020). Given the co-occurrence in rural locations of these multiple risk factors for under-surveillance, expanded

surveillance or improved healthcare provisioning in these areas may greatly increase observation of current Lassa fever disease burden. Improvement to healthcare access in rural locations has been identified as a high importance activity for Nigeria (Riman & Akpan 2012). If implemented, a secondary benefit of these improvements to access will be an increase in passive observations of Lassa fever and associated disease burden.

By modelling the changing population sizes of West African countries since 1960, I estimated a greater incidence of infection than Basinski *et al.* (2021) when using their estimated human seroprevalence model. For a model in which birth rate is greater than death rate – as has been the case for all countries in the focal area since 1960 – a larger force of infection is needed to attain the same seroprevalence when compared with a model where birth and death rate are equal. This is because the susceptible population is increased at a relatively greater rate than the infected and recovered compartments due to the influx of births into the susceptible compartment. Under the no-seroreversion model, I estimated 3.03 million infections annually, more than three times the 897,000 estimated by Basinski *et al.* using the constant population size model. Under the assumption of seroreversion after a mean period of 15.6 years, I estimated 6.97 million infections, almost twice Basinski *et al.*'s 4.38 million. Assuming 80% of infections are asymptomatic (WHO 2017), this implies approximately 600,000 (assuming no seroreversion) and 1.4 million (assuming seroreversion after an average of 15.6 years) instances of disease annually. However, these values should be considered lower bounds since seropositivity does not necessarily imply protection against reinfection (Bond *et al.* 2013).

The human seroprevalence model used to estimate infection incidence was based upon the spatial model created by Basinski *et al.* which assumed that zoonotic hazard was proportional to reservoir host occurrence and virus occurrence (RH1V1), while different zoonotic hazard models would have led to different human seroprevalence estimates used for infection incidence. Since different assumptions of zoonotic hazard and exposure led to very large differences in estimated disease incidence, it would be valuable to assess the effect of different spatial zoonotic hazard and exposure models on the estimated human seroprevalence model and resulting infection incidence estimates. Furthermore, the human seroprevalence model did not include uncertainty, meaning that estimates of infection incidence produced here could not incorporate uncertainty in the seroprevalence model. In the future, uncertainty in the model of Basinki *et al.* to correlate human serosurvey data with

modelled zoonotic hazard could be used to produce further spatial models of human seroprevalence by sampling from the model distribution of the estimated correlation between zoonotic hazard and observed human seroprevalence. By performing the analysis carried out here on a large number of these samples, uncertainty in the spatial human seroprevalence model could be propagated into disease incidence estimates.

When estimating disease incidence, I found that the per-capita annual probability of disease in contact tracing was $p_A = 0.098\ [0.087, 0.11]$, but this reflected a population which is at increased risk of disease since individuals in contact tracing are a contact of someone who has been infected. I used models of the spatial distribution of zoonotic hazard and exposure to relate this observed risk in contact tracing in Nigeria to the total risk across West Africa, with a wide range of estimated disease incidence across the nine zoonotic exposure models used. These disparities reflect the ways in which these models estimate that risk is clustered; for instance, the zoonotic exposure model with the greatest estimated disease incidence assumed that risk was proportional to virus prevalence in the reservoir host (V2), estimated by weighted interpolation of rodent serosurveys, rather than estimated occurrence of Lassa virus (V1), estimated by environmental niche modelling. Because V2 estimates that virus prevalence is relatively uniform across West Africa, while V1 predicts that Nigeria and the Mano River Union are hotpots, the latter resulted in a much lesser incidence because the resulting zoonotic hazard model predicted more spatial clustering of disease incidence. Reducing the susceptible population to only those in rural areas unsurprisingly greatly reduced total estimated disease incidence.

While connecting health data with spatial models of zoonotic hazard has here enabled absolute infection and disease incidence estimation, this blended approach is particularly sensitive to the assumed spatial structure of risk because the spatial model is used both to estimate relative risk at health data points and to project absolute risk across the rest of the space. Disparities in estimates across models highlight the existing knowledge gaps around spatial distribution of disease risk and how assumptions influence our understanding of overall incidence. Moving forwards, if the goal is the improvement of incidence estimation precision, longitudinal seroprevalence studies in stratified samples might be helpful; however, this is challenging in endemic communities since observations of infection events and seroreversion can be muddied by reinfection (Bond *et al.* 2013). Combining health surveys with longitudinal studies in a similar model to the contact tracing procedure could

help identify instances of disease as opposed to asymptomatic infection. The models of infection and disease incidence used here should be updated as different spatial maps are developed to, hopefully, improve precision of estimates as knowledge of spatial Lassa fever risk accumulates. For many zoonoses, incidence of infection and disease are thought to be underestimated (Grace *et al.* 2012; Halliday *et al.* 2015) despite their value for resource prioritisation (Mathers *et al.* 2013), therefore methods to combine spatial ecological approaches with human health records such as presented here may be valuable in improving incidence estimates, and consequently health outcomes, for other neglected zoonotic diseases.

Chapter 3: Population dynamics of *Mastomys natalensis* are driven nonlinearly by seasonality and inter-annual precipitation variability

3.1 Abstract

The processes driving human exposure to zoonotic pathogens are still not fully understood, hindering development of accurate zoonotic disease models and forecasting systems. While zoonotic hosts often exhibit climatically-sensitive population dynamics which underpin transmission of the pathogen between hosts, these dynamics are typically understudied. This gap is of particular concern when contemplating future changes to zoonotic host dynamics due to climate change, leaving the effect of climate change on zoonotic diseases unknown. Here I tailor a demographic population modelling method to *Mastomys natalensis,* a host of a number of pathogens across its range, one of which is known to be zoonotic, to explore the effects of seasonal and inter-annual climatic changes on its demographic processes, analysing data from a long-term capture-mark-recapture study in Tanzania. I show that *M. natalensis* exhibits nonlinear population dynamics and that inter-annual variability of precipitation seasonality is a significant driver of recruitment. This finding provides evidence that observed population increases in this species following rainfall are likely due to increased resources for individuals to invest in pregnancy due to vegetation growth. Additionally, analysis of demographic processes confirm that *M. natalensis* population dynamics can change year-on-year and that this may be due to changing weather patterns. I then simulate the population dynamics of *M. natalensis* and show that the model derived from demographic processes is a good predictor of population size by comparing with non-demographic approaches. This process-based model with climatic inputs can be used to forecast both short- and long-term changes to *M. natalensis* population dynamics which ultimately underpin pathogen transmission dynamics. This development also facilitates the incorporation of ecological reservoir host processes into models of wildlife disease dynamics and zoonotic hazard.

3.2 Introduction

While the understanding, forecasting, and control of zoonoses is critical, for many of these diseases the processes leading to human infection are not understood sufficiently well to tackle these challenges (Plowright *et al.* 2017). Although the transmission of zoonotic pathogens takes place within complex ecosystems (Cunningham 2005; Engering *et al.* 2013; Plowright *et al.* 2017; Gibb *et al.* 2020a), such complexities are rarely considered when exploring drivers of zoonotic spillover or disease. Host ecology is often missing from the picture, despite offering potentially valuable insights (Karesh *et al.* 2012; Gibb *et al.* 2020a). Drivers of zoonotic hazard are likely to respond idiosyncratically to global environmental change, making it increasingly important to incorporate ecological complexity into models of zoonotic disease systems.

Climate change has had and continues to have a dramatic effect on biodiversity and ecosystem function (Portner *et al.* 2021). Zoonotic host-pathogen interactions, as with all other ecological relationships, will be impacted by environmental change in a variety of ways (Karesh *et al.* 2012; Glidden *et al.* 2021). Changes to climatic variation can impact animal population dynamics across several temporal scales, with intra-annual climatic variation giving rise to seasonal population patterns and inter-annual variation leading to differences in population patterns between years. Dependence on climate has been found to be important in driving pathogen dynamics in many vertebrate populations (Kausrud *et al.* 2007; Dobson 2009; Luis *et al.* 2010; Lal *et al.* 2012; Peel *et al.* 2014; Scherer *et al.* 2019). Therefore, better quantifying the climatic dependence of zoonotic host dynamics will enable a better understanding of short- and long-term patterns of zoonotic disease.

Rodents are particularly likely to be influenced by climatic patterns. Typified by "fast" life histories, rodents invest in different traits under different environmental conditions to maximise the number of offspring produced over their relatively short lifespan (Dobson 2007). Often rodents will reproduce during one portion of the year and not reproduce during the rest of the year, resulting in seasonal birth pulses which can have important consequences for disease dynamics (Kallio *et al.* 2009; Peel *et al.* 2014). Additionally, many rodent species have been shown to have shifting demographic structures, causing nonlinear dynamics which can lead to population surges (Andreassen *et al.* 2021) since demographic traits can have knock-on effects on vital rates (rates relating to birth and death). For instance, the sexual maturation of *Mastomys natalensis* varies inter-annually, with implications for

recruitment (Leirs *et al.* 1993). Similarly, population cycles in voles and lemmings have periods in which adults become heavier, resulting in faster sexual maturation and larger litter sizes (Oli 2019). With rodents making up a disproportionate number of zoonotic hosts and often found at high abundance close to humans (Gibb *et al.* 2020b), it is important to build these nonlinear demographic dynamics into models of rodent host and host-pathogen dynamics, and those of other hosts with similar population dynamics.

A popular way of operationalising demographic models is through the matrix population model, in which a demographic trait is described by discrete stages within a population, and vital rates vary across stages (Caswell 2006). The integral projection model (IPM) generalises this approach by allowing the demographic trait to be continuous rather than discrete (Easterling *et al.* 2000; Ellner & Rees 2006). Not only does this continuity allow for more accuracy in inferring the trait-dependent vital rates and improve precision of the model itself, but it also means that inferred rates are more interpretable when the animal does not present distinct and discrete life stages. The IPM framework also has the flexibility to incorporate other processes such as pathogen transmission, by specifying additional traits and how their distributions change or by creating additional compartments and describing movement between compartments (Metcalf *et al.* 2016). This flexibility makes the IPM an excellent candidate for constructing a demographic model of zoonotic hosts upon which to base a process-driven model of the host-pathogen system.

Given the importance of Lassa fever, caused by infection with Lassa virus, as a neglected zoonotic disease, understanding the population dynamics of the principal zoonotic reservoir host *M. natalensis* could yield insights helpful for disease forecasting and mitigation. *M. natalensis* is thought to be strongly influenced by climate, potentially underpinning the strong seasonal trends in Lassa fever incidence (Leirs *et al.* 1989). There is strong evidence that the rodent exhibits seasonal population dynamics (Leirs *et al.* 1990, 1994; Christensen 1993; Sluydts *et al.* 2007; Olayemi *et al.* 2018), and further evidence suggests that abundance is greater following periods of increased rainfall (Mayamba *et al.* 2021) and that the demographic structure of the population changes depending on precipitation patterns (Leirs *et al.* 1993). This link between reservoir host population dynamics and precipitation has been proposed as an explanation for the outbreaks of Lassa fever in West Africa which are observed in November to February each year. While precipitation and other environmental factors have been found to be predictive of temporal patterns of Lassa fever incidence

(Redding *et al.* 2021), it is not yet clear whether this link is limited to seasonal rodent population dynamics (i.e., rodent population size varying within years on an annual pattern), or whether inter-annual variability in precipitation (i.e., deviations from the seasonal pattern) could also influence rodent population dynamics and Lassa fever incidence. With its importance as an agricultural pest and zoonotic reservoir host, *M. natalensis* is therefore an excellent candidate for which to examine and model climate-driven demography.

Here, I develop an IPM-based approach to model observed demographic processes as functions of lagged climatic variables. I first ask how the climate influences *M. natalensis* demographic processes, testing for and parameterising the effect of lagged seasonal and inter-annually variable climate on the demographic processes of *M. natalensis* using data from a capture-mark-recapture (CMR) study in East Africa. Next, I seek to characterise the dependence of demographic processes on the climate by developing a predictive population dynamic model. I use the fitted demographic processes to derive a dynamic demographic population model based on the IPM framework which can be used to simulate *M. natalensis* population dynamics based on climate-driven demographic processes. Finally, I ask whether climate-driven demographic processes are important and valuable in understanding *M. natalensis* population dynamics, or whether existing standard approaches would be just as suitable. I therefore compare the climatic-demographic model with non-demographic approaches to assess whether the new approach is an appropriate model of *M. natalensis* population dynamics.

3.3 Methods

3.3.1 Data

I used trapping data of *M. natalensis* captured in a CMR study on a mosaic field at Sokoine University of Agriculture, Morogoro, Tanzania (6.51°S, 37.38°E) (Figure 3.1A–B) from 1994 to 2019, provided by the University of Antwerp and Sokoine University of Agriculture, to infer demographic processes. Capture sessions ($n = 296$) were approximately every 28 days, usually consisting of three consecutive nights, recording 41,349 captures of 17,877 tagged individuals. For each individual, data were collected on sex (male or female), body weight *w* (in grams), reproductive condition, and tag identification. For female individuals, reproductive condition was defined as sexually mature or immature (perforate or closed vagina, respectively), lactating or not lactating (nipples lactating or small, respectively), and pregnant or not pregnant (based on visible pregnancy, which is detectable in roughly the

latter half of the 21-day gestation period). I removed data in which weight was missing or erroneously recorded as 0, or in which tagging information was not provided, leaving 7,995 tagged females recorded in 18,611 captures.

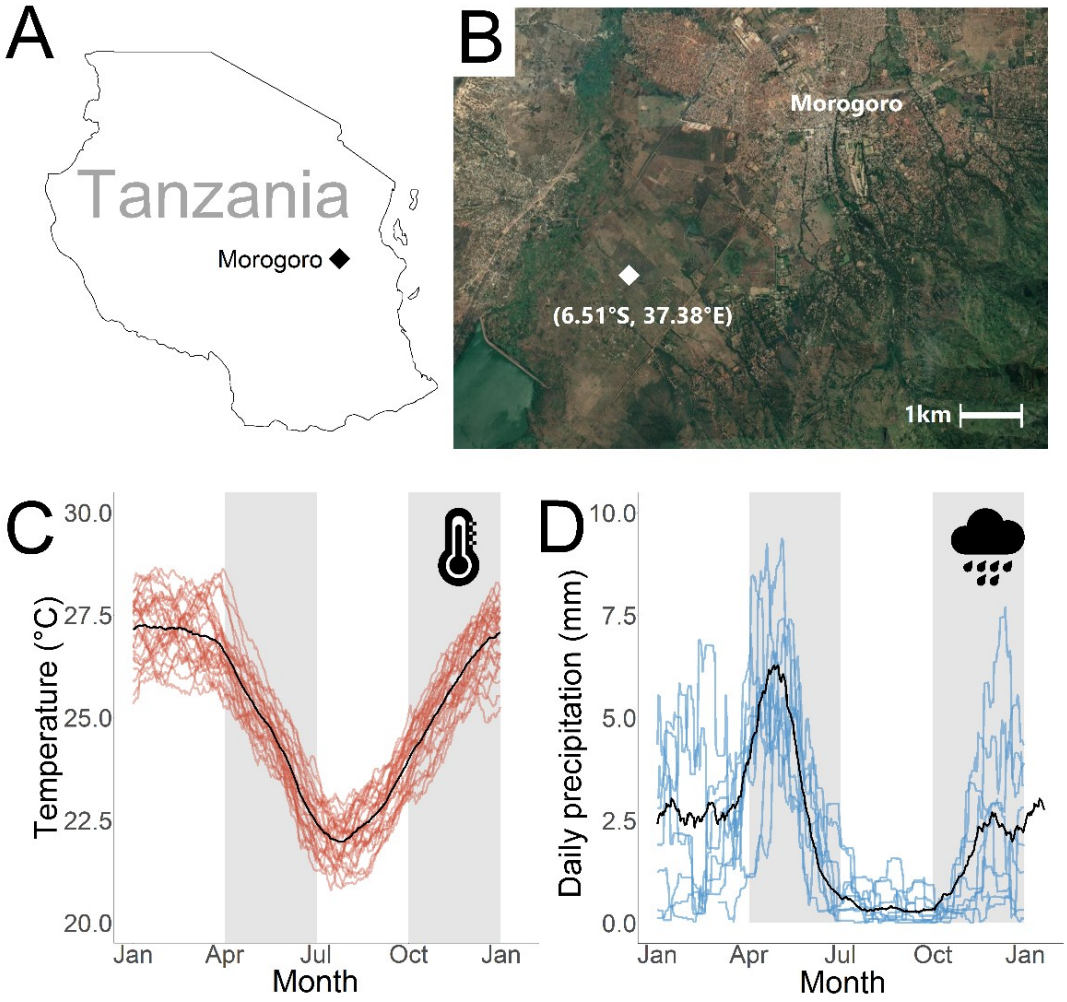


Figure 3.1: CMR study location and patterns of weather and seasonal climate at the site. A: The spatial context of the study location Morogoro within the United Republic of Tanzania. B: Satellite image of the area south-west of the city of Morogoro, with the exact location of the study site shown. Map data © Google, CNES / Airbus, Maxar Technologies; 2021. Reproduced in line with Google Earth guidelines. C–D: Daily mean climate metrics at the study site across the study period (1994–2019) averaged with a 28-day rolling average. C: Temperature seasonality (T_{seas}) shown with a black line, with the inter-annual variability around this seasonal trend shown by the overlaid red lines ($T_{seas} + T_{var}$). Seasonal temperature is highest in the period from January to March and lowest in the months of July and August. D: Precipitation seasonality (P_{seas}) shown with a black line, with the inter-annual variability around this seasonal trend shown by the overlaid blue lines ($P_{seas} + P_{var}$). April to June is usually the period with the most rainfall, with little rain between July and October. There is substantial inter-annual variability around the seasonal trend in the months November to April.

I extracted mean daily temperature (°C) and total daily precipitation (mm) at the study site from the National Oceanic and Atmospheric Administration Climate Prediction Centre (NOAA CPC) total daily precipitation and daily temperature global gridded datasets (NOAA/OAR/ESRL PSD 2021a, b). For both temperature and precipitation I computed a 28-day rolling average to reflect the temporal resolution of the rodent capture sessions, and then decomposed these values into seasonal and variable components (see Table 3.1, Figure 3.1C–D).

Table 3.1: Methods for deriving seasonal and variable climate metrics for precipitation and temperature at the CMR study site in Morogoro, Tanzania. Before calculation of seasonal components, values were imputed for non-existent leap days based on the last day of February and the first day of March, e.g., T_2(29/02/2001) = (T_2(28/02/2001) + T_2(01/03/2001))/2. This was done to avoid a discontinuity in the climate metrics on leap days.

	Temperature	Precipitation
Extracted value	$T_1(d) =$ mean temperature at the study site on day d (°C)	$P_1(d) =$ total precipitation at the study site on day d (mm)
Averaged value	$T_2(d) =$ mean temperature across the period $[d-27, d]$. $= \text{mean}_u(T_1(u) \mid d-27 \leq u \leq d))$	$P_2(d) =$ mean daily precipitation across the period $[d-27, d]$ $= \text{mean}_u(P_1(u) \mid d-27 \leq u \leq d)$
Component of seasonality	$T_{\text{seas}}(d) =$ mean of T_2 on day d of the year across the study period $= \text{mean}_u(T_2(u) \mid u$ is the same day of the year as $d)$	$P_{\text{seas}}(d) =$ mean of P_2 on day d of the year across all years of the study period $= \text{mean}_u(P_2(u) \mid u$ is the same day of the year as $d)$[1]
Component of variability	$T_{\text{var}} =$ averaged temperature metric minus the seasonal component $T_{\text{var}}(d) = T_2(d) - T_{\text{seas}}(d)$	$P_{\text{var}} =$ averaged precipitation metric minus the seasonal component $P_{\text{var}}(d) = P_2(d) - P_{\text{seas}}(d)$

3.3.2 Climate-driven demographic processes

Models of survival, pregnancy, and body growth

With the aim of characterising the survival, pregnancy, and body growth of *M. natalensis*, I designed models to test for and quantify the dependence of these processes on climatic variables and population size (both of which are functions of time t), and the individual's body weight w. The vectors of covariates for each process were denoted $\boldsymbol{X}_S(w,t)$ (survival), $\boldsymbol{X}_P(w,t)$ (pregnancy), and $\boldsymbol{X}_G(w,t)$ (body growth), with corresponding vectors of coefficients (to be found) $\boldsymbol{\beta}_S$, $\boldsymbol{\beta}_P$, and $\boldsymbol{\beta}_G$. These processes were estimated based on recapture,

observable signs of recent pregnancy, and changes in body weight in the data (Figure 3.2) as described in further detail below. The timescale used was 28 days to match the timescale of the CMR study, so that $t + 1$ is 28 days after t, and if t is the date of one capture session then $t + 1$ would be the expected date of the next capture session.

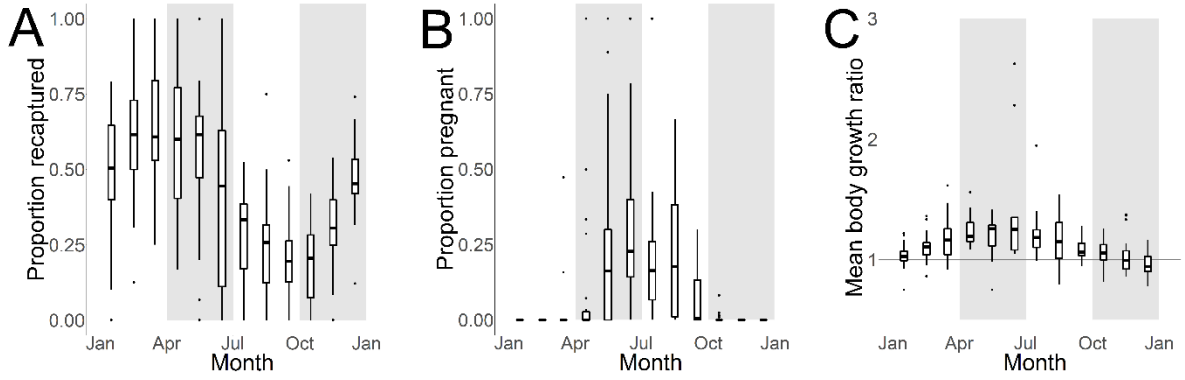


Figure 3.2: Observed demographic changes in the female CMR study population by month. For each calendar month, the distribution of the observed value across different years is shown, highlighting the seasonal trends and inter-annual variation. Boxes show the median and the 25% to 75% interquartile range (IQR). Whiskers show the range of data within 1.5 times IQR outside the box, with outliers shown as individual points. Three-month intervals are shaded alternately in grey and white to aid readability. A: Observed proportion recaptured (at each capture session, proportion of females which are seen at a future session). B: Observed proportion pregnant or recently pregnant (proportion of females at a given capture session which are visibly pregnant or lactating). C: Observed mean body growth ratio (new weight divided by old weight for females caught at consecutive capture sessions).

I modelled survival $S(w, t)$ (whether an individual of body weight w at time t survives to time $t + 1$) as a random variable drawn from a Bernoulli distribution with probability of success (survival) $p_S(w, t)$ (equation 3.1).

$$\begin{aligned} S(w,t) &\sim \text{Bern}(p_S(w,t)) \\ p_S(w,t) &= \mathbb{P}(\text{alive at time } t+1 \mid w,t) \end{aligned} \tag{3.1}$$

However, survival was not directly observable in the data. Since dead individuals cannot be trapped (though on rare occasions individuals will move into the trap while alive and die overnight before the traps are checked), the only way death of an individual is indicated in the data is that they are not recaptured. However, lack of observation could be accounted for by migration or simple stochasticity. Therefore, I defined an additional random variable; recapture, $R(w, t)$ – whether an individual of body weight w at time t is recaptured (Figure 3.2 A), with probability of success (recaptured any time after time t) $p_R(w, t)$. Survival

probability p_S can be written in terms of recapture probability p_R by conditioning the probability of survival on recapture. Suppose an individual was weight w when they were captured at time t, then we can write their survival probability as

$$\begin{aligned} p_S(w,t) &= \mathbb{P}(\text{alive at } t+1) \\ &= \mathbb{P}(\text{alive at t} + 1 \mid \text{recaptured after } t)\mathbb{P}(\text{recaptured after } t) \\ &+ \mathbb{P}(\text{alive at } t+1 \mid \text{never recaptured after } t)\mathbb{P}(\text{never recaptured after } t) \\ &= 1 \cdot p_R(w,t) + \mathbb{P}(\text{alive at } t+1 \mid \text{never recaptured after } t) \cdot (1 - p_R(w,t)) \end{aligned}$$

$$\begin{aligned} \implies p_S(w,t) &= p_R(w,t) + \kappa(w,t)\big(1 - p_R(w,t)\big) \\ &= \kappa(w,t) + (1 - \kappa(w,t))p_R(w,t) \end{aligned} \tag{3.2}$$

where $\kappa(w,t)$ is the probability that an individual captured at time t with weight w is still alive at time $t+1$ even though they were never recaptured, accounting for the factors of migration and stochasticity. For simplicity, I assumed this was constant ($\kappa(w,t) \equiv \kappa$), and termed this parameter the recapture modifier. Although this likely varies due to temporal patterns of migration, and may also depend on body weight, it would not be possible to distinguish between κ and p_R if κ were allowed to vary.

I restructured the data to observe the binary recapture outcomes of the random variable R. For each record, the time t and measured body weight w was noted. The observed outcome of $R(w,t)$ was then recorded as 1 if the individual was recaptured, and 0 otherwise. Since recapture was a binary outcome, I modelled recapture probability using a generalised linear model (GLM) with a binomial error structure, so that $\text{logit}(p_R(w,t))$ was a linear function of the covariates in $\boldsymbol{X}_S(w,t)$ (equation 3.3). Equations 3.1–3.3 combined then provide the distribution of $S(w,t)$.

$$\text{logit}(p_R(w,t)) = \boldsymbol{X}_S(w,t)\boldsymbol{\beta}_S \tag{3.3}$$

Pregnancy $P(w,t)$ (whether a female of body weight w at time t is pregnant at time $t+1$) was also modelled as a Bernoulli random variable, with probability of success (pregnancy) $p_P(w,t)$ (equation 3.4).

$$\begin{aligned} P(w,t) &\sim \text{Bern}(p_P(w,t)) \\ p_P(w,t) &= \mathbb{P}(\text{pregnant at time } t+1 \mid w,t) \end{aligned} \tag{3.4}$$

Recent pregnancy could be directly observed from the data, since individuals' reproductive status are noted. Depending on whether an individual is pregnant, lactating, or both, the

windows in which pregnancy began can be inferred by constructing a timeline based on the reproductive timings of *M. natalensis* (Figure 3.3).

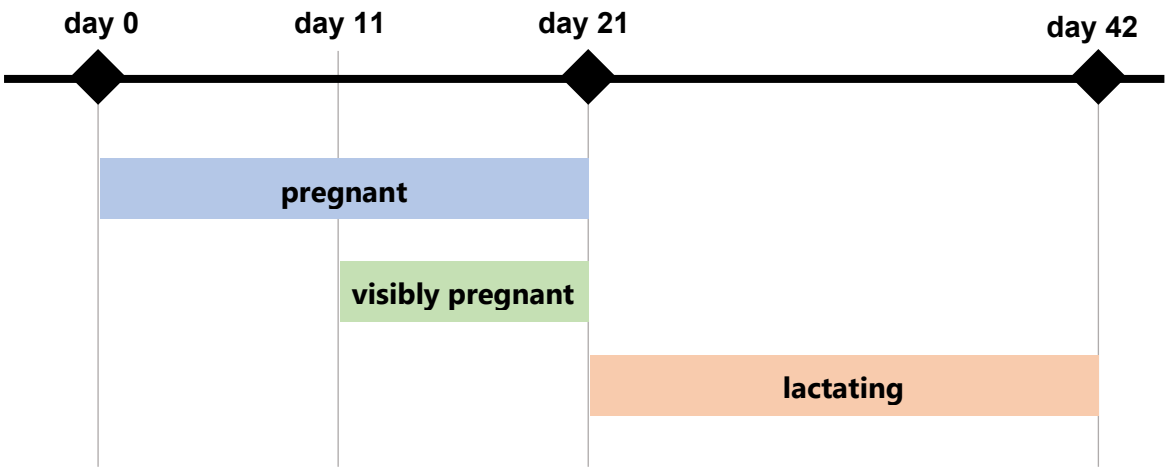


Figure 3.3: The approximate reproductive timings of M. natalensis. The average gestation period is 21 days, and pregnancy is visible in approximately the latter half of this period. Lactation then occurs for around a further 21 days after birth. Therefore, an individual which becomes pregnant on day 0 has the approximate reproductive timings shown by this timeline.

In terms of the relationship between pregnancy and body weight, the objective was to quantify how intrinsic ecological traits impacted an individual's likelihood of pregnancy. For instance, an individual with a larger body weight may be more likely to become pregnant since their body size could be an indicator of sexually maturity, or because individuals who are healthier or better fed are more likely to invest in pregnancy. If individuals which were currently pregnant were included in this analysis, then the increased weight due to pregnancy would unfairly bias the model. Therefore, to observe pregnancy in the data, I first made a dataset containing all records of females which are not currently pregnant or lactating, noting the time t and body weight w. I then restricted these records to ones which had a corresponding consecutive capture record; defined as a record between three and five weeks after the current record (since capture sessions were often not exactly 28 days apart). Lactation U was recorded as 1 if the individual was lactating in the consecutive record, and 0 otherwise. Visible pregnancy V was recorded as 1 if the individual was visibly pregnant in the consecutive record, and 0 otherwise. Based on the observed binary outcomes U and V and the reproductive timings of *M. natalensis* (Figure 3.3), the expected date that pregnancy began was inferred (Table 3.2).

Table 3.2: Inferred date of pregnancy based on observed signs of pregnancy. U is the binary outcome of lactation, with $U = 1$ indicating that on date d the individual captured was visibly lactating. V is the binary outcome of visible pregnancy.

Date of second record	U	V	Expected date pregnancy began
d	0	0	No pregnancy
d	0 or 1	1	$d - 16$ days
d	1	0	$d - 32$ days

As with recapture, the probability of pregnancy $p_P(w,t)$ was modelled using a GLM with a binomial error structure so that $\text{logit}(p_P(w,t))$ was a linear function of the covariates associated with pregnancy (equation 3.5).

$$\text{logit}(p_P(w,t)) = \boldsymbol{X}_P(w,t)\boldsymbol{\beta}_P \tag{3.5}$$

I modelled body weight $G(w,t)$, the body weight at time $t+1$ given weight w at time t, as a normally distributed random variable (equation 3.6), denoting the probability density function of $G(w,t)$ by $p_G(u|w,t)$ (equation 3.7). The mean of this distribution, μ_G, was modelled as a linear function of the covariates in $\boldsymbol{X}_G(w,t)$, and the standard deviation, σ_G, was a linear function of body weight only (equations 3.8–3.9).

$$G(w,t) \sim \text{N}\big(\mu_G(w,t), \sigma_G(w)\big) \tag{3.6}$$

$$\Rightarrow p_G(u|w,t) = \frac{1}{\sigma_G(w)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{v-\mu_G(w,t)}{\sigma_G(w)}\right)^2\right) \tag{3.7}$$

$$\mu_G(w,t) = \boldsymbol{X}_G(w,t)\boldsymbol{\beta}_G \tag{3.8}$$

$$\sigma_G(w) = a_G + b_G w \tag{3.9}$$

Body weight was observed by identifying all pairs of consecutive capture records for each individual. Given a body weight w at time t and a body weight of v in the consecutive record, the observed value of $G(w,t)$ was v.

Population size

To allow for density dependence, I included estimated relative population size as a covariate in the models of demographic processes, and used the minimum number alive (MNA) method as a measure of relative population size. For each capture session, the MNA was equal to the total number of individuals observed, plus the total number of individuals which

were not observed during that capture session but were observed both before and after. MNA can provide a robust indicator of relative population size (Graipel *et al.* 2014), while estimates of the absolute population size as obtainable from more involved probabilistic approaches (e.g., Laake *et al.* 2013) were not needed. Estimated relative population size $A(t)$ was standardised to have a mean value of 1.

Climate lags and covariates

To model the delayed effect of climatic variables on the demographic processes due to ecological processes such as vegetation growth, I introduced lags in the climatic variables. I allowed precipitation and temperature to each have a different climatic lag for each of the three demographic processes, meaning there were six lags in total; $L_{P,S}$ and $L_{T,S}$ (precipitation and temperature lags, respectively, for survival), $L_{P,P}$ and $L_{T,P}$ (pregnancy), and $L_{P,G}$ and $L_{T,G}$ (body growth).

The vectors of covariates then took the following form:

$$\boldsymbol{X}_S(w,t) = \left(1, w, A(t), P_{\text{seas}}(t - L_{P,S}), P_{\text{var}}(t - L_{P,S}), T_{\text{seas}}(t - L_{T,S}), T_{\text{var}}(t - L_{T,S})\right) \quad (3.11)$$

$$\boldsymbol{X}_P(w,t) = \left(1, w, A(t), P_{\text{seas}}(t - L_{P,P}), P_{\text{var}}(t - L_{P,P}), T_{\text{seas}}(t - L_{T,P}), T_{\text{var}}(t - L_{T,P})\right) \quad (3.12)$$

$$\boldsymbol{X}_G(w,t) = \left(1, w, A(t), P_{\text{seas}}(t - L_{P,G}), P_{\text{var}}(t - L_{P,G}), T_{\text{seas}}(t - L_{T,G}), T_{\text{var}}(t - L_{T,G})\right) \quad (3.13)$$

Model fitting

The grid of tested climate lags for temperature and precipitation had a resolution of seven days, with a range of zero to 168 days for both lags. For each demographic process, at each point on the climate lag grid a model was fitted for every combination of covariates. Only those models for which all tested covariates were significant ($\alpha = 0.05$) were kept, and the remaining models were assessed using the Akaike information criterion AIC (Burnham & Anderson 1987). I used the principle that a difference in AIC of at least 2 ($\Delta\text{AIC} \geq 2$) between a candidate model and the model with the lowest AIC indicates that the candidate model should be rejected in favour of the model with the lowest AIC, while a difference of less than 2 ($\Delta\text{AIC} < 2$) indicates that the candidate model cannot be ruled out and should be considered alongside the model with the lowest AIC.

Validation of the demographic models was carried out by fitting the models to a randomly sampled 80% of the rodent demographic data and validating on the remaining 20%. Validation using a chi-squared test or similar was not appropriate since every observation

was modelled to have been drawn from a unique distribution dependent on individual body weight and time of capture. Additionally, a measure of the size of residuals such as root mean squared error was not appropriate for recapture and pregnancy probabilities since the observed binary outcomes (0 or 1) would differ substantially from the predicted probabilities (between 0 and 1). I therefore assessed model performance by generating predicted probabilities and computing the likelihood that the observed data resulted from these models of recapture, pregnancy, and body growth. I then compared this with the likelihood that the observed data resulted from null models of the processes, in which the probability of each event is fixed at the observed proportion of times the event occurred across the validation dataset, ignoring body weight and time. Using the likelihood ratio test, I then calculated the significance level of my model being a better explanation for the data than the null model for each of the three processes ($\alpha = 0.05$; $\mathrm{df} = 6$ since the climatic-demographic model had 6 additional covariates for each process compared with the null model).

Model robustness was tested by comparing the resulting candidate models on two further random 80% subsets of data, identifying whether climate lags and covariates were the same between the three models and, if so, comparing estimated coefficients. All figures were produced using the *R* package *ggplot2*. For visualisation only, distributions were smoothed with kernel density estimations using *ggplot2::geom_density*, using the default bandwidth calculation method.

3.3.3 Climatic-demographic simulation

Model specification

The full IPM takes the form of a set of equations to project the distribution of body weights in the population forwards in discrete time steps based on modelled survival, recruitment, and body weight. The distribution of body weights w at time t is denoted by $D(w,t)$, so that the number of individuals at time t whose weights are between v and z is given by $\int_v^z D(w,t)\,\mathrm{d}w$.

To model the demographic process of recruitment, I described how the previously modelled pregnancy probability $p_P(w,t)$ would lead to recruitment of offspring by creating a sub-compartment of pregnant individuals $J(w,t)$ within the full distribution $D(w,t)$. Individuals determined to become pregnant in the time $(t, t+1]$ would be in compartment J at time $t+1$. I then assumed that if the pregnant individual survives to time $t+2$ they recruit l

offspring, where l is a constant, which follow a distribution of body weights given by $v(w)$. Importantly, individuals in compartment J at time t can return to compartment J at time $t+2$, to reflect that female *M. natalensis* can become pregnant approximately 4 weeks after giving birth.

Table 3.3: List of quantities in the full climatic-demographic model.

Quantity	Name	Value / Method of calculation	Meaning
$S(w,t)$	Survival	Modelled (equations 3.1–3.2)	Whether an individual of body weight w captured at time t survives to time $t+1$.
$R(w,t)$	Recapture	Modelled (equation 3.3)	Whether an individual of body weight w captured at time t is ever captured again.
$P(w,t)$	Pregnancy	Modelled (equations 3.4–3.5)	Whether an individual female of body weight w captured at time t is pregnant at time $t+1$ (given that she is still alive at time $t+1$).
$G(w,t)$	Growth rate	Modelled (equations 3.6–3.9)	Body weight at time $t+1$ of an individual with body weight w at time t.
$A(t)$	Population size	Estimated (equation 3.10)	Estimated abundance of rodents in the population at time t.
$\boldsymbol{X}_S(w,t)$ $\boldsymbol{X}_P(w,t)$ $\boldsymbol{X}_G(w,t)$	Vectors of covariates	From the data, using the estimated climatic lags (equations 3.11–3.13)	Covariates for survival (S), pregnancy (P) and new body weight (G) for an individual with body weight w at time t.
$\boldsymbol{\beta}_S, \boldsymbol{\beta}_P, \boldsymbol{\beta}_G$	Vectors of coefficients	From the GLMs	Coefficients corresponding to the covariates in $\boldsymbol{X}_S, \boldsymbol{X}_P, \boldsymbol{X}_G$.
$J(w,t)$	Distribution of pregnant females	Simulated (equation 3.14)	Abundance of pregnant females of body weight w at time t. Note that projected body weight due to growth is calculated ignoring pregnancy.
$D(w,t)$	Distribution of all rodents	Simulated (equation 3.15)	Abundance of rodents with body weight w at time t.
$v(w)$	Juvenile weight distribution	Assumed; probability density function for the distribution $N(\mu = 13, \sigma = 0.3)$	Probability density function for the body weight w (in grams) of an individual joining the population 35 days after birth. This distribution was chosen based upon a study which recorded the weight of female *M. natalensis* 28 days after birth (Jackson & van Aarde 2003).
l	Litter size	Free parameter	Expected number of offspring in a litter.
κ	Recapture modifier	Fitted as a function of l to produce stable population	Probability of an individual who was recorded at some time t as not having been seen again in the study still being alive at time $t+1$. (Assumed constant).

The full model (Table 3.3) is then given by the integral projection equations:

$$J(w,t+1) = \int_0^\infty p_G(w|y,t)p_S(y,t)p_P(y,t)\big(D(y,t) - J(y,t)\big)\,\mathrm{d}y \tag{3.14}$$

$$D(w,t+1) = \int_0^\infty p_G(w|y,t)p_S(y,t)D(y,t) + lp_S(y,t)J(y,t)v(w)\,\mathrm{d}y \tag{3.15}$$

Simulation

The full model was simulated on a grid with time step of 28 days and body weight step of 1 gram, with the integrals discretised into sums and the probability density function for body growth (p_G) converted into a probability mass function as follows.

A body weight step of 1 gram was chosen to provide precision while maintaining computational speed. Smaller values of body weight step were tested and no discernible differences in resulting simulations were found. An upper limit of 100 grams was chosen because there were very few records of individuals above this body weight, and it was found that this body weight was never reached during simulations.

Equations 3.14–3.15 were discretised by converting the continuous probability density functions $p_G(w|y,t)$ and $v(w)$ into discrete probability mass functions using equations 3.16–3.17, and then by converting integrals into sums over the set of possible weights $W = \{1, 2, 3, \ldots, 100\}$. Equations 3.18–3.19 then give the discretised form of the model to produce the simulations.

$$p_g(w|y,t) = \frac{p_G(w|y,t)}{\sum_{z\in W} p_G(z|y,t)} \tag{3.16}$$

$$p_v(w) = \frac{v(w)}{\sum_{z\in W} v(z)} \tag{3.17}$$

$$J(w,t+1) = \sum_{y\in W} p_g(w|y,t)p_S(y,t)p_p(y,t)(D(y,t) - J(y,t)) \tag{3.18}$$

$$D(w,t+1) = \sum_{y\in W} p_g(w|y,t)p_S(y,t)D(y,t) + lp_S(y,t)J(y,t)v(w) \tag{3.19}$$

The initial population size was set as 10,000 and the distribution of weights was chosen by randomly sampling across the capture data. Since the modelled population size is only relative, a mean population size can be set without affecting the dynamics. The litter size l was set as 5 females (Leirs *et al.* 1993), and recapture modifier κ was estimated by calibrating the model with different values of κ until the mean population size was the stipulated (arbitrary) value, 10,000.

The simulations were only of females, assuming that there would be enough males in the population for recruitment to occur. A burn-in period of 24 time steps (nearly 2 years) was used, and free parameters α (capture success) and l (litter size) were varied (Table 2).

Comparison with non-demographic approaches

To compare the climatic-demographic model with approaches which do not consider demographic processes, I defined four alternative models as follows. (1) Null – an unstructured model in which the relative population size at each time is as a normal distribution with mean 1 and variance equal to the variance across the relative population sizes A at each capture session. (2) Linear – a linear model in which the population size could depend linearly on lagged climatic variables, with lags selected and the final model fit using linear regression and minimising mean squared error. I then defined two mass-action population models, in which population growth rates depended linearly on lagged climatic variables, fit in the same manner as the linear model, where (3) linear growth was a linear growth model, with population growth rate at an observation at time t calculated as $A(t) - A(t-1)$, and (4) exponential growth assumed an exponential growth rate, so that the population growth rate was $\log(A(t)/A(t-1))$. I wanted to compare models 1–4 with my climatic-demographic model (5), so I computed residuals (predicted relative population size minus estimated relative population size), variance of residuals, and median squared error.

3.4 Results

3.4.1 Climate-driven demographic processes

The temperature lags L_T were estimated as 70 days for survival ($\Delta\text{AIC} = 3.76$ compared with the next best model) and 154 days for pregnancy ($\Delta\text{AIC} = 2.20$). The precipitation lags L_P were estimated as 98 days for survival ($\Delta\text{AIC} = 6.67$) and 63 days for pregnancy ($\Delta\text{AIC} = 4.44$). For body growth, the best model ($\Delta\text{AIC} = 1.64$) had $L_{T,G}$=161 days and $L_{P,G}$=133 days, and was accepted despite having $\Delta\text{AIC} < 2$ since the next best model was similar (Supplementary Table S3.1). For the climate lags selected, the best model (Supplementary Table S3.2) for survival ($\Delta\text{AIC} > 10$) included body weight w, lagged temperature seasonality T_{seas}, lagged precipitation seasonality P_{seas}, and lagged precipitation variability P_{var} ($p < 0.0001$ for all covariates) (Figure 3.4). The best model for recruitment ($\Delta\text{AIC} = 4.21$) included w ($p = 0.01$), relative population size A ($p < 0.0001$), T_{seas} ($p < 0.0001$), and P_{var} ($p < 0.0001$) (Figure 3.5). The best model for growth ($\Delta\text{AIC} > 10$) had w, A, T_{seas}, and P_{seas} as significant

covariates for mean body growth μ_G, and standard deviation σ_G depended on w ($p < 0.0001$ for all covariates) (Figure 3.6).

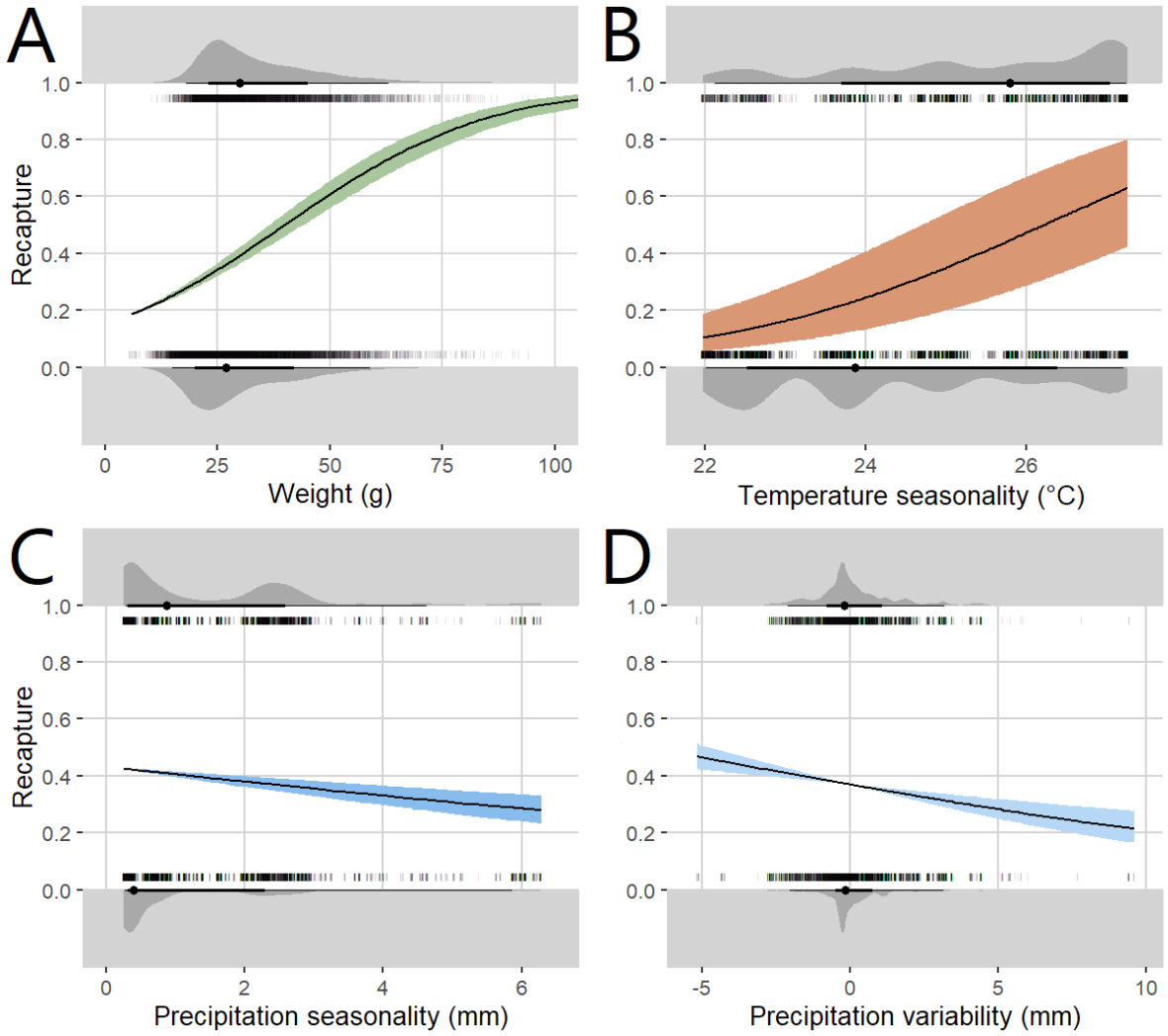


Figure 3.4: Model fit for recapture. Each plot shows predicted recapture probability for varying values of a given covariate, with data (binary recapture response) shown with a density plot, box plot, and points along the top (recaptured again) and bottom (never recaptured again) of the graph. For each given covariate, the predicted probability of recapture is shown with a black line, while the coloured band around this line shows the 95% confidence interval around this prediction based on the uncertainty around the predicted coefficient around the given covariate. A: Body weight (g). B: Seasonal component of temperature (°C). C: Seasonal component of precipitation (mm). D: Interannual variability in precipitation (mm).

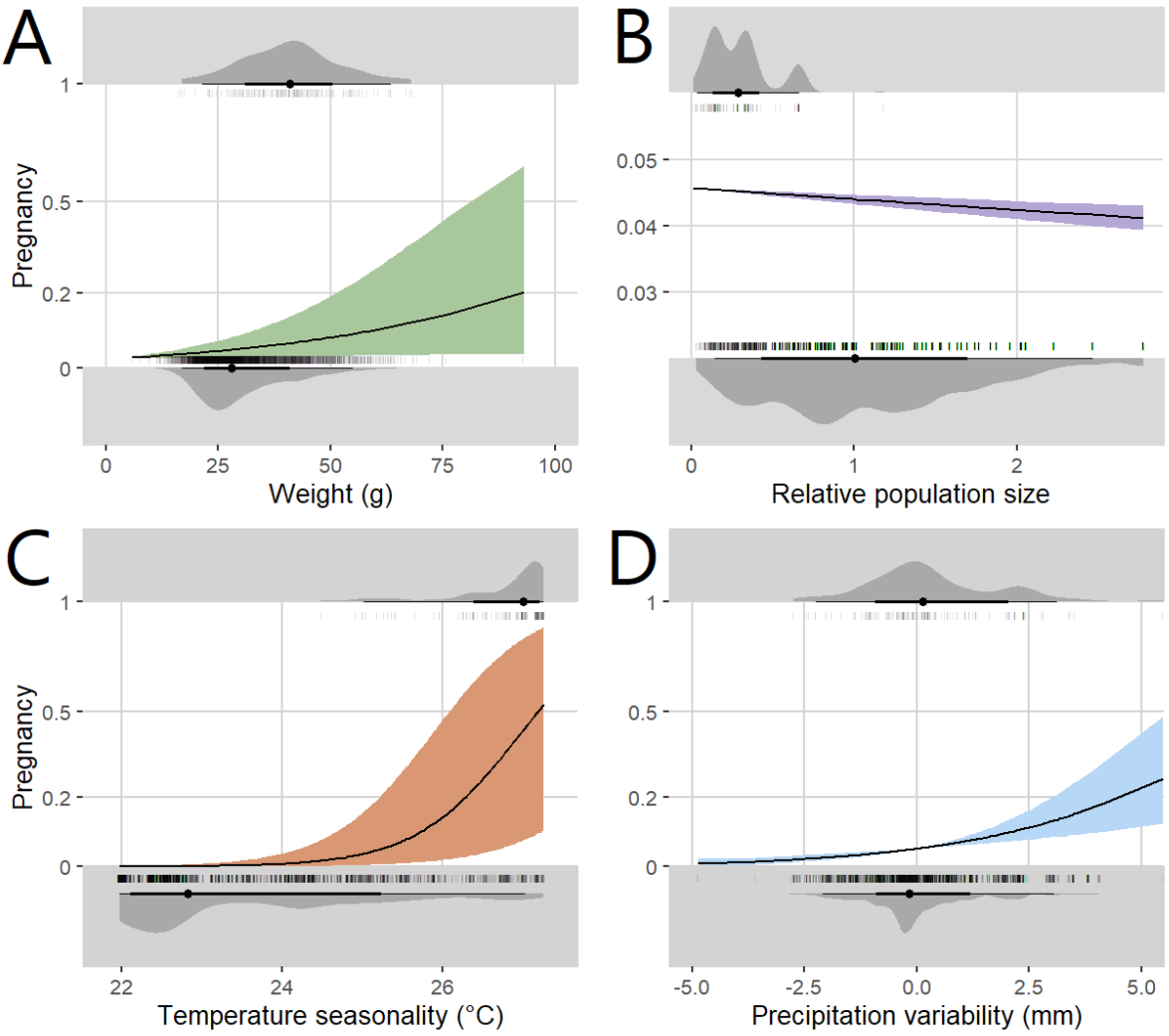


Figure 3.5: Model fit for pregnancy. Each plot shows predicted pregnancy probability for varying values of a given covariate, with data (binary pregnancy response) shown with a density plot, box plot, and points along the top (became pregnant) and bottom (did not become pregnant) of the graph. For each given covariate, the predicted probability of pregnancy is shown with a black line, while the coloured band around this line shows the 95% confidence interval around this prediction based on the uncertainty around the predicted coefficient around the given covariate. Note that in panels A, B, and D, probability of pregnancy is presented on a logarithmic scale to aid readability of small probabilities, while in panel C the axis for probability of pregnancy is cropped since the effect of the covariate on pregnancy probability is small. A: Body weight (g). B: Relative population size. C: Seasonal component of temperature (°C). D: Interannual variability in precipitation (mm).

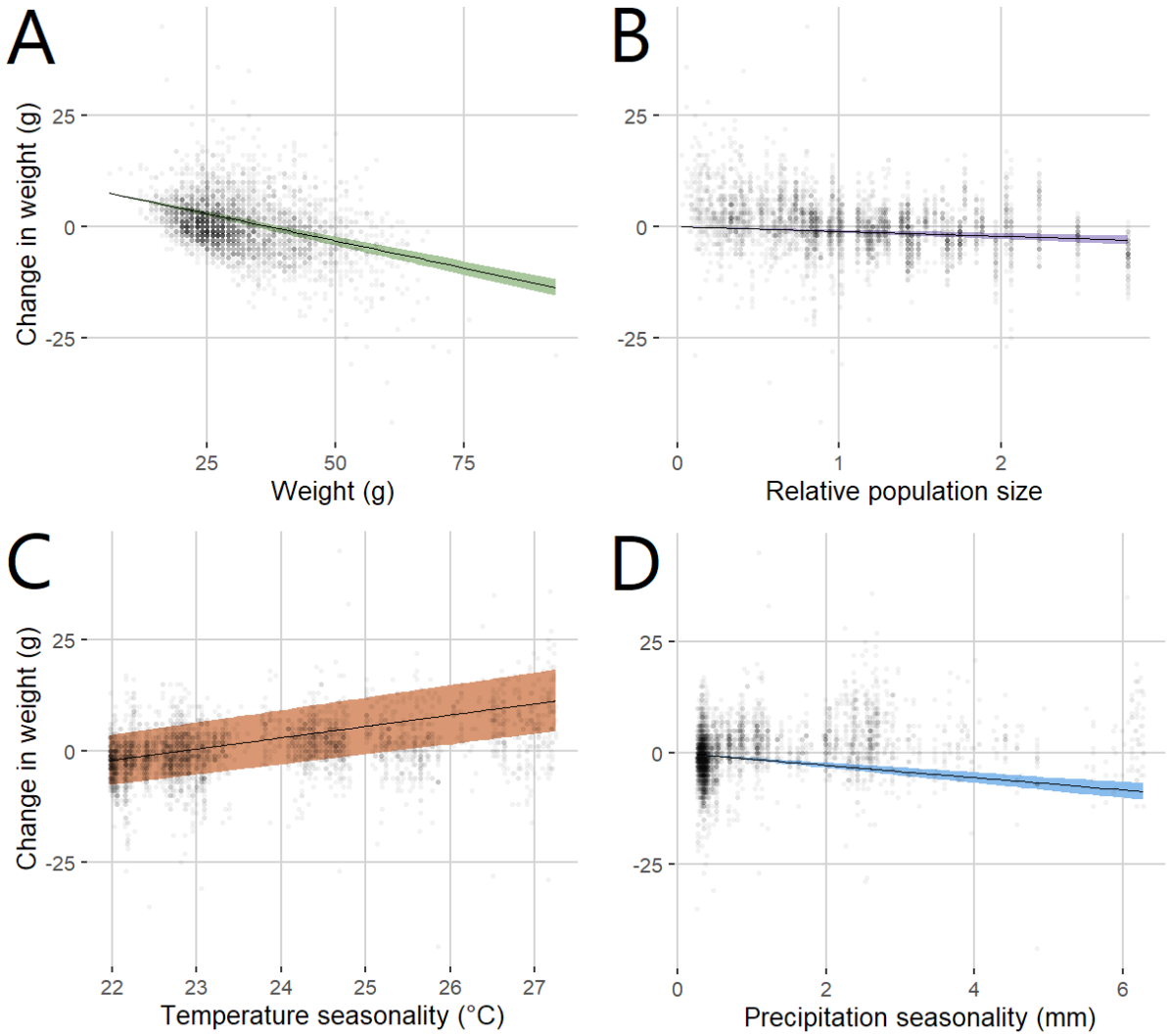


Figure 3.6: Model fit for mean body growth. Each plot shows predicted change in body weight for varying values of a given covariate, with data shown with black points. For each given covariate, the predicted change in body weight is shown with a black line, while the coloured band around this line shows the 95% confidence interval around this prediction based on the uncertainty around the predicted coefficient around the given covariate. A: Body weight (g). B: Relative population size. C: Seasonal component of temperature (°C). D: Seasonal component of precipitation (mm).

Temperature seasonality T_{seas} was the strongest driver of survival probability, with survival rate increasing following the hottest time of the year with a 70-day lag, and was also the strongest driver of recruitment, with pregnancy probability increasing following the hottest time of the year with a 154-day lag (Figure 3.7A–B). Contribution of precipitation seasonality, temperature seasonality, and body size towards mean growth ratio were comparable, and body weight increased standard deviation of growth ratio (Figure 3.7C–D).

The contribution of precipitation variability in the models indicated that in years with more precipitation, recruitment rates are higher and survival rates lower on average. The year with the highest median precipitation variability was 2006, with $\text{median}(P_{\text{var}}) = 0.774$mm, while the year with the lowest median precipitation variability was 2005, with $\text{median}(P_{\text{var}}) = -0.577$mm. Based on these values and with other variables held constant, high precipitation would see an average monthly pregnancy probability of 0.0617 compared with 0.0392 in a low-precipitation year. Average survival probability would be 0.356 in a high-precipitation year, and 0.381 in a low-precipitation year. This means that in a high-precipitation year, *M. natalensis* are on average 57.6% more likely to become pregnant every month, and only 6.51% less likely to survive every month compared with a low-precipitation year.

I found that models for all three processes were significantly better than constant parameter null models ($p < 0.0001$ for survival, $p = 0.0002$ for recruitment, $p = 0.0031$ for growth), indicating a good fit to the observed demographic processes.

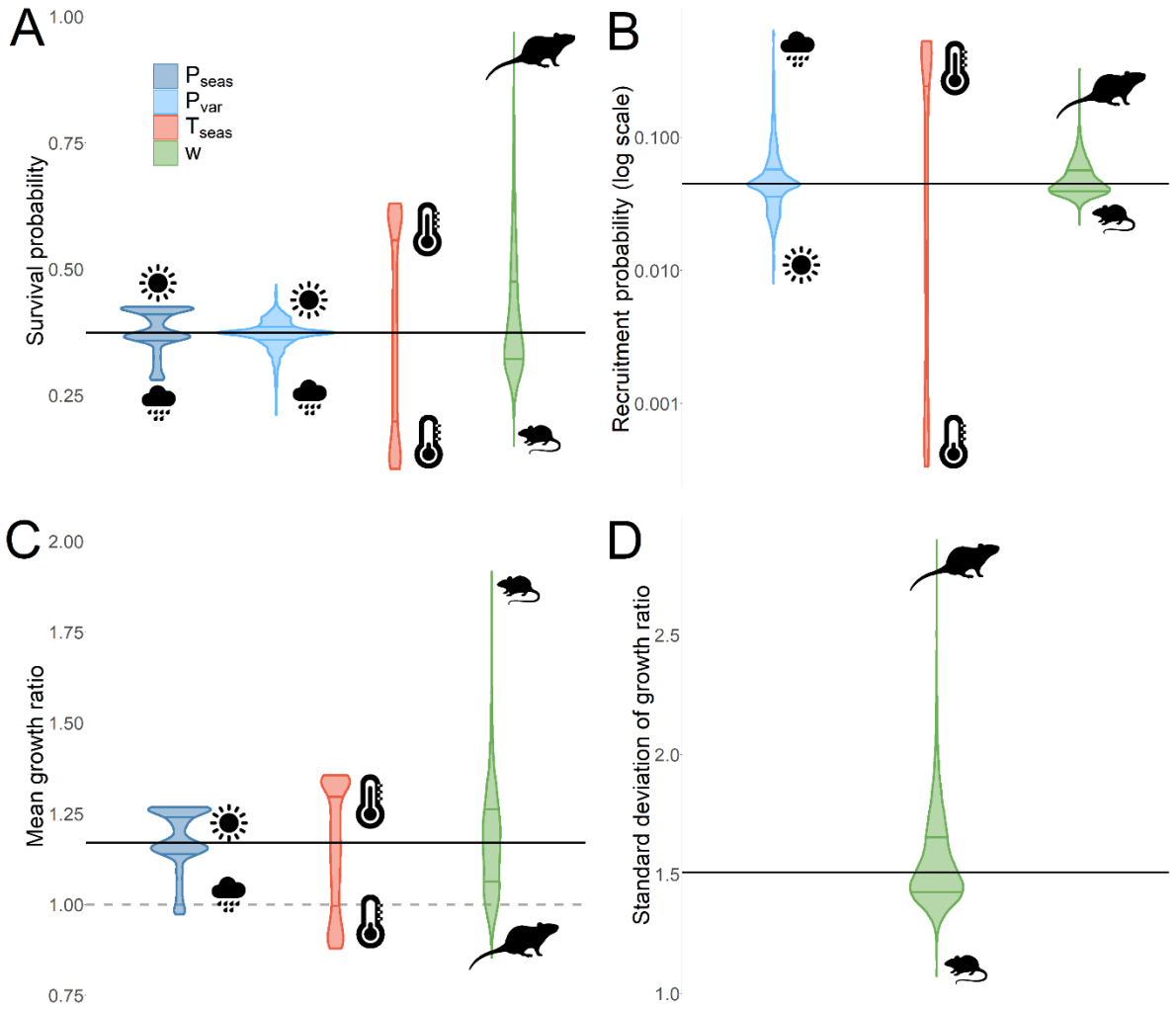


Figure 3.7: Distribution of covariate effects on demographic processes survival, pregnancy, and body growth in M. natalensis. Horizontal black lines show the baseline value of the variable when all covariates are equal to their mean. Violin plots show how the actual distribution of each covariate affects the variable while other covariates are held constant. Icons indicate direction of increasing and decreasing values of the covariate. For all processes, while the effect of population size was significant, the contribution was so small that the density plot was unreadable and therefore this contribution was omitted from the graphs. A: Probability of survival is mainly driven by temperature seasonality; increasing following hot times of the year and decreasing following cooler times of year. Survival is also increased greatly in particularly large individuals. B: The baseline pregnancy probability is low and is especially low following colder times of the year. Probability of pregnancy is greatly increased following periods with more than average precipitation and following hot times of year. C: The mean growth ratio (mean new body size divided by old body size) is similarly affected by all three shown covariates, with growth rate larger for smaller body size and following dry and hot times of the year. D: The standard deviation of body growth ratio increases with body size.

3.4.2 Climatic-demographic simulation

The climatic-demographic model greatly outperformed (1) the null model and both mass-action population models, (3) linear growth and (4) exponential growth (Figure 3.8, Supplementary Table S3.4). The best model was (2) the linear model, with a median squared error of 0.171, while (5) the climatic-demographic model had a median squared error of 0.276, both of which were much lower than that of models 1, 3, and 4.

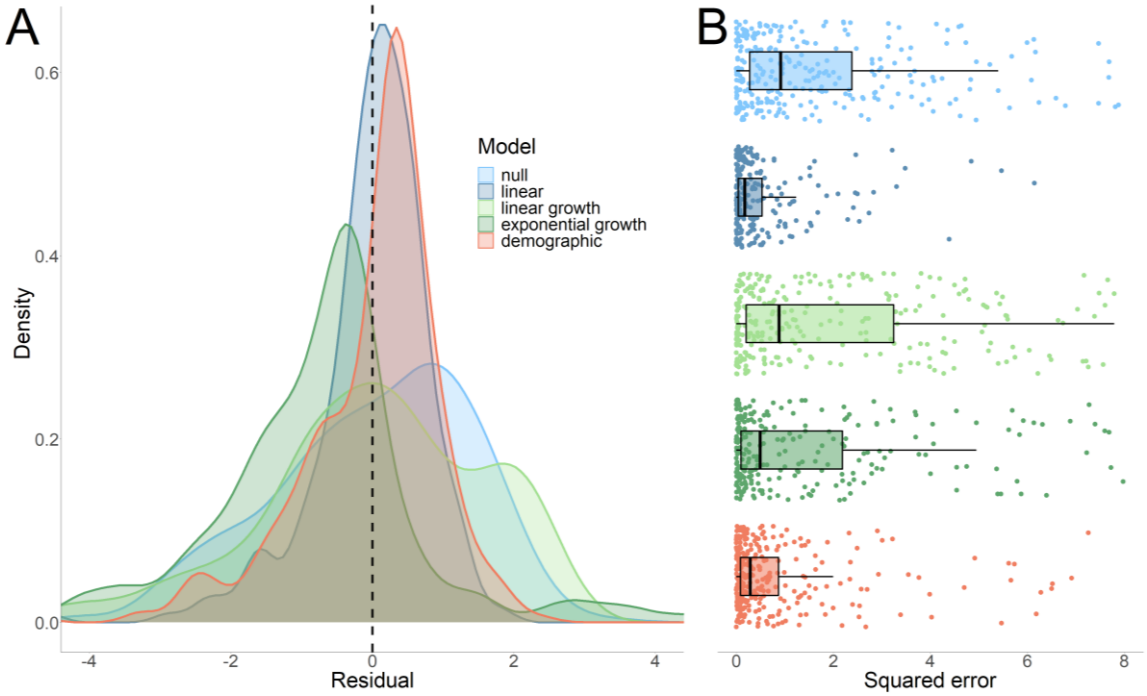


Figure 3.8: Comparison of climatic-demographic simulation performance with non-demographic approaches to predicting population size. The models were (1) null, (2) linear, (3) linear growth, (4) exponential growth, and (5) climatic-demographic. A: Density of residuals (predicted relative population size minus interpolated observed relative population size). The linear model (2) and the climatic-demographic model (5) have the least spread of residuals around 0. B: Squared error (residuals squared) as points, with box-and-whisker plot showing 5%, 25%, 50%, 75% and 95% quantiles. The distribution of squared error is closer to 0 for the linear (2) and climatic-demographic (5) models compared with the other models, with the linear method (2) having the lower median squared error.

3.5 Discussion

By testing for and parameterising the effect of lagged seasonal and inter-annually variable climate on the demographic processes of *M. natalensis*, I found that climate seasonality, inter-annual precipitation variability, and individual body weight are all significant factors in the species' demographic processes. These results imply the presence of nonlinear population dynamics with a seasonal component, but which change year-on-year depending

on climate patterns. Using these results, I adapted the IPM method to simulate the population dynamics of *M. natalensis,* and simulated population size using both this climatic-demographic approach and non-demographic approaches which model population size as driven by the same climatic variables. I found that the climatic-demographic approach was comparable to non-demographic approaches, but with the important advantage of having been fit to more directly-observed processes (rather than variable population size which is challenging to measure in a CMR study) and predicting demographic changes such as births and deaths which are relevant to disease dynamics. The climatic-demographic model can be used as an experimental environment in which to test the impact of different climate scenarios on *M. natalensis* population dynamics and has the potential to be further adapted as a model of infectious disease transmission among the reservoir host population. Importantly, I found that inter-annual variation in rainfall greatly increases pregnancy rate, with high precipitation years predicted to have much larger peak population sizes than average precipitation years.

The finding of a dependence of demographic processes on lagged precipitation seasonality and inter-annual variability is consistent with the suggestion that precipitation is a key driver of *M. natalensis* population size (Fichet-Calvet *et al.* 2007; Makundi *et al.* 2007; Olayemi *et al.* 2018), with the lag implying an ecological delay such as vegetation growth, with precipitation as an indirect factor in recruitment rate. The precipitation lag in pregnancy I found of 63 days is similar to that found in previous analyses of the same study population (Leirs *et al.* 1994) and other analyses and observations of *M. natalensis* (Massawe *et al.* 2015; Mayamba *et al.* 2021). The increase in inferred recruitment in a high-precipitation year greatly outweighed the associated decrease in survival probability, the latter of which could perhaps be explained by a decreased survival rate associated with pregnancy. I found that extremes of precipitation could substantially influence recruitment rate, with a predicted 57% increase in the average rodent population size in the highest-precipitation year compared to the lowest-precipitation year. The dependence of demographic processes on precipitation variability but not temperature variability could suggest that precipitation is an environmental limiting factor in resource availability (Christensen 1993).

The seasonal climate metrics which drove *M. natalensis* population demography were temperature and precipitation for survival, and temperature alone for pregnancy and body growth. The strong positive relationship between lagged seasonal temperature and these

three processes could be indicative of an increased resource availability following periods of warmth and increased sunshine (due to vegetation growth). However, rather than interpreting this relationship as temperature playing a measurable role in driving specific processes, it may be more appropriate to view the rodent's dependence on seasonal temperature as the result of environment-driven phenology. Through this interpretation, the effect of seasonal climate patterns on population demography is simply an underlying annual structure determining timing of population changes. The strong link between temperature and rodent phenology found here echoes that found in a large-scale analysis of the North American deer mouse (Luis *et al.* 2010).

Validation of the population size simulation was imperfect due to the challenges in accurately estimating time-varying population size from CMR data. I estimated the relative population size of the study population using MNA, which was used to fit the linear climate-driven population model. Meanwhile, in the climatic-demographic model, individual processes were fit to observed changes in traits such as pregnancy, which in turn were used to describe population changes and model resulting population size. In other words, the linear model was fit to and tested on the same dataset, while the climatic-demographic model was blind to the relative population size it would be used to simulate. It is therefore unsurprising that the linear model was a better predictor of population size, however, it should not be discounted that under some conditions a linear approach may provide a more appropriate model of relative population size due to its reduced complexity. It would be of interest to apply new developments to MNA estimation (Bright Ross *et al.* 2022) to reassess the relative performance of the climatic-demographic and linear models in predicting time-varying population size.

The method developed here to model the host *M. natalensis* has several advantages over existing approaches. Fitting demographic processes to climate metrics creates an in silico experimental environment which is more suitable for forecasting population and infectious disease dynamics under novel climate scenarios than a population-level approach. With rodent control found experimentally to be ineffective at reducing number of *M. natalensis* rodent hosts on the long term (Mariën *et al.* 2019), this model provides an experimental environment within which to test other interventions. Additionally, by treating demographic processes and climate metrics as continuous variables, I have been able to characterise the processes underpinning *M. natalensis* population dynamics with more precision than earlier

studies of the species which used more categorical descriptions, such as "rainy season" or "dry season" (Leirs *et al.* 1993; Sluydts *et al.* 2007). A demographic, process-based, approach to modelling the population also circumnavigates the challenges in estimating longitudinal, temporally variable population size from CMR studies.

With the changing climate likely to impact dynamics of *M. natalensis*, by quantifying the climatic dependence of its demography the climatic-demographic model creates an opportunity to explore how these impacts might manifest. However, my model does not take into account some important determinants of spatiotemporal *M. natalensis* distribution, such as land-use and land cover, movement and migration, and agricultural patterns. The estimated precipitation lag for recruitment estimated here is consistent with analyses linking climate to Lassa fever incidence (Redding *et al.* 2016; 2021, Supplementary Figure S1.1), after taking into account additional delays due to gestation and disease incubation, suggesting that this model for *M. natalensis* could shed light on the epidemiology of Lassa fever. However, when seeking to apply this model to understand Lassa virus dynamics, it should be emphasised that Lassa fever is endemic in West Africa, not Tanzania, with differing habitat types and genetic divergence (Colangelo *et al.* 2013) making applicability of this model to different population of the same species limited. However, by linking demographic processes to decomposed seasonal and inter-annually variable climate metrics, this model of *M. natalensis* captures separately the rodents' phenology and response to inter-annual changes in precipitation. Given the ethical and health and safety implications of carrying out capture-release studies on an animal which carries a zoonotic pathogen, a model such as this may be the only way to gain demographic insights into known hosts and is a promising direction to explore towards the development of a process-based climate-driven model of Lassa virus in the reservoir host.

The novel approach of modelling continuous demographic traits with lagged climate metrics opens avenues of exploration into the demographic processes of other species and populations of interest, including other rodent hosts of zoonotic diseases. Taking the climatic-demographic model forwards, it would be valuable to incorporate transmission dynamics among the host population, made possible by the demographic structure of the model which, for example, would enable distinction between vertical and horizontal transmission. This in turn could be applied to develop forecasting and intervention models for zoonotic diseases which are driven by climate-sensitive host population dynamics.

Chapter 4: Process-based modelling links climate-driven reservoir host infectious disease dynamics with Lassa fever incidence

4.1 Abstract

Transmission of zoonotic pathogens is an essential and dynamic part of the zoonotic disease system and, alongside reservoir host ecology and other processes, underpins risk of human disease. In Chapter 3 I found that the demographic processes of *Mastomys natalensis* in Tanzania were highly seasonal, and additionally found an important effect of inter-annual variability in precipitation on recruitment rate. These findings and the resulting climatic-demographic model present an opportunity to model this species in the context of pathogen hosting and transmission dynamics. Here, I develop a mathematical model of rodent-to-rodent arenavirus transmission dynamics based upon field observations of arenavirus infection in *M. natalensis*, selecting a model framework in which horizontally-infected and vertically-infected rodents have different shedding and recovery dynamics. I then incorporate this transmission model into the climatic-demographic model to create a full climate-driven model of population and pathogen dynamics. I hypothesise that this model, when applied to Nigeria where *M. natalensis* hosts the zoonotic pathogen Lassa virus, could be used to predict temporal patterns of infection incidence. To test this hypothesis, I first estimate a probabilistic distribution of the incubation period of Lassa fever and transform observed human cases of Lassa fever in Nigeria into inferred infection incidence. I then model the relationship, finding that for the five states with the highest reported incidence of Lassa fever in 2020, temporal patterns of inferred infection incidence were consistent with the climate-driven, process-based model of zoonotic hazard. These results indicate that reservoir host and pathogen dynamics, driven by seasonal and inter-annually variable climate patterns, may be useful in predicting Lassa virus spillover risk, and that these dynamics could explain observed temporal and spatiotemporal patterns of Lassa fever incidence. Given that incidence of Lassa fever is predicted to increase due to climate change, this model provides the opportunity to explore inter-annual and intra-annual changes to temporal dynamics of the zoonotic disease system.

4.2 Introduction

Models of risks associated with zoonotic and vector-borne disease can lead to improved understanding of risk factors (Seidahmed *et al.* 2018), advance forecasting to enable resource allocation (Anyamba *et al.* 2009; Kshirsagar *et al.* 2013) and exploration of mitigation and intervention strategies (Reis *et al.* 2008; Smith *et al.* 2008). Given that zoonotic diseases result from complex multi-species, multi-scale systems, process-based modelling approaches are needed to accurately capture the dynamics underpinning zoonotic disease risk (Plowright *et al.* 2017), especially when considering how a zoonotic disease system might be impacted by novel circumstances such as following climate or land-use change (Gibb *et al.* 2020a).

Host ecology plays a key role in the zoonotic disease system, fundamentally determining the availability of zoonotic hosts in the environment which are capable of transmitting the pathogen to other zoonotic hosts and humans. Nested within zoonotic host population dynamics is pathogen transmission, often resulting variable pathogen prevalence in the host population. Therefore, temporal abundance of infectious hosts is not necessarily proportional to the overall host abundance. Such nonlinear dynamics have been observed in hantavirus dynamics in bank voles (Voutilainen *et al.* 2016; Reil *et al.* 2017), rabies virus dynamics in bats (George *et al.* 2011), and arenavirus dynamics in *Mastomys natalensis* (Mariën *et al.* 2020). Pathogen dynamics vary between disease systems; for instance, density-dependent transmission (e.g., Mariën *et al.* 2019) describes a rate of infection proportional to the spatial density of infected individuals, while in frequency-dependent transmission (e.g., Tompros *et al.* 2021) the rate of infection is proportional to the proportion (frequency) of infected individuals among the entire population. Therefore, estimates of temporal zoonotic hazard – which in this thesis has been defined as the relative abundance of the zoonotic pathogen in hosts which can transmit the pathogen to humans (Gibb *et al.* 2020a) – should ideally combine zoonotic host ecology with transmission dynamics, rather than assuming zoonotic hazard is simply proportional to zoonotic host abundance.

The modelling of process-based population ecology and transmission dynamics is well established in the field of vector-borne disease ecology (Franklinos *et al.* 2019), often through stage-structured models (Ciota & Keyel 2019), a type of demographic population model. In these models of vectors, pathogens and their transmission are explicitly incorporated into the modelling framework to allow for nonlinear dynamics. Modelling the climate-driven dynamics of vectors and their pathogens using demographic models has

allowed forecasting of disease risk and incidence and the recommendation of targeted interventions based on climatic monitoring (Hancock *et al.* 2009; Wang *et al.* 2016; Tjaden *et al.* 2018). However, climate-driven host-pathogen dynamics are rarely modelled for zoonotic disease systems.

For Lassa virus, the climate-driven integral projection model (IPM) of the reservoir host *M. natalensis* developed in Chapter 3 presents an opportunity to build a process-based model of zoonotic hazard. Although the climatic-demographic model was fit to a capture-mark-recapture (CMR) study of a population in Tanzania – outside of the Lassa virus endemic area, the model describes dependence of demographic processes on seasonal and inter-annually variable climate, capturing an increased degree of ecological detail than was previously included. This ecological detail should increase the likelihood of the model being appropriate for describing *M. natalensis* under varying climate patterns. By applying the model to climate patterns in West Africa, it may therefore be possible to predict *M. natalensis* population dynamics within the endemic area of Lassa virus. However, the climatic-demographic model is of reservoir host population dynamics only, and does not include pathogen dynamics. In order to best describe the Lassa virus system given potential nonlinearities in transmission dynamics, it would be valuable to include a host-to-host transmission model within the climatic-demographic model. The resulting model could then potentially be used to forecast temporal Lassa virus zoonotic hazard and explore changes to the Lassa virus system, which in turn could inform mitigation and intervention strategies.

Here, I explore the Lassa virus host-pathogen system and assess the potential for climate-driven reservoir host demography and transmission dynamics to explain human disease incidence. First, I investigate how the climatic-demographic IPM of *M. natalensis* can be extended to include host-to-host arenavirus dynamics constructed from observations based on field data. Then I apply this model of host-pathogen dynamics to Nigeria to predict Lassa virus zoonotic hazard. To link observed cases of human disease with predicted zoonotic hazard, I estimate the human incubation period of Lassa fever using collated nosocomial outbreak data and use this to infer human infection incidence from observed patterns of disease in Nigeria. Finally, I construct a generalised linear model (GLM) to test whether predicted zoonotic hazard is a significant and positive correlate in patterns of inferred human infection incidence and assess to what extent the model created can inform temporal zoonotic hazard and infection risk.

4.3 Methods

First (Section 4.3.1) I develop a process-based model of climate-driven zoonotic hazard by combining an arenavirus transmission model with the climatic-demographic model of *M. natalensis* in Chapter 3. Next (Section 4.3.2), I infer Lassa virus infection incidence in Nigeria by estimating the incubation period distribution of Lassa fever in humans and using this distribution to transform observed Lassa fever cases in Nigeria. Finally, I assess how my model of climate-driven zoonotic hazard can be used to understand and predict patterns of Lassa virus infection incidence in Nigeria by constructing a GLM.

4.3.1 Developing a process-based model of climate-driven zoonotic hazard

The IPM framework (seen in Chapter 3) takes the form

$$J(w, t+1) = \int_0^\infty p_G(w|y,t) p_S(y,t) p_P(y,t) D(y,t)\, \mathrm{d}y \tag{3.14}$$

$$D(w, t+1) = \int_0^\infty p_G(w|y,t) p_S(y,t) D(y,t) + l p_S(y,t) J(y,t) v(w)\, \mathrm{d}y \tag{3.15}$$

where $D(w,t)$ is the abundance of individuals of body weight w at time t, and $J(w,t)$ is the subset of these individuals who are currently pregnant or recently gave birth. This framework was constructed to reflect the reproductive timings of *M. natalensis*; for instance, individuals who become pregnant at time t are then in the pregnancy compartment at time $t+1$ (where t measures time in 28-day steps) but can still become pregnant again at time $t+2$, reflecting that the post-partum oestrous of *M. natalensis* occurs only around 4 weeks after the initial pregnancy (Johnston & Oliff 1954) and observations that in the wild some females are pregnant most months of a breeding season (Leirs *et al.* 1993).

To expand this model to include pathogen transmission dynamics, I constructed a compartmental infectious disease model of the form Susceptible (S)-Infectious$_{\text{horizontal}}$ (H)-Infectious$_{\text{vertical}}$ (V)-Recovered (R). This model structure and later parameter values were based on studies of Morogoro virus, Gairo virus, and Lassa virus in *M. natalensis*; it was not possible to produce a transmission model of Lassa virus alone since there are no CMR studies of *M. natalensis* in the Lassa virus endemic area. Instead, CMR studies of populations which host other (non-zoonotic) arenaviruses provide an approximate framework for the transmission of the related Lassa virus. The vertical and horizontal transmission routes are informed by observed arenavirus dynamics in *M. natalensis* wherein vertically-infected individuals remain infectious for life, while horizontally-infected individuals are only transiently infectious and

recover after a period of up to 30 days (Borremans *et al.* 2015). The mass-action SHVR model (Figure 4.1) takes the form

$$\frac{\mathrm{d}N_S}{\mathrm{d}t} = b\big(N_S + N_R + (1-\alpha)(N_H + N_V)\big) - \beta N_S(N_H + N_V) - \mu N_S \quad (4.1)$$

$$\frac{\mathrm{d}N_H}{\mathrm{d}t} = \beta N_S(N_H + N_V) - \gamma N_H - \mu N_H \quad (4.2)$$

$$\frac{\mathrm{d}N_V}{\mathrm{d}t} = \alpha b(N_H + N_V) - \mu N_V \quad (4.3)$$

$$\frac{\mathrm{d}N_R}{\mathrm{d}t} = \gamma N_H - \mu N_R, \quad (4.4)$$

where β is the horizontal transmission rate, γ is the recovery rate following horizontal transmission, μ is the intrinsic mortality rate, b is the intrinsic birth rate, and α is the probability of vertical transmission between an infected individual and its offspring.

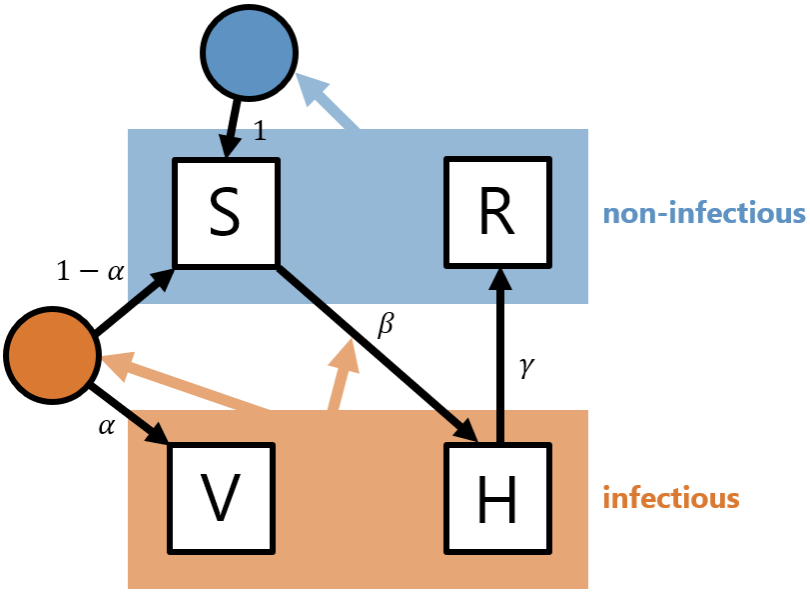


Figure 4.1: Diagram of the SHVR transmission model. Black squares represent compartments, and coloured backgrounds represent super-compartments. The blue and orange circles represent births from non-infectious and infectious individuals respectively. Black arrows represent movement between compartments, while coloured arrows indicate the contribution of super-compartments on transmission and birth. The blue non-infectious super-compartment contains compartments S (susceptible) and R (recovered and no longer infectious). Non-infectious individuals contribute to births which enter the susceptible compartment. The orange infectious super-compartment contains the compartments V (vertically-infected) and H (horizontally-infected). Vertically-infected rodents remain infectious for life, while horizontally-infected rodents recover at rate γ, after which they are no longer infectious. Infectious individuals infect susceptible individuals at rate β per year, and contribute to births which enter the susceptible compartment S with probability $1-\alpha$ and the vertically-infected compartment V with probability α.

Analysis performed using field studies of Lassa virus, Morogoro virus, and Gairo virus in *M. natalensis* found evidence of density-dependence (rather than frequency-dependence) (Mariën *et al.* 2020) which informed the infection term of the SHVR model. I did not incorporate a latent period between horizontal exposure and infectiousness because the latent period is believed to be only around 6 days (Borremans *et al.* 2015), making it short and therefore not very influential as well as being challenging to include in the IPM which has a 28-day time step. Although maternal antibodies from recovered mothers may protect some offspring for a period of up to around a month (Fichet-Calvet *et al.* 2014; Mariën *et al.* 2020), I did not include this in the model because it is not clear what proportion of recovered pregnant females pass antibodies vertically, or how long these antibodies persist in the wild.

M. natalensis arenavirus dynamics observed by Mariën *et al.* (2020) were best explained by a small number of chronically infected individuals and an increase in infected rodents due to horizontal infection following annual increases in population size. To verify that my SHVR model qualitatively matched these observations before incorporating transmission into the climatic-demographic model, I produced a theoretical simulation. I simulated equations 4.1–4.4 for a seasonally varying birth rate $b(t^*) = 10 + 5\sin(2\pi t^*)$ where t^* is in years, with intrinsic death rate set as $\mu = \bar{b} = 10$, using the forward Euler method with time step $\Delta t^* = {}^{1}/_{365}$ (one day). Initial conditions were $S = 600$, $H = 200$, $V = 100$, $R = 100$. Period before recovery following horizontal infection was set as 28 days, since this timescale would aid incorporation of the SHVR model into the IPM and is reasonable based on observed dynamics (Borremans *et al.* 2015). The transmission parameters were then recovery rate $\gamma = (28\text{ days})^{-1}$, probability of vertical transmission α was varied from 0.1 to 1, and infectiousness β was varied from 0 to 0.1. Simulated compartment sizes were then visualised for a choice of values representative of the range of simulations produced, informing selection of the transmission parameters in later IPM simulations.

Since infection by Lassa virus and other arenaviruses is not believed to impact – or has little impact – on the vital rates of *M. natalensis* (Mariën *et al.* 2017), I nested this compartmental transmission model within the climatic-demographic IPM without implementing any changes to demographic processes for infected rodents. I treated H as a transient compartment which individuals are only in for one time step (28 days). To implement the SHVR model into the IPM it only remained to estimate the number of horizontal infections between time steps. Local truncation error from a first-order numerical solution method for the nonlinear

SHVR model could be large if the time step of 28 days were used. A truncation error would mean that if the infected population were small, the force of infection would be greatly underestimated, delaying the resulting simulated disease dynamics. Therefore, I used the Runge-Kutta method to compute the estimated horizontal force of infection across each 28-day time step (Supplementary Text S4.3).

I denoted the distributions in each compartment using subscripts; for instance, D_S is the distribution of weights in the susceptible compartment. The complete model then took the form of the system of equations 4.5–4.8.

$$Z_S(w|y,t) = p_G(w|y,t)p_S(y,t)D_S(y,t) \tag{4.5a}$$

$$Z_H(w|y,t) = p_G(w|y,t)p_S(y,t)D_H(y,t) \tag{4.5b}$$

$$Z_V(w|y,t) = p_G(w|y,t)p_S(y,t)D_V(y,t) \tag{4.5c}$$

$$Z_R(w|y,t) = p_G(w|y,t)p_S(y,t)D_R(y,t) \tag{4.5d}$$

$$B_N(t+1) = \int_0^\infty lp_S(y,t)(J_S(y,t) + J_R(y,t))\,\mathrm{d}y \tag{4.6a}$$

$$B_I(t+1) = \int_0^\infty lp_S(y,t)(J_H(y,t) + J_V(y,t))\,\mathrm{d}y \tag{4.6b}$$

$$D_S(w,t+1) = \int_0^\infty \big(1 - F(t)\big)Z_S(w|y,t)\,\mathrm{d}y + \big(b_S(t+1) + (1-\alpha)b_I(t+1)\big)v(w) \tag{4.7a}$$

$$D_H(w,t+1) = \int_0^\infty F(t)Z_S(w|y,t)\,\mathrm{d}y \tag{4.7b}$$

$$D_V(w,t+1) = \int_0^\infty Z_V(w|y,t)\,\mathrm{d}y + \alpha b_I(t+1)v(w) \tag{4.7c}$$

$$D_R(w,t+1) = \int_0^\infty Z_R(w|y,t) + Z_H(w|y,t)\;\mathrm{d}y \tag{4.7d}$$

$$J_S(w,t+1) = \int_0^\infty p_P(y,t)\big(1 - F(t)\big)\big(Z_S(w|y,t)\big)\,\mathrm{d}y \tag{4.8a}$$

$$J_H(w,t+1) = \int_0^\infty p_P(y,t)F(t)\big(Z_S(w|y,t)\big)\,\mathrm{d}y \tag{4.8b}$$

$$J_V(w,t+1) = \int_0^\infty p_P(y,t)\big(Z_V(w|y,t)\big)\,\mathrm{d}y \tag{4.8c}$$

$$J_R(w,t+1) = \int_0^\infty p_P(y,t)\big(Z_R(w|y,t) + Z_V(w|y,t)\big)\,\mathrm{d}y. \tag{4.8d}$$

Equations 4.5a–d define shorthand for the density of weights at time $t+1$ based on the application of body growth and survival processes to a particular compartment at time t. For example, $Z_S(w|y,t)$ is the distribution of weights w at time $t+1$ resulting from susceptible

individuals with weight y at time t. Equations 4.6a–d define B, the number of juveniles entering the population from mothers in the different compartments. Finally, equations 4.7a–d project forwards the distribution of weights D in the compartments S, H, V, and R, while equations 4.10a–d describe the distribution of weights J in the pregnant/lactating sub-compartments. I assumed a litter size of $l = 5$ females, based on observed litter sizes of wild *M. natalensis* (Leirs *et al.* 1993), while juvenile weights were distributed as in Chapter 3 (Table 3.1).

4.3.2 Application to Lassa fever case data in Nigeria

Data

For the purpose of estimating the Lassa fever incubation period distribution, I collated papers on nosocomial outbreaks. From the papers identified, data relating to each individual case were included if the cases (1) were confirmed to have Lassa fever through a laboratory test, (2) were believed to have been infected in a healthcare setting, (3) reported a finite and informative period during which exposure occurred, and (4) reported a finite period during which symptom onset occurred. A "finite and informative" period during which exposure occurred was defined as one with a total time window of between 0 and 30 days, since the incubation period is currently believed to have a maximum period of 21 days. For each included case ($n = 30$), I extracted study location, earliest and latest possible date of exposure, earliest and latest possible date of symptom onset, and fatality including date of fatality (if recorded) (Figure 4.2, Supplementary Table S4.1).

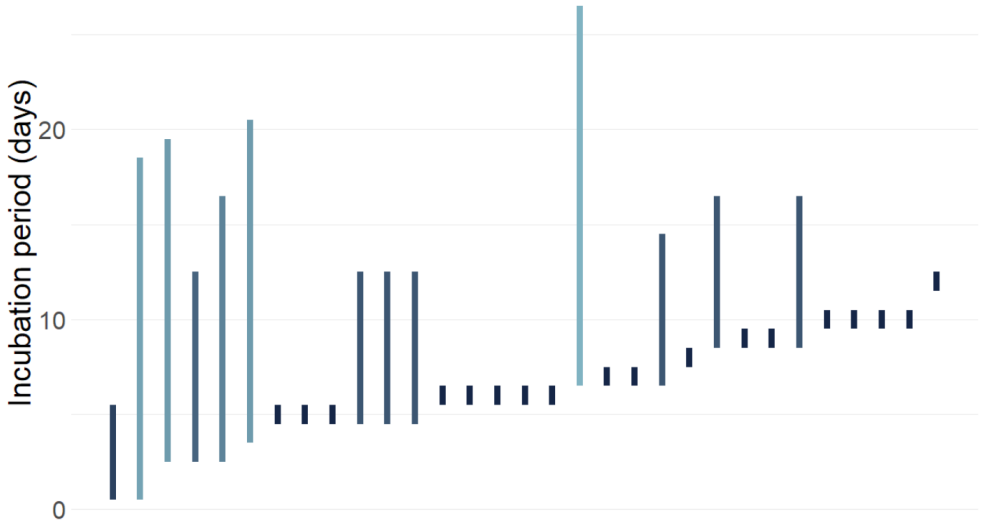


Figure 4.2: Inferred incubation period ranges for observed nosocomial infections. Lines show possible ranges of the incubation period for individual nosocomial cases of Lassa fever. Darker lines indicate a smaller possible range (Supplementary Table S4.1).

I obtained state-level Lassa fever weekly incidence data for the period 01/01/2018 to 31/12/2021 from the Nigeria Centre for Disease Control (NCDC) situational reports (NCDC 2022). I selected the states of Edo, Ondo, Ebonyi, Taraba, and Bauchi for analysis (Figure 4.2) since they had the highest reported confirmed Lassa fever case numbers in 2020, the year which appears to have the least severity bias (Chapter 2; Figure 2.2A and Supplementary Table S2.4). Therefore, these states should have temporal patterns of cases most representative of true Lassa fever incidence.

I obtained mean daily temperature (°C) and total daily precipitation (mm) for the period 2010 to 2021 from the National Oceanic and Atmospheric Administration Climate Prediction Centre (NOAA CPC) total daily precipitation and mean daily temperature global gridded datasets (NOAA/OAR/ESRL PSD 2021a, b). I then extracted these bilinearly at the centroid of each of the selected states (administrative boundary level 1) in Nigeria. For each of these quantities I computed a 28-day rolling average to reflect the temporal resolution of the climatic-demographic model, and then decomposed these quantities into seasonal and variable components as in Chapter 3 (Table 3.1) (Figure 4.3). The component of inter-annual variability in temperature was not needed since this was not present in the fitted climatic-demographic model (Supplementary Table 3.2).

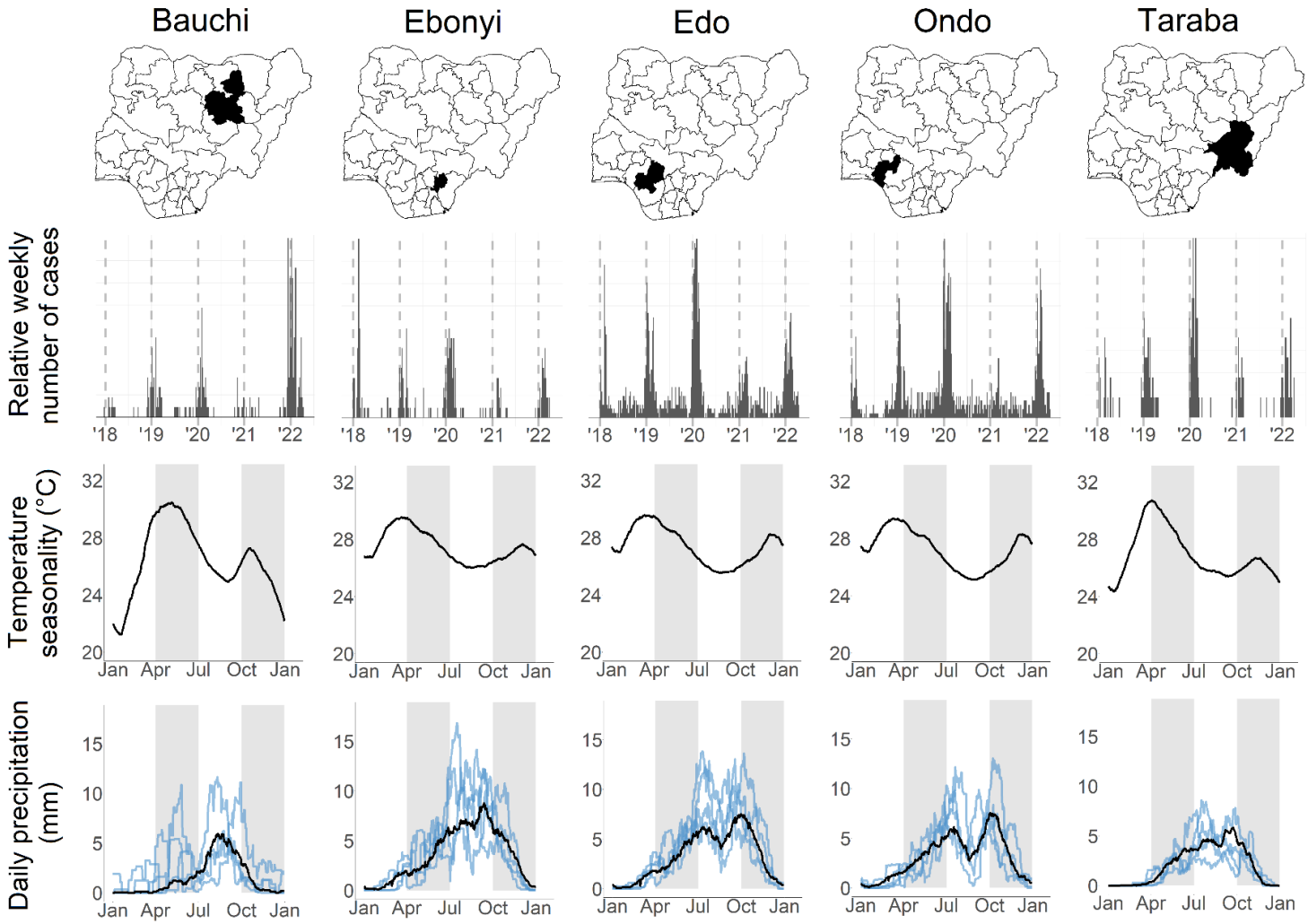


Figure 4.3: Weekly confirmed cases and seasonal and variable components of climate metrics for the states Bauchi, Ebonyi, Edo, Ondo, and Taraba in Nigeria. Top row: Outline of Nigeria and states of Nigeria with the state in question shaded. Second row: Weekly confirmed cases per state for the period January 2018 to April 2022 inclusive, obtained from NCDC Lassa fever situation reports. Vertical dashed lines indicate the first day of each year. Third row: Temperature seasonality (T_{seas}). Vertical shaded bars indicate three-month periods to aid readability. Bottom row: Precipitation seasonality (P_{seas}) shown with a black line, with the inter-annual variability around this seasonal pattern shown by the overlaid blue lines ($P_{seas} + P_{var}$).

Estimation of human Lassa fever incubation period

Zoonotic hazard and incidence of zoonotic disease are separated by several processes, with a potentially large time delay introduced by human incubation period before symptoms appear. Therefore, to link predicted zoonotic hazard from Lassa virus with observed Lassa fever cases, I estimated an incubation period distribution. No recent analysis has confirmed the commonly quoted incubation period of 3 to 21 days with a mean of 10 days (Guerrant *et al.* 2011), and this range does not provide a probabilistic distribution. I assumed that the incubation period I followed a discretised lognormal distribution, assuming a maximum incubation period of $N = 30$ days (equations 4.11–4.12), since the maximum possible incubation period from the collated data was 26 days. The distribution was discretised since the observed data recorded exposure period and symptom onset in discrete days, whereas the domain of a standard lognormal distribution is continuous.

$$p_I(d) = P(I = d) = \frac{F_I\left(d + \frac{1}{2}\right) - F_I\left(d - \frac{1}{2}\right)}{F_I\left(N + \frac{1}{2}\right) - F_I\left(\frac{1}{2}\right)} \tag{4.11}$$

$$F_I(d) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\log(d) - \mu_L}{\sigma_L\sqrt{2}}\right)\right) \tag{4.12}$$

I used a Bayesian approach so that uncertainty in model incubation period could be propagated forwards later in the full model linking predicted zoonotic hazard to observed disease. I estimated the log-mean μ_L and log-standard deviation σ_L of the distribution using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm based on the data collated from nosocomial outbreak reports, with fully uninformative prior distributions. Chains were inspected for convergence and initial values were varied to confirm that the resulting distributions were not dependent on initial conditions (initial μ_L was varied from $\log(2)$ to $\log(30)$ and initial σ_L was varied from 0.01 to 2). For the purpose of visualisation only, I used a kernel density estimate to produce a 95% credible region contour. To relate the result to existing knowledge of incubation period, I also estimated the predicted 95% central region, and the predicted probability of an incubation period being 21 days or fewer.

I performed sensitivity analysis by additionally fitting incubation period as a discretised gamma distribution, estimating shape parameter α_I and rate parameter β_I (equation 4.13), where subscripts are used to differentiate these incubation period parameters from the

transmission parameters α and β. (Here, γ denotes the incomplete gamma function such that F_I in equation 4.13 is the cumulative distribution function for the gamma distribution).

$$F_I(d) = \frac{\gamma(\alpha_I, \beta_I d)}{\Gamma(\alpha_I)} \tag{4.13}$$

Simulation of zoonotic hazard and comparison with Lassa fever incidence in Nigeria

As in Chapter 3, the full model of *M. natalensis* dynamics was simulated on a grid with time step of 28 days and body weight step of 1 gram, with the integrals discretised into sums and probability density functions into probability mass functions (Chapter 3; equations 3.16–3.19). The initial population size in each compartment was 6000 (S), 2000 (H), 2000 (V), and 0 (R) with an initial distribution of weights chosen by randomly sampling across the Morogoro CMR data (Chapter 3). Again, the simulations were only of females, assuming that there would be enough males in the population for recruitment to occur. A burn-in period of 24 time steps was used, and the recapture modifier parameter within the survival probability function (κ seen in Chapter 3) was estimated for each state by calibrating the model to a stipulated (arbitrary) mean population size of 10,000. Transmission parameters α and β were chosen based on the theoretical SHVR model simulations, with β scaled based on the ratio of mean population size in the theoretical SHVR simulation and the climatic-demographic simulation. The climatic inputs were the temperature and precipitation metrics for the different selected states in Nigeria. From the resulting simulations I computed the infectious reservoir host population size $X(t) = \sum_w (D_H(w,t) + D_V(w,t))$ at each time step t in the simulation. For each state, I then computed the relative infectious host population size at simulation points by dividing the predicted infectious host population size by the maximum observed infectious host population size in the simulation. From here on, this quantity will be referred to as zoonotic hazard. For the purposes of comparison with human disease data, I used linear interpolation to estimate the zoonotic hazard between simulation points. This temporally re-scaled zoonotic hazard is denoted $H_k(d)$, where d is time measured in days and k is the state.

Let $C_k(T)$ denote the number of observed cases in state k during epi-week T. For simplicity, I assumed that cases were uniformly distributed throughout the epi-week, and so divided this value by 7 and assigned the daily number of cases to each day of the epi-week. I defined the estimated number of cases on day d in state k by $Y_k(d)$. Using the incubation period distribution to infer date of infection, I transformed the daily cases Y_k into infection incidence

I_k (equation 4.13). I then tested for a relationship between daily zoonotic hazard H_k and daily human infection incidence I_k by fitting a GLM with a Gaussian error structure (equation 4.14) for each day d in the period 2018 to 2021 inclusive.

$$I_k(d) = \sum_{i=1}^{30} Y_k(d+i)p_I(i) \tag{4.13}$$

$$I_k(d) \sim \mathrm{N}(\mu = a_k + b_k H_k(d), \sigma^2 = {\sigma_k}^2) \tag{4.14}$$

I fit the linear model using the Metropolis-Hastings MCMC algorithm with completely uninformative (uniform) prior distributions, estimating a chain of values for a_k, b_k, and σ_k. For each state k I performed this analysis for 1000 points in (μ_L, σ_L)-space drawn from the incubation period distribution Markov chain, thereby obtaining a posterior distribution for (a_k, b_k, σ_k) incorporating uncertainty from the incubation period model. Using the posterior distributions, I estimated 95% credible intervals for the parameters and computed the p-value for $b_k > 0$ to test whether human infection incidence I_k was significantly positively correlated with predicted zoonotic hazard H_k (significance level $\alpha = 0.05$).

4.4 Results

4.4.1 Developing a process-based model of climate-driven zoonotic hazard

Numerical solutions of the theoretical SHVR model with a seasonally forced birth rate showed how transmission parameters could affect timing and size of peaks in the infectious population size, compartments H and V (horizontally-infected and vertically-infected, respectively) (Figure 4.4). When comparing these theoretical results with the inference of arenavirus transmission among *M. natalensis* (Mariën *et al.* 2020), a smaller horizontal transmissibility β appears to reflect dynamics in which infection is sustained by a low-level vertically-infected, chronically infectious population V between breeding seasons, and is amplified by horizontal transmission during population booms. Additionally, peaks in seroprevalence appear to be delayed following peaks in population size, similar to analysis of observed Morogoro virus dynamics in Tanzania (Mariën *et al.* 2020), as opposed to model simulations with larger values of β in which seroprevalence peaks almost simultaneously with population size. Therefore, when incorporating the SHVR transmission dynamics into the climatic-demographic framework of *M. natalensis*, I assumed the values $\beta = 0.008\ (\mathrm{year})^{-1}$ (which would need to be scaled inversely to average population size in the climatic-demographic simulation) and $\alpha = 0.6$.

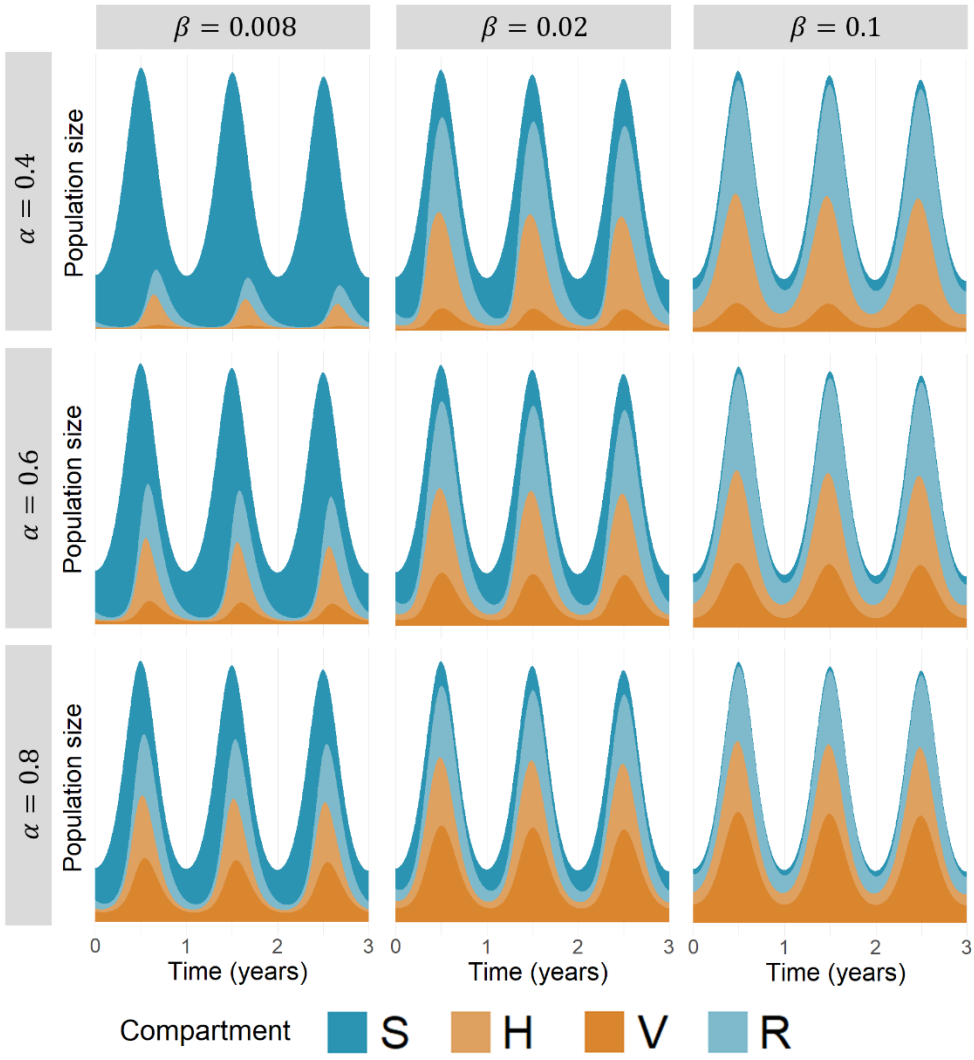


Figure 4.4: Numerical solutions to the theoretical SHVR transmission model with seasonal population fluctuations. Increasing probability of vertical transmission, α, (lowest in top row, highest in bottom row) created a larger infectious pool, increasing the number of both horizontally- and vertically-infected rodents throughout. Increasing horizontal transmission rate, β, (lowest in left column, highest in right column) made the non-susceptible population increase more quickly during population size increases. Compartment sizes are stacked vertically on top of each other.

4.4.2 Application to Lassa fever case data in Nigeria

Estimation of human Lassa fever incubation period

From the generated chain of incubation period distribution parameter values, log-mean μ_L was estimated as 2.02 (95% CrI [1.89, 2.15]) and the estimate of log-standard deviation σ_L was 0.301 [0.199, 0.441] (Figure 4.4A). The estimated mean, median, and mode incubation period were 7.82 [6.85, 9.12] days, 7 [6, 8] days, and 7 [6, 8] days respectively. For the incubation period distribution predicted by μ_L and σ_L, the central 95% incubation period range was from 3 [2, 4] days to 13 [11, 18] days, and the probability of an incubation period

being less than or equal to 21 days was 99.98% [99.14%, 100%], indicating that the 21-day duration of contact tracing employed by the NCDC (NCDC 2018) is appropriate. However, the average estimated incubation periods were significantly lower than the commonly quoted mean of 10 days (Guerrant *et al.* 2011).

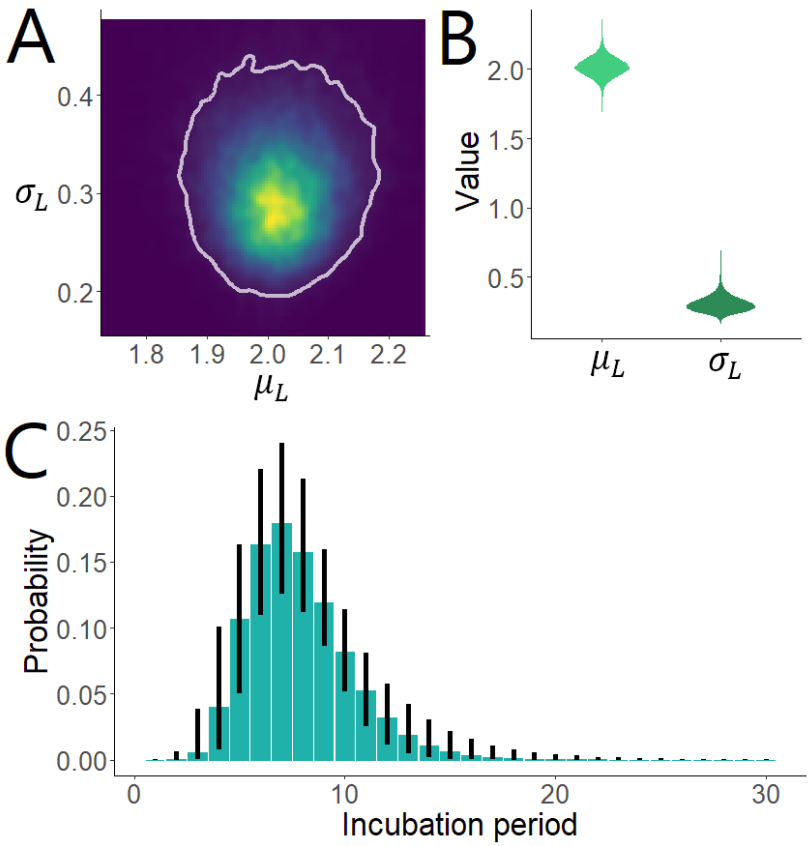


Figure 4.5: Credible region for the lognormally distributed incubation period, parameter estimates, and predicted distribution with credible intervals. A: Kernel density estimate for the density in the MCMC chain across parameter space, with lighter colours representing higher density and 95% credible region enclosed by a white line. B: Kernel density estimates for the values of the log-mean and log-standard deviation. C: Bar graph showing predicted probability of an incubation period being the given duration. Each black line range indicates the 95% credible interval for the probability mass function of that incubation period duration in isolation.

In the sensitivity analysis (Supplementary Figure S4.3), the estimated mean, median, and mode incubation period following the gamma distribution were 7.79 [6.84, 8.82] days, 7 [6, 8] days, and 7 [6, 8] days respectively. Other derived metrics also indicated that the incubation period distribution following a gamma distribution was similar to that following the lognormal distribution, indicating a robust model (Supplementary Table S4.4).

Simulation of zoonotic hazard and comparison with Lassa fever incidence in Nigeria

Since the mean rodent population size in the SHVR model theoretical simulation was 2400, and the mean rodent population size in the IPM simulations were set to 10,000, I multiplied the chosen horizontal infectiousness parameter β by a factor of 2400/10000, giving $\beta = 0.00192$ and $\alpha = 0.6$ as the selected transmission parameters in the full climatic-demographic transmission model. The model predicted population and transmission dynamics across 2018 to 2021 for each of the four states of Nigeria (Figure 4.6).

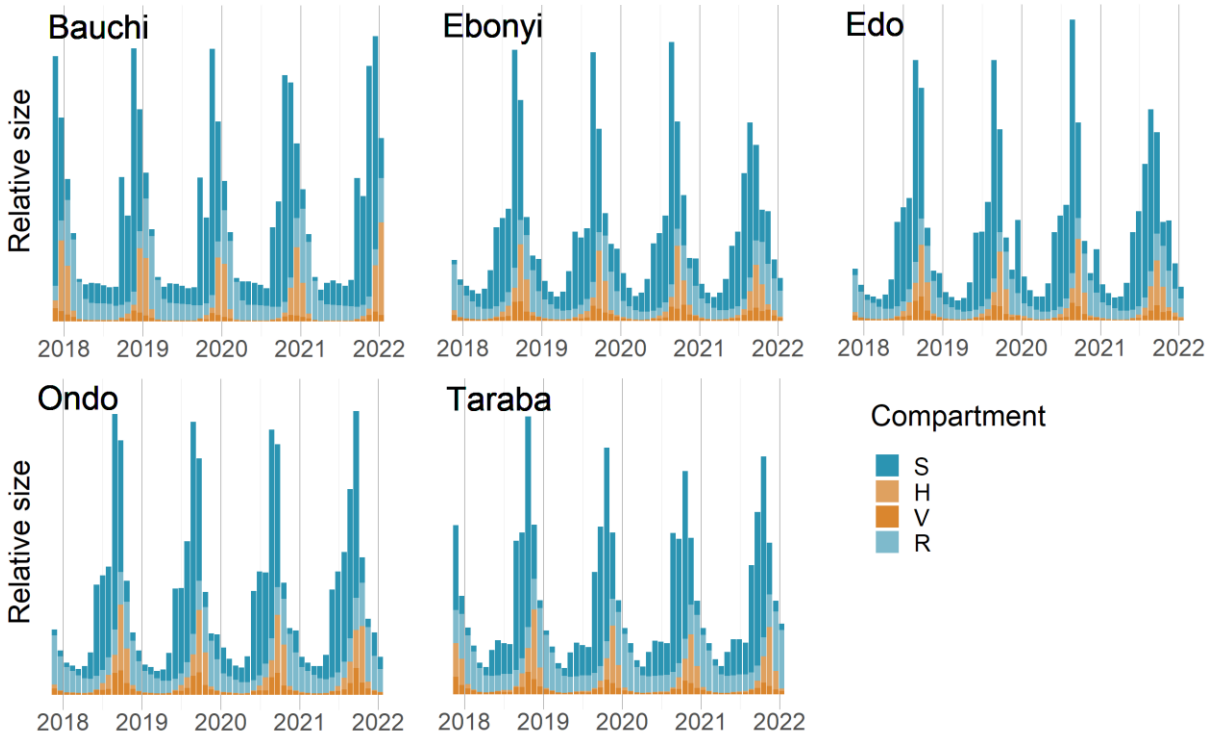


Figure 4.6: Predicted population and transmission dynamics of M. natalensis *in Nigeria.* Stacked bar plots show simulations of the climatic-demographic model with nested Lassa virus transmission dynamics in the five study states of Nigeria. Each bar is one time step of the model (every 28 days). Each distinct colour represents a compartment; S (susceptible) in dark blue, H (horizontally infected) in light orange, V (vertically infected) in dark orange, and R (recovered) in light blue. The orange colours combined represent infectious rodents and therefore the orange region of the graphs are predictions of temporal zoonotic hazard.

Inferred infection incidence was significantly positively correlated ($b > 0$) with predicted zoonotic hazard for all states (Bauchi: $p = 0.0003$; Ebonyi, Edo & Ondo: $p < 0.0001$; Taraba: $p = 0.0243$) (Supplementary Table S4.5). However, visual assessment of model fit (Figure 4.7) indicates a delay between predicted peaks in zoonotic hazard and peaks in inferred human infection incidence, especially for the southern states of Ebonyi, Edo, and Ondo.

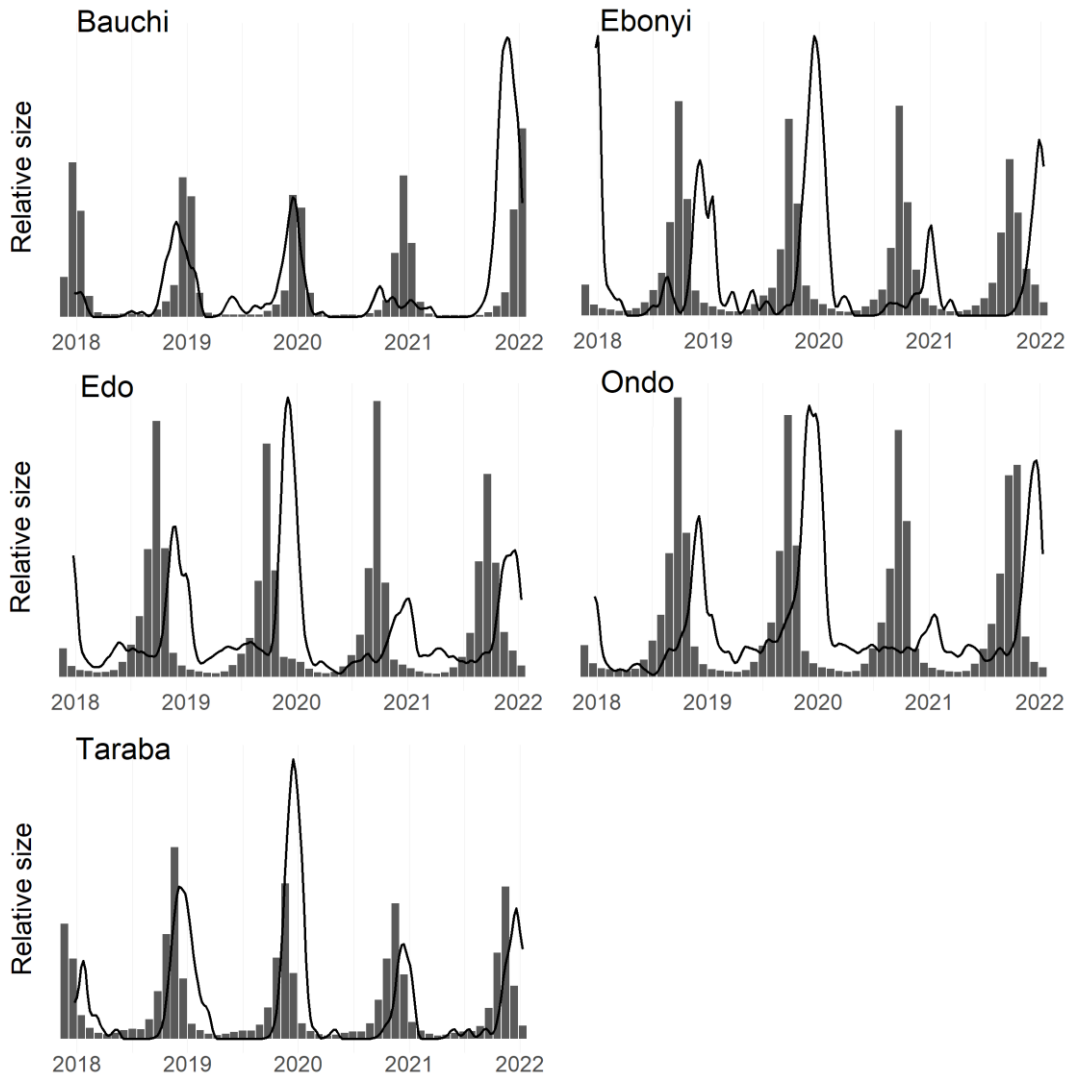


Figure 4.7: Predicted Lassa virus zoonotic hazard and inferred human infection incidence. Grey bar plots show simulations of zoonotic hazard (abundance of infectious hosts) from the climatic-demographic model with nested Lassa virus transmission dynamics in the five study states of Nigeria. Each bar is one time step of the model (every 28 days). Black lines show inferred daily incidence of infection in humans in each state, computed by transforming observed weekly cases of Lassa fever using the incubation period distribution.

4.5 Discussion

Here, I described a method to link climatic-demographic host processes to observations of human zoonotic disease through host-pathogen dynamics and incubation period in humans. First, I incorporated pathogen transmission into the climatic-demographic model from Chapter 3, informed by observations of arenavirus dynamics in *M. natalensis*. Next, I estimated human Lassa fever incubation period distribution from data relating to nosocomial cases and used this distribution to infer patterns of human Lassa virus infection from weekly case data in five states in Nigeria. Finally, I simulated zoonotic hazard (relative abundance of

infectious hosts) using the climatic-demographic model with nested pathogen dynamics and, using a GLM, modelled the relationship between predicted zoonotic hazard and inferred human infection. I found that predicted zoonotic hazard was a significant and positive factor in patterns of inferred human infection for all five states, but for the southern states of Edo, Ondo, and Ebonyi, visual inspection showed that an additional time lag of one to two months separated peaks in predicted zoonotic hazard and peaks in inferred human infection.

My analysis of the Lassa fever incubation period produced the first estimates of incubation period distribution from collated nosocomial case data. I fitted a lognormal and gamma distribution, finding that both similarly described incubation period distribution from the case data ($n = 31$). My finding that the probability of an incubation period being less than or equal to 21 days was 99.98% (95% CrI [99.14%, 100%]) indicates that the 21-day duration of contact tracing used by the NCDC (NCDC 2018) is appropriate. In my estimated incubation period distribution, the central 95% incubation period range was from 3 [2, 4] days to 13 [11, 18] days indicated that the commonly quoted range of 3 to 21 days (Guerrant *et al.* 2011) may be too high at the upper end. The estimated mean incubation period was 7.82 [6.85, 9.12] days, significantly lower than the commonly quoted mean of 10 days (Guerrant *et al.* 2011). These findings imply that the incubation period of Lassa fever may be lower than is currently clinically advised. However, these findings are limited by the scope of the data which were restricted to apparently nosocomial cases of Lassa fever and not cases caused by zoonotic spillover. Due to the potential for within-human evolution of Lassa virus, it is possible that the nosocomial incubation period is longer or shorter than the incubation period following infection by a zoonotic host. It is also possible that patients included in the nosocomial data could have been infected by a zoonotic host prior to being admitted to hospital, thereby shortening the apparent incubation period.

In this chapter I developed a model for rodent arenavirus transmission dynamics informed by observations of Morogoro virus, Gairo virus, and Lassa virus prevalence in wild rodents and by climate-driven demography of *M. natalensis* (Chapter 3). This model differs from existing models of *M. natalensis* arenavirus transmission dynamics (Mariën *et al.* 2019) because my climatic-demographic model explicitly models seasonal rodent body growth, survival, and pregnancy probabilities dependent on body weight, climate seasonality and inter-annual variability in precipitation. The model of Mariën *et al.* (2019) found through model simulation that annual rodent control in Guinea would be insufficient to substantially reduce abundance

of Lassa virus infected rodents. While Mariën *et al.* (2019) assumed a constant total rodent population size in the absence of control (informed by observations in Guinea), my climatic-demographic model (informed by a CMR study in Morogoro, Tanzania) simulated rodent population size as seasonally and inter-annually variable. For areas in the Lassa virus endemic region in which *M. natalensis* dynamics are thought to undergo temporal fluctuation (Fichet-Calvet *et al.* 2007), it would therefore be valuable to explore rodent control through the in silico experimental environment provided by my full climatic-demographic arenavirus transmission model.

My approach made use of existing ecological studies of a species instead of requiring in-situ studies for the chosen locations of interest in Nigeria. However, it is important to note that this modelled zoonotic hazard is predicted from observed climate-driven demographic dynamics of a population of *M. natalensis* in an agricultural landscape in Tanzania, and not in Nigeria. Differing habitat types likely lead to differing population dynamics (Fichet-Calvet *et al.* 2007), and given that Tanzania and Nigeria have different lineages of *M. natalensis* (Olayemi *et al.* 2016b), it is possible that – though unknown whether – these populations also have phenotypic differences. Still, given the ethical and health and safety implications of carrying out capture-release studies on an animal which carries a zoonotic pathogen, in some cases this translocation of a species' demographic processes as carried out here may be the best or only way to gain demographic insights into known hosts, and could reduce the number of such studies needed. It would be challenging to assess the performance of the climatic-demographic model in predicting population changes in Nigeria due to the ethical considerations of collecting suitable data.

Modelled temporal patterns of zoonotic hazard in all five states of Nigeria were significant and positive correlates in explaining patterns of human cases, after accounting for a delay before observation of symptoms due to incubation period. Visual comparison of the model with inferred infection indicated that patterns of predicted zoonotic hazard aligned well with timing of peak human infection in the more northern states of Bauchi and Taraba, but that an additional time lag of one to two months exists between zoonotic hazard and human infection in the southern states of Edo, Ondo, and Ebonyi. If modelled zoonotic hazard is a good indication of abundance of infected reservoir hosts in Nigeria which – as previously discussed – is not necessarily true, then this time lag could indicate one or more of a number of additional processes. For instance, if probability of human infection by rodents was

frequency- instead of density-dependent, this would explain at least some of the observed lag since simulated peak predicted abundance of infectious rodents followed peak in total abundance with a delay. It is also likely that agricultural practices (Sluydts *et al.* 2009) or resource-seeking behaviour (Fichet-Calvet *et al.* 2007; Bonwitt *et al.* 2017) could drive the entering of *M. natalensis* into homes. From the viewpoint of the observed time delay representing a delay between peak zoonotic hazard and peak zoonotic exposure, this analysis suggests that further quantification of temporal patterns in rodent behaviour and incorporation of these patterns into a model such as this could help explain and predict patterns of Lassa fever incidence.

The integrated disease model presented here represents a step forward in process-based modelling of temporal zoonotic disease dynamics, and also how these dynamics might vary across space due to differing climate patterns. Temporal models of Lassa fever and Lassa virus infection have previously been limited to incorporating reservoir host dynamics through theoretical simulation (Akhmetzhanov *et al.* 2019), or have not included these dynamics at all (Musa *et al.* 2020; Barua *et al.* 2021; Ibrahim & Dénes 2021; Abidemi *et al.* 2022). With the changing climate likely to impact the dynamics of *M. natalensis,* Lassa virus prevalence in the reservoir host, and risk of spillover to humans, by quantifying the dependence of *M. natalensis* and arenavirus transmission dynamics on the climate I have created an opportunity to explore this complex system. My model is a promising step towards human disease forecasting and can predict rodent infectious disease dynamics two months in advance due to climate lags. While there is scope to develop an early warning system for the timing and magnitude of human Lassa virus infection through this model, the disparities in timing between predicted zoonotic hazard and inferred human infection incidence are consistent with a role of other factors, such as agricultural practices, in determining timing of risky human-rodent contact and human infection. Host-pathogen ecology – including transmission dynamics, movement, behaviour, and pathogen shedding – should be a clear focus when seeking to understand patterns and drivers of zoonotic disease, with the methods developed here enabling the modelling of several aspects of this system.

Chapter 5: Discussion

5.1 Summary of findings and implications

In this thesis, I sought to extend understanding of ecological and environmental drivers of the Lassa fever by developing methods to describe the processes underpinning the chain of Lassa virus transmission. Importantly, these methods were context-specific – aiming to capture the necessary complexity to explore key processes in the Lassa virus system – but also adaptable to other zoonotic disease systems, thereby presenting methodological advancements which can be used to explore drivers of other zoonoses.

I first asked (Chapter 2) whether mathematical modelling to leverage zoonotic niche models and different data sources could provide insights into the absolute incidence of Lassa virus infection and disease. Until recently, infection incidence has been underestimated due to the use of outdated and spatially limited serosurveys which predict incidence prior to substantial increases in the population of West Africa (McCormick *et al.* 1987). Additionally, Lassa fever case data are believed to suffer from under-ascertainment (Arruda *et al.* 2021), with the number of cases observed in each country in its endemic range far fewer than the likely disease incidence. Analysing publicly available case data from the Nigeria Centre for Disease Control (NCDC) (NCDC 2022), I identified for the first time a severity bias which supports the theory of under-ascertainment and spatial heterogeneity in surveillance. While this severity bias persists throughout the years 2018 to 2021, there is evidence that the bias has reduced, adding to the observation of Redding *et al.* (2021) that surveillance in Nigeria measurably improved between 2012 and 2019 (Supplementary Figure S5.1). The most conservative estimates from my model of total infection incidence suggest between 600,000 and 4.5 million symptomatic infections of Lassa fever per year, implying 30,000 to 225,000 annual deaths due to Lassa fever. Model estimates varied substantially depending on the choice of spatial model of zoonotic hazard and exposure, with other models predicting many more infections.

Until now, the only estimates of Lassa virus infection or disease incidence have been based on serosurvey data (McCormick *et al.* 1987; Basinski *et al.* 2021). However, current serosurvey data are limited in value due to a lack of knowledge around seroreversion and underrepresentation of much of West Africa. The results of Chapter 2 show a potential to use other health data in addition to serosurveys to contrast incidence produced through

different approaches and thus increase result robustness. To maximise the value of serosurveys, uncertainties in my findings indicate that longitudinal serosurveys to identify whether seroreversion of Lassa virus-specific antibodies occurs, and serosurveys targeted at new locations to reduce the clustering and bias of existing data, would be valuable in clarifying the overall incidence of Lassa virus infection and the extent of spatial heterogeneity across West Africa. Development of longitudinal health surveys in a similar format to the contact tracing protocol (NCDC 2018), for example testing anyone in a given study population who presents with a fever, may provide an opportunity to bridge the gap between indicators of disease and infection incidence. This would provide a valuable dataset against which to verify incidence estimates based on serosurvey data and, if stratified across West Africa, could provide additional insight into spatial heterogeneity in Lassa fever risk. Given my new estimate that Lassa virus infection incidence is on the scale of millions (at least 3 million), rather than hundreds of thousands (between 100,000 and 300,000) as previously estimated, unreported disease incidence across much of West Africa potentially represents a very large unrecognised disease burden. Given that the goal of public health is to reduce disease burden, …

Next (Chapter 3) I began to explore the dynamics of the Lassa virus system from the bottom-up, starting with the reservoir host, *Mastomys natalensis*. Since rodents are highly seasonal and often opportunistic, their demography (here, recruitment rate, survival, and body growth), and how demographic processes are affected by environmental factors, are important in determining resulting population dynamics (Ostfeld & Mills 2007). Studying demographic processes requires longitudinal individual-based field studies, which are not appropriate in West Africa where the human disease risk posed by *M. natalensis* makes it unethical to capture and release rodents. Therefore, I studied the demographic processes of *M. natalensis* from a long-term capture-mark-recapture (CMR) study in Morogoro, Tanzania, outside of the Lassa virus endemic range. Modelling observed proxies for recruitment, survival, and body growth as functions of lagged precipitation and temperature, I found that population demography of *M. natalensis* is highly seasonal. Further, by decomposing climate variables into components of seasonality and inter-annual variability, I was able to identify that precipitation may drive survival and recruitment rates beyond seasonality. The most important effect of inter-annual variability in precipitation was on recruitment, with periods in which rainfall is greater than the seasonal average leading to periods with significantly and

substantially greater recruitment rates than the seasonal average. Finally, I constructed an integral projection model (IPM) to describe and simulate the population dynamics of *M. natalensis*, with precipitation and temperature as inputs.

The results of Chapter 3 indicated that years with higher-than-average precipitation may have an increased population size of *M. natalensis*. While previous studies have linked climatic variables with *M. natalensis* demographic processes (Leirs *et al.* 1990; Christensen 1993; Fichet-Calvet *et al.* 2007; Mulungu *et al.* 2016; Mayamba *et al.* 2021), this is the first time the effects of lagged seasonal and inter-annually variable temperature and precipitation on these processes have been separately quantified. The resulting climatic-demographic IPM model provides a tool through which to explore determinants of *M. natalensis* population dynamics, and on which to build a process-based model for Lassa virus transmission dynamics in the reservoir host.

In Chapter 4 I investigated how to model viral transmission dynamics in a theoretical *M. natalensis* population. Based upon field studies of Lassa virus and other arenaviruses, I defined a susceptible-horizontally infected-vertically infected-recovered (SHVR) compartmental transmission model. By creating compartmental distributions of body weights in the IPM, I was able to incorporate the SHVR arenavirus transmission model into the climatic-demographic model. Using documented nosocomial cases of Lassa fever, I estimated a probability distribution for the incubation period of Lassa virus in humans and used this distribution to transform patterns of observed Lassa fever cases in Nigeria into inferred infection incidence. I then modelled the relationship between predicted zoonotic hazard and inferred Lassa virus infection using a generalised linear model (GLM). I found that predicted zoonotic hazard was a significant and positive correlate in the model of inferred infection incidence in all five of the studied states (chosen as the states with the largest number of observed cases in 2020). However, visual inspection of predicted zoonotic hazard compared with inferred infection incidence indicated an additional time delay of around one to two months.

The work of Chapter 4 demonstrates that it is possible to integrate a context-specific disease transmission model into a climate-driven reservoir host population model to predict infectious population size based on demographic processes rather than mass-action assumptions. The results imply that even though the climatic-demographic model was fit to a population of *M. natalensis* in Tanzania, fitting of demographic processes to seasonal and

inter-annually variable climate produced a feasible estimate of zoonotic hazard in Nigeria where Lassa fever is endemic but where *M. natalensis* cannot be studied in this way. While the relationship between Lassa fever incidence and rainfall has already been modelled by Redding *et al.* (2021), this new process-based model of the reservoir host provides an opportunity to explore zoonotic hazard through host ecological processes. The resulting temporal climate-driven reservoir host transmission model can be used to forecast zoonotic hazard several months in advance and could be used to predict the effects of future changes to climate patterns on human risk of Lassa fever.

5.2 Limitations and outstanding questions

In Chapter 2, I utilised a variety of spatial models of Lassa virus zoonotic hazard and exposure to estimate the absolute incidence of infection and disease. Uncertainty is inherent but challenging to quantify for these models, therefore by varying the fundamental assumptions I introduced some of this uncertainty into my estimates of incidence. The substantial differences in results from these models indicate that the uncertainty in spatial zoonotic hazard – due to unknown spatial distribution of Lassa virus and the reservoir host – and exposure – due to the unknown scaling of contact with human and rodent population density and – leads to a large amount of uncertainty in estimated absolute disease incidence. While this uncertainty limits the work of Chapter 2 in terms of reducing the precision of disease incidence estimation, it is also an important conclusion providing a benchmark for the state of spatial zoonotic hazard and exposure modelling. The large uncertainty caused by differing assumptions implies that more work is needed to verify the spatial structure and components underpinning risk of zoonotic spillover, and that such uncertainties should be included in future spatial models of Lassa fever rather than assuming one "correct" spatial zoonotic hazard and exposure model.

The infection incidence estimate in Chapter 2 was also limited by two further types of uncertainty, which are important to characterise and quantify when interpreting results. The assumed rate of seroreversion leads to differing relationships between force of infection and seroprevalence; therefore, uncertainty in the occurrence and rate of seroreversion is an important determinant of uncertainty in the Lassa virus infection incidence estimate. To an extent I quantified this uncertainty by producing estimates assuming that no seroreversion occurred, and assuming a rate of seroreversion informed by findings that seroreversion occurs at a rate of 6.4 per 100 seropositive individuals per year (McCormick *et al.* 1987).

However, this seroreversion estimate did not include reported uncertainty, potentially limiting interpretations of the finding. This lack of reported uncertainty and the fact that populations have likely substantially changed since the dates of the serosurveys analysed suggest that further investigation into seroreversion would be of value. Further uncertainty originated from the component model of Basinski *et al.* (2021) which predicts human seroprevalence from modelled zoonotic hazard, but as the spatial human seroprevalence model did not incorporate uncertainty, it was not possible to quantify the resulting uncertainty in my estimates of infection risk. In future, it would be valuable to generate spatial human seroprevalence models sampled from the underlying model distribution in order to better quantify uncertainty.

The main limitation of the climatic-demographic model constructed in Chapter 3 and applied in Chapter 4 is the scope of the CMR study used. Long-term CMR is logistically challenging and resource-intensive, leading to trade-offs in study design. The scope of the Morogoro CMR study was primarily agricultural mosaic land, providing opportunities to assess effects of agricultural practices on *M. natalensis* as a pest (Massawe *et al.* 2005; Sluydts *et al.* 2009) in addition to observing demographic processes (Sluydts *et al.* 2007) and later exploring Morogoro virus transmission (Mariën *et al.* 2020). Therefore, the data used in this thesis were limited to agricultural land and did not include forest, villages, urban areas, or other land cover types and uses, even though demographic processes could vary across these environments, as has been found in other animal species (Davison *et al.* 2019). While studies looking at *M. natalensis* population changes in other habitats do exist, none have a long enough temporal extent and resolution to study how population size could link to climate, especially not through demographic modelling of individual-based data. This limitation highlights the deficit of small rodent studies in sub-Saharan Africa adequate for studying climate-driven population dynamics in different habitat types. When answering questions about zoonotic hosts, the resource requirements of setting up and maintaining longitudinal studies – especially of wildlife – may outweigh the research benefits. However, for hosts which are also agricultural pests, such as *M. natalensis* and other rodents (Morand *et al.* 2015), studies designed to answer questions relating to both disease ecology and agricultural management could enable a more efficient use of resources. From a One Health perspective, such studies could inform measures to protect human health through both food security and disease mitigation. Indeed, the climatic-demographic model produced in

Chapter 3 is not only limited to zoonotic hosts but could be useful in describing climate-linked population dynamics of any rodent species, including agricultural pests.

In Chapter 4, when using predicted temporal zoonotic hazard in a model for Lassa fever risk, I held other processes between zoonotic hazard and infection incidence temporally constant. These intermediate processes include vulnerability (such as susceptibility to disease given infection), human and rodent behaviour which could lead to increased contact, and human-to-human disease transmission. The factor which appears most likely to be temporally variable is human and rodent contact, as it is possible that *M. natalensis* enter homes during the dry season in search of food (Fichet-Calvet *et al.* 2007). Since human-to-human transmission is believed to be rare (Siddle *et al.* 2018) and probability of human-to-human transmission is indirectly related to zoonotic spillover risk since most human cases will originate from spillover, I assessed neglecting human-to-human transmission to be a reasonable simplifying assumption. To limit the effects of simplifying assumptions, my process-based model of temporal Lassa fever risk should be used to estimate or forecast spillover risk rather than total infection or disease risk. If human-to-human transmission increases or is found to be more substantial, or if super-spreading events occur, this will no longer be the case. A simple compartmental human-to-human infection transmission model could be appended to my model of Lassa fever risk and could be used to test the effect of different levels of human-to-human transmission on disease dynamics.

A major new direction made methodologically possible through the work of Chapters 3 and 4 is the prediction of patterns and trends in Lassa virus spillover risk under future changes to climate patterns. It is believed that precipitation will decrease in the eastern Sahel (north of the Mano River Union) and increase in the western Sahel (north of Nigeria) (Gaetani *et al.* 2020), potentially implying that parts of the Lassa virus endemic area will also experience changes in precipitation. These changes could alter seasonal patterns and variability in *M. natalensis* outbreaks between years (Chapter 3), meaning that some future years could potentially experience larger population outbreaks and resulting Lassa fever outbreaks than previously seen. It would be of great value to apply the model of Chapter 4 to possible scenarios of intra- and inter-annual weather patterns; however, such models have large inherent uncertainty and stochasticity and no such models have been produced for West Africa. If, as some models suggest, climatic seasonality also alters in the future, it could be imagined that this could alter the length of the breeding season or the speed of maturation

and drastically alter population dynamics. If climate seasonality is found to significantly change in the future, I suggest that the climatic-demographic model be re-fit under the new climate regime to maintain the distinction between rodent phenology and responses to inter-annual climatic variability.

While this thesis focussed on a case study of Lassa fever, the tools developed can be applied to other zoonotic disease systems. Since the climatic-demographic model is more appropriate for rodent-like small mammals than, for example, predator species, large animals, and species with complex movement ecology such as birds and bats, it would be valuable to apply the approach to other zoonotic disease systems with rodent or rodent-like hosts. Different rodent species have different demographic population dynamics (Sluydts *et al.* 2009), therefore fitting the climatic-demographic model to other zoonotic hosts would yield similarly informative but disease-specific insights into climate-driven processes within zoonotic disease systems. One such system is Puumala hantavirus and its reservoir host in central Europe, the bank vole (*Myodes glareolus*). This system is well-studied in terms of estimated abundance and viral seroprevalence of the bank vole population and how these quantities relate to human infection risk (Reil *et al.* 2017). This evidence makes it possible to predict human infection risk based on observed bank vole population patterns. However, observed population dynamics have not yet been linked to climatic patterns. Applying the climatic-demographic approach to the bank vole could therefore provide an opportunity to improve understanding of the timing of bank vole population processes through demography, and to predict human infection risk further in advance from observed climate patterns and under different climate scenarios.

5.3 Spatial determinants of Lassa virus zoonotic hazard

The central region of West Africa, here defined as Côte d'Ivoire, Ghana, Togo, southern Burkina Faso, and southern Mali, appears to have a lower incidence of Lassa fever and Lassa virus infection as compared with eastern West Africa (Nigeria and Benin) and western West Africa (countries of the Mano River Union) (Figure 5.1). This geographic heterogeneity in observed cases has not been well-explained, which is symptomatic of a lack of understanding around the geographic distribution of Lassa fever. Limited understanding of the spatial structure of Lassa fever risk across West Africa is multi-faceted since geographic heterogeneity in observed cases could be a manifestation of one or more disease ecological processes, an artefact of surveillance and reporting processes, or a combination of both.

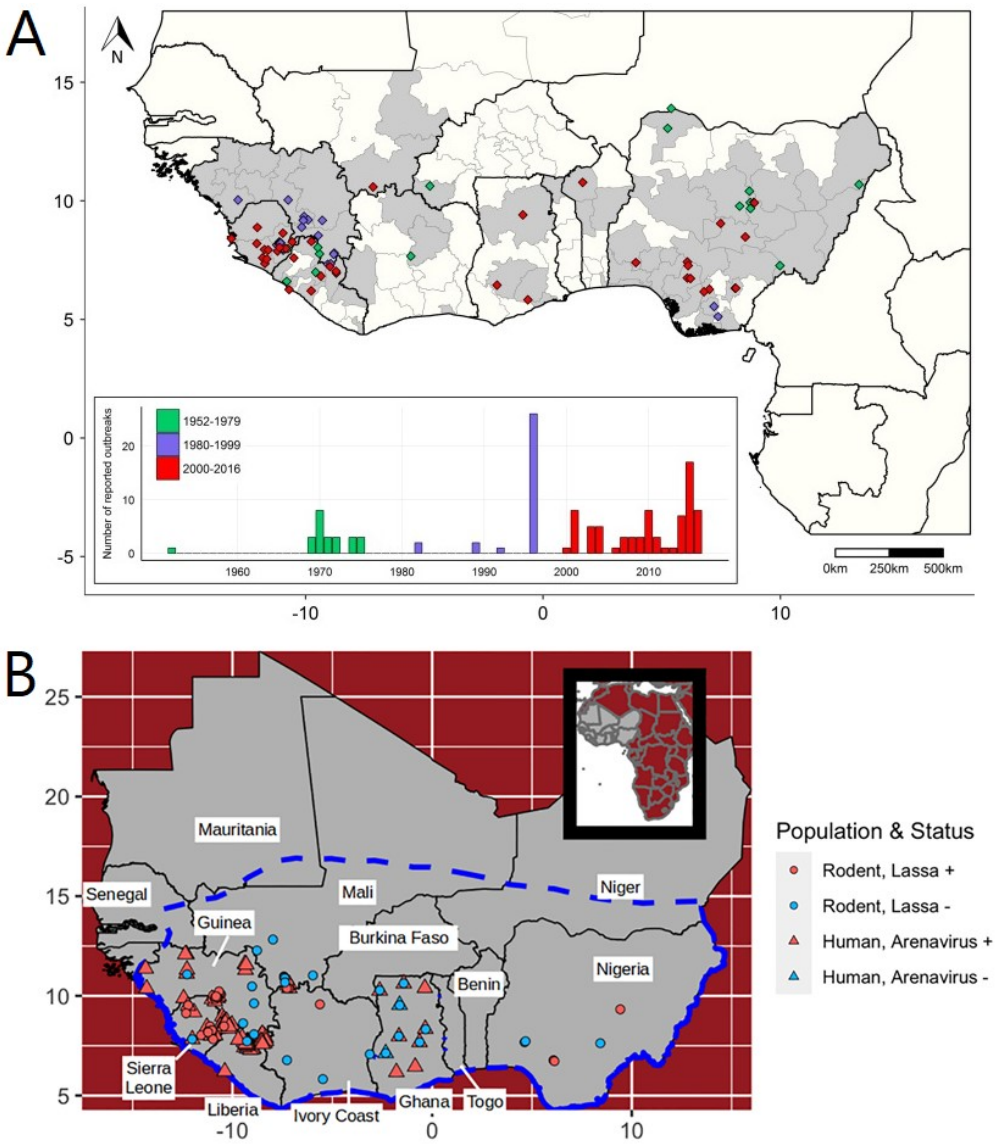


Figure 5.1: Spatial heterogeneity in observations of Lassa fever and Lassa virus infection in humans and rodents. Reporting has been biased towards the Mano River Union and Nigeria, with one seroprevalence study (Kronmann *et al.* 2013) representing the majority of data points in the central region of West Africa. A: Human Lassa fever outbreak reports in West Africa from 1952 to 2016, with each point representing a separately reported cluster of cases. Each point is coloured by time period in which the outbreak began, with the bar graph showing number of reported outbreaks beginning in each year. The grey lines indicate borders of local administrative areas with grey shaded areas indicating any evidence of Lassa virus or similar arenavirus in humans or rodents. Adapted from Gibb *et al.* (2017). © Informa UK Limited, trading as Taylor & Francis Group; 2017. Reproduced with permission. B: Collated rodent and human serosurveys. Circles denote rodent trapping studies which performed Lassa virus testing and are coloured red if at least one tested rodent was Lassa positive. Triangles denote human serosurveys and are coloured red if at least one tested human was seropositive for arenavirus antibodies. The blue dashed line indicates the region of interest. Adapted from Basinski *et al.* (2021). © Basinski *et al.;* 2021. Licence: CC BY.

While spatial determinants of Lassa fever risk are outside of the scope of this thesis, uncertainty in the assumed structure of zoonotic hazard and exposure varied my estimates of disease and infection incidence substantially (Chapter 2) and has important implications for the extension of my climate-driven reservoir host and virus model (Chapter 4) into a spatiotemporal model for Lassa fever. Therefore, I will briefly explore this topic in terms of current knowledge, open questions and data gaps, and the potential for reservoir host ecology (as modelled in Chapter 3) to improve understanding of spatial determinants of Lassa fever risk.

As previously discussed, differing assumptions of *M. natalensis* distribution, human-rodent contact scaling, and viral prevalence in the reservoir host lead to differing projected spatial incidence of Lassa virus infection. The distribution of *M. natalensis* has historically been studied only through environmental niche modelling approaches, with models predicting probability of occurrence through presence or presence-absence models. While this niche modelling approach may inform on a species' likely distribution – commonly used to aid conservation planning (Costa *et al.* 2010), assess invasive potential (Thompson *et al.* 2011), and predict effects of climate change (Wiens *et al.* 2009) – it does not give the full picture of potential for disease transmission. The presence of a species does not imply that it occurs at a high abundance, and the probability of occurrence does not necessarily correlate well with abundance (Estrada & Arroyo 2012). Yet, the use of environmental niche models and similar zoonotic niche models as a proxy for the risk posed by a zoonotic host is common across zoonoses (Pigott *et al.* 2014, 2015; Chalghaf *et al.* 2016) and rarely disputed. Abundance of *M. natalensis* may be a more appropriate measure of Lassa virus risk, but – while possible – is more challenging and data-intensive to model (Gibb 2020). In a rodent capture study sampling several sites across Ghana, *M. natalensis* were found to only represent 27% of the 764 captured rodents, often outnumbered by *Praomys daltoni* which is not known to host Lassa virus (Kronmann *et al.* 2013). Meanwhile, a study in Guinea found 54% of 1123 captured rodents were *M. natalensis* (Fichet-Calvet *et al.* 2007), and in southern Mali, 80% of 103 rodents (Safronetz *et al.* 2010). If *M. natalensis* is the most effective host of Lassa virus but is outcompeted by other rodents in some areas, then this implies the potential for spatial heterogeneity of Lassa virus risk driven by competitor rodent abundances and distributions. Thus far, these complex ecosystem dynamics have not been considered in spatial models of *M. natalensis* or Lassa virus.

Another key component of the spatial structure of disease risk is the distribution of Lassa virus within the reservoir host. Lassa virus is limited by presence of a reservoir but may not be present entirely or uniformly across the reservoir host distribution. One possible explanation for the apparent spatial heterogeneity in Lassa fever incidence could be heterogeneity in the prevalence of Lassa virus in the reservoir host. However, given the spatial biases in rodent serosurveys which have historically focussed on the Mano River Union area where Lassa fever is already known to be endemic (Figure 5.1B), this hypothesis is currently impossible to test. The rodent trapping study in Ghana found no *M. natalensis* positive for arenaviruses out of a total of 209 individuals captured and tested (Kronmann *et al.* 2013), providing evidence for the heterogeneous Lassa virus distribution hypothesis, but only represents ten sites in one country. Additionally, without sampling rodents for Lassa virus at an extremely high spatial resolution, it will be unclear to what degree rodent seroprevalence is locally variable, or "patchy". Choice of spatial model leads to substantially different estimates of spatial virus distribution within the reservoir host; my model (Chapter 2) based on weighted interpolation led to a relatively homogeneous map of viral prevalence (Figure 5.2A), whereas the model of Basinski *et al.* (2021) assuming environmental factors were correlates led to a much more heterogeneous viral prevalence map (Figure 5.2B). Understanding of Lassa virus distribution in *M. natalensis* is currently severely hampered by biases in data and a lack of knowledge around the most appropriate spatial model for these data, which in turn limit the precision and certainty of any broad-scale spatial analyses of Lassa fever.

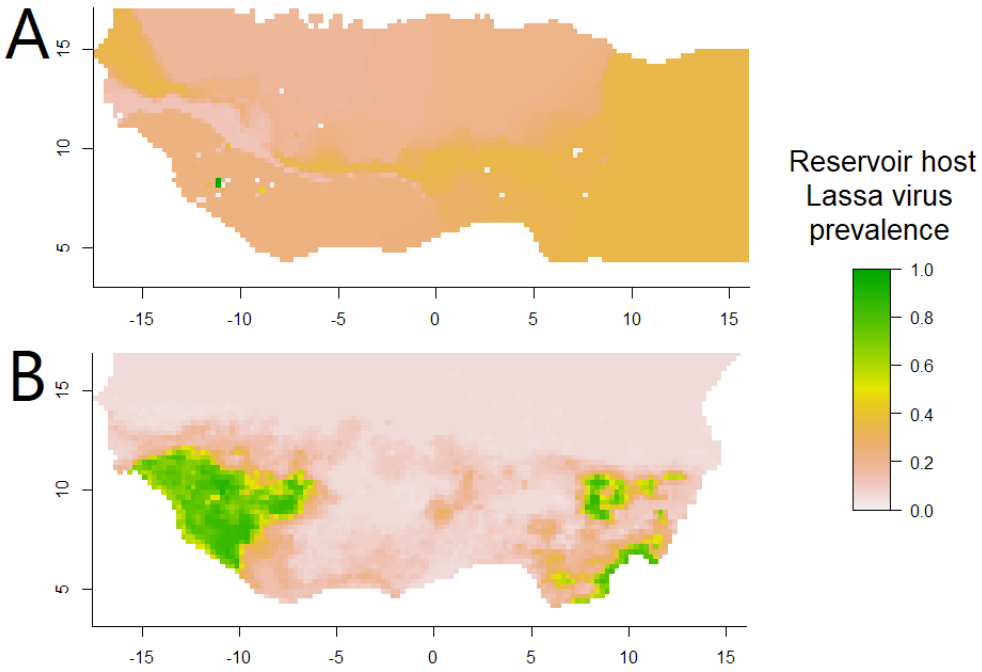


Figure 5.2: Differing models of Lassa virus prevalence in M. natalensis. While both models are based on largely the same rodent serosurvey data, differing methods lead to substantially differing maps of viral prevalence in the reservoir host. A: Virus prevalence (proportion of rodents seropositive), interpolated with a contact-based weighting dependent on estimated relative abundance map (virus seroprevalence map "V2" in Chapter 2). B: Probability of Lassa virus occurrence in *M. natalensis* estimated using an environmental niche approach by Basinski et al. (2021) (virus occurrence map "V1" in Chapter 2).

Phylogenetic evidence has been used to explore origins and dispersal of Lassa virus across West Africa and can help inform on the potential for spatial heterogeneity of Lassa virus prevalence in the reservoir host and resulting observed spatial patterns in human infection. Phylogeographic clustering analysis points to ancient origins of Lassa virus in Nigeria with historic exportation reaching Côte d'Ivoire, Guinea, Liberia, and Sierra Leone in the last centuries (Andersen *et al.* 2015). It is difficult to imagine how this westward exportation could have occurred without first establishing in the countries Benin, Togo, and Ghana, a distance on the scale of around 500km. Following the 2014–2016 Lassa fever outbreak in Benin, phylogenetic evidence was found for a distinct strain of Lassa virus in Benin and Togo (Whitmer *et al.* 2018; Yadouleton *et al.* 2020) which is distantly related to those already identified in Nigeria (at least three strains) (Ehichioya *et al.* 2019), Côte d'Ivoire and southern Mali (one strain) (Manning *et al.* 2015) and the Mano River Union (one strain) (Bowen *et al.* 2000). This newly-identified lineage in Benin and Togo provides evidence for Lassa virus infections and outbreaks originating in these countries, rather than only being imported

from other countries. The phylogeny of Lassa virus suggests that the distribution of Lassa virus is fairly homogeneous across West Africa and does not support an absence of Lassa virus in the central region.

Passive surveillance of Lassa fever is spatially heterogeneous because definitive diagnosis relies on central laboratory testing and high-containment safety (Raabe & Koehler 2017), usually using reverse transcription polymerase chain reaction (RT-PCR) (Asogun *et al.* 2012). Only two West African countries have these permanent clinical facilities, which include the Lassa Diagnostic Laboratory at Kenema Government Hospital, Sierra Leone (Khan *et al.* 2008), the Institute of Lassa Fever Control at Irrua Specialist Teaching Hospital, Nigeria (Asogun *et al.* 2012), and six other diagnostic facilities in Nigeria (NCDC 2021b). In the absence of laboratory testing, Lassa fever cases can be misdiagnosed as malaria or other febrile diseases (Bausch *et al.* 2004). It is therefore possible that Lassa fever occurs more frequently than observed in some parts of West Africa (Figure 5.1A), and that surveillance may be underlying observed spatial patterns of cases. Given that case reporting in Nigeria, with the greatest spatial density of Lassa virus testing and treatment facilities, likely suffers from under-ascertainment and spatial bias (Chapter 2), it seems probable that cases are even more severely under-ascertained in areas of West Africa outside of Nigeria and the Mano River Union.

Active surveillance of Lassa virus infection is carried out heterogeneously throughout West Africa through serological testing (Figure 5.1B). It is notable that there are no currently published studies of human Lassa virus seroprevalence in Nigeria (Figure 5.1B; Basinski *et al.* 2021) despite belief that this country has the greatest Lassa fever burden. Spatial seroprevalence of Lassa virus is therefore relatively well-understood in the Mano River Union, but poorly elsewhere. The study in Ghana sampling humans and rodents at multiple sites found only two Lassa virus infected rodents out of 764 captured (Kronmann *et al.* 2013), and found that 5% of 657 tested human samples were positive for Lassa virus using an enzyme-linked immunosorbent assay (ELISA) but subsequently tested negative for Lassa virus using a plaque reduction neutralisation test (PRNT) (Nimo-Paintsil *et al.* 2019). These findings could suggest ELISA cross-reactivity following infection with a different – potentially novel – arenavirus (Nimo-Paintsil *et al.* 2019). It is also possible that a novel strain of Lassa virus exists in Ghana to which the PRNT was not sensitive (Nimo-Paintsil *et al.* 2019). To better assess the spatial risk of Lassa virus spillover, it would be valuable to explore human and

rodent seroprevalence across under-sampled areas by increasing active surveillance in countries such as Côte d'Ivoire, Benin, and Togo.

The final potential determinant of spatial Lassa virus risk which I will explore here is spatial variation of climatic patterns. In Chapter 3, I found that both climate seasonality and inter-annual precipitation variability are linked with *M. natalensis* demographic processes, therefore if these climate patterns vary across West Africa, this implies that *M. natalensis* population dynamics may vary, too. I therefore explored spatial variation in climate patterns by computing precipitation and temperature seasonality and inter-annual precipitation variability (method in Table 3.1) from global gridded temperature and precipitation data (NOAA/OAR/ESRL PSD 2021a, b) across the spatial extent from 21°W to 22°E and 2°N to 17°N to include the believed Lassa virus endemic region of West Africa, at a resolution of 1 degree, from the years 2010 to 2021 inclusive. First, I computed the variance in seasonal temperature and precipitation to observe how intra-annual variability in climate varies across West Africa. Then, using the model fit in Chapter 3, I estimated the seasonal patterns of survival, recruitment, and body growth based on temperature and precipitation seasonality (holding all other covariates constant) and computed the variance of these values, to observe the degree to which these processes may be varied by seasonal climate patterns. I also computed the mean absolute precipitation variability to observe the extent to which population dynamics may vary year-on-year in different locations.

The results (Figure 5.3) showed substantial heterogeneity in climate and weather patterns across West Africa and the ways these patterns may affect *M. natalensis* demographic processes. Variance in temperature seasonality increases with latitude (Figure 5.3A), while variance in precipitation seasonality is more spatially heterogeneous (Figure 5.3B), with increased variance appearing to occur closer to the coast. The resulting seasonal demographic processes displayed individualistic spatial patterns of variance. While predicted seasonal variance in survival probability (Figure 5.3C) and mean body growth (Figure 5.3E) present a similar spatial pattern to variance in temperature seasonality (increasing with latitude), variance in predicted recruitment probability (Figure 5.3D) is more spatially heterogeneous. This result implies that in some areas of West Africa, the breeding season of *M. natalensis* is highly seasonally restricted, with clearly delineated periods of high and low recruitment and others experiencing less differentiated breeding and non-breeding periods. Mean absolute inter-annual precipitation variability (Figure 5.3F) was highest in areas closer

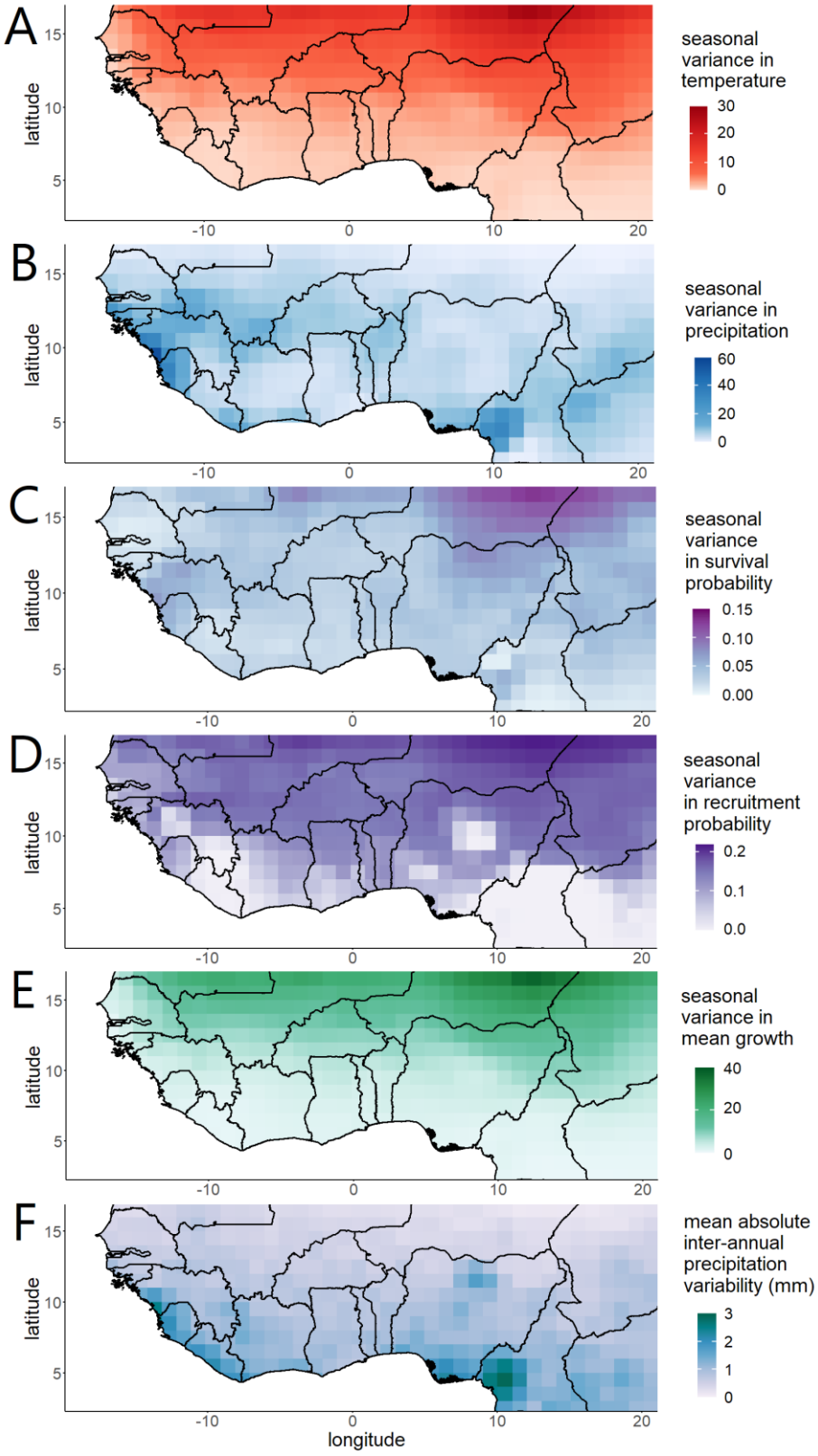


Figure 5.3: Spatial patterns of variability in climate and rodent demographic processes across West Africa computed using global gridded temperature and precipitation data (NOAA/OAR/ESRL PSD 2021a, b) and the climatic-demographic model developed in Chapter 3. A: Variance of the seasonal temperature metric (°C). B: Variance of the seasonal precipitation metric (mm). C–E: Variance of the predicted demographic processes (C: survival probability, D: recruitment probability, E: mean body growth in grams)). F: Mean absolute inter-annual precipitation variability (mm).

to the coast, and in part of northern Nigeria, with these areas predicted to have more variability in year-on-year recruitment rates and resulting population booms. Together these results may begin to explain a degree of spatial heterogeneity in Lassa fever cases, and could indicate that in areas closer to the coast Lassa fever outbreaks are likely to become even more variable year-on-year should inter-annual variability in precipitation increase under climate change.

5.4 Future directions for Lassa fever mitigation

Since the majority of Lassa virus spillover is believed to occur in and around the home (Bonwitt *et al.* 2017), there is a potential for Lassa fever risk to be mitigated by individuals adapting their behaviours and household infrastructure. However, the relative risks of transmission routes are unknown, making it difficult to identify what mitigation steps would be effective. Several factors are believed to lead to risky rodent-human contact, including consumption of rodents (Ter Meulen *et al.* 1996), poor housing quality (Bonner *et al.* 2007), unsecure food storage (Clark *et al.* 2021), and certain agricultural practices (Grant *et al.* 2016). However, without quantifying to what extent these factors – and others which may not have been identified – contribute to disease risk, it is difficult to have confidence in specific strategies. Creating better-informed mitigation strategies at the point of spillover requires further studies within communities to understand risk factors (Clark *et al.* 2021), such as testing of environmental samples for Lassa virus, and further interviews with residents of Lassa virus endemic areas about observations of *M. natalensis* in and around the house.

Current top-down mitigation strategies are limited to education and awareness. The NCDC advises secure food storage, heightened hygiene during food preparation, and community-level rodent culling and other steps as effective personal mitigation actions (NCDC 2021a), and in Nigeria, television and radio advertisements are important sources raising awareness of Lassa fever (Oladeinde *et al.* 2014; Awosanya 2018). Studies indicate variable awareness of Lassa fever in Nigeria, with generally low awareness in rural communities (Oladeinde *et al.* 2014; Ilesanmi Olayinka *et al.* 2015; Akinwumi *et al.* 2016; Awosanya 2018; Usuwa *et al.* 2020), with few people fully aware of transmission risk factors and preventative measures. The two most recent studies listed here (Awosanya 2018; Usuwa *et al.* 2020) have the highest estimated awareness level, potentially indicating improved awareness following high-profile outbreak events in 2018 and 2019.

Community-based rodent control studies have improved Lassa fever awareness and have been welcomed by communities (Mari Saez *et al.* 2018; Clark *et al.* 2021). These studies focussed on rodent control using rodenticide but in both instances communities identified that the building of communal storage structures would also be beneficial but limited by resources, and in one study people were concerned that control is only effective if all members of the community take part. Rodenticide community-based interventions are likely too expensive for widespread uptake by rural communities and include substantial set-up costs (Mari Saez *et al.* 2018). Therefore, successful rodent control-based mitigation strategies would need to be funded by top-down programmes made available to communities by governments or external agents. It would be complementary for such programmes to also provide grants for community storage structures and to fund household rodent-proofing improvements (Clark *et al.* 2021).

In clinical settings, the time delay in diagnosis caused by the need to send samples to central diagnostic laboratories may contribute to poor patient outcomes (Raabe & Koehler 2017). However, the commercial availability of ELISAs, which are cheaper than the "gold standard" RT-PCRs, may enable more cost-effective initial diagnostic testing in primary and secondary health facilities (Hallam *et al.* 2018). Increased development of rapid tests such as lateral flow immunoassays may soon enable even cheaper and quicker initial diagnosis of Lassa fever (Hartnett *et al.* 2015), available outside of laboratory settings at the point-of-case, potentially improving both surveillance coverage and clinical outcomes (Happi *et al.* 2019).

Top-down policies and allocation of resources to reduce Lassa fever burden fundamentally require an understanding for spatiotemporal patterns of zoonotic disease. While at a broad scale, models such as the climatic-demographic model developed in this thesis may be suitable for forecasting timing of increased Lassa fever risk, spatiotemporal risk varies in different ways at the community and household level. For example, precipitation appears to be an important driver of Lassa fever risk (Chapter 3; Leirs *et al.* 1989; Redding *et al.* 2021), but at the community level timing of agricultural practices may be more significant (Grant *et al.* 2016). Therefore, research and response priorities must differ on differing spatial scales (Figure 5.4), with community level research enhanced by participatory modelling which has so far been limited for Lassa fever (Scoones *et al.* 2017). Carrying out monitoring programmes at these different scales will not only enable coordinated and appropriate responses to Lassa fever, but will also make it easier to engage stakeholders such as

communities and healthcare providers; an important component for successful mitigation (Grant *et al.* 2016; Bardosh *et al.* 2017).

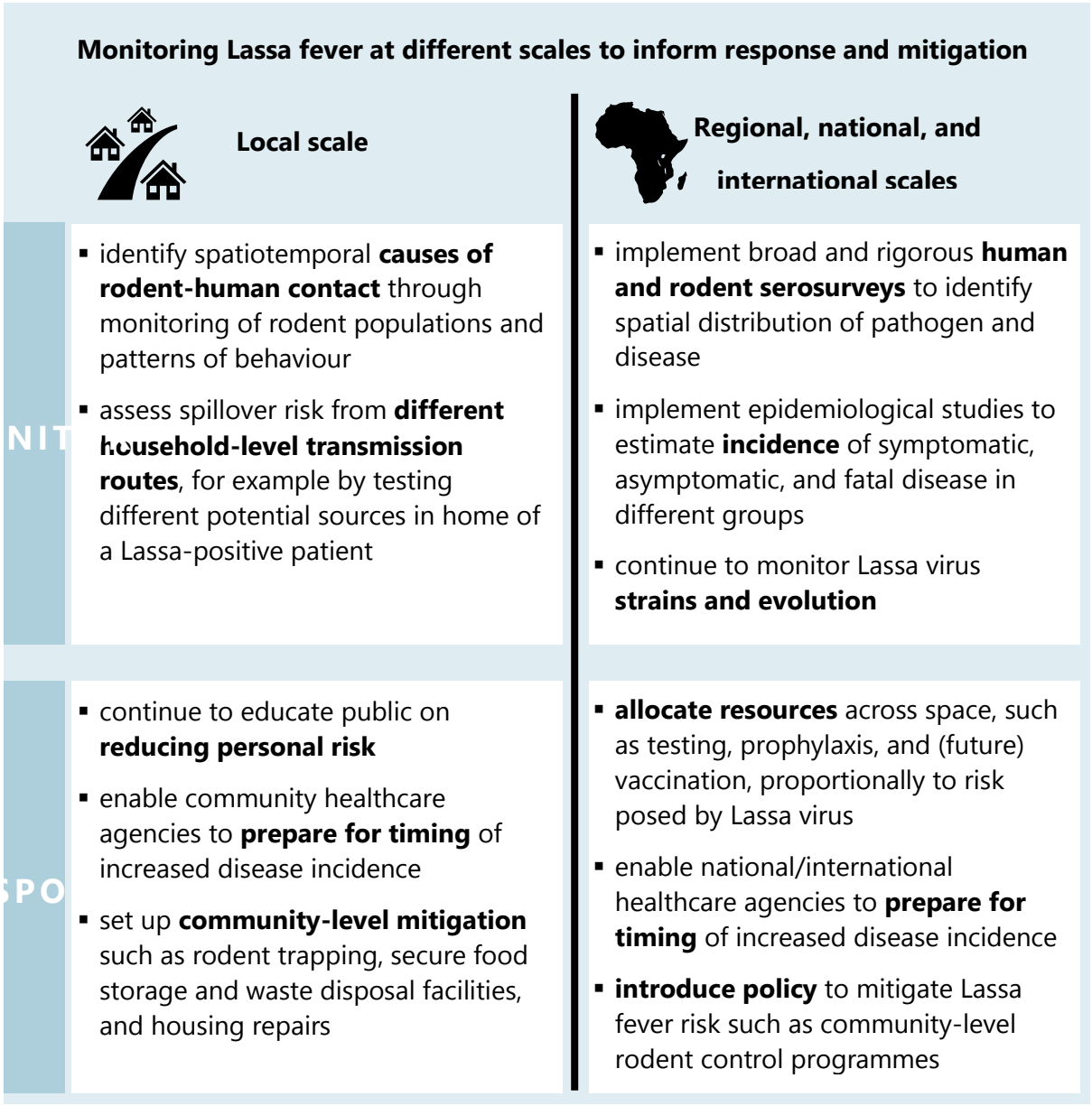


Figure 5.4: Proposed Lassa fever monitoring priorities at different scales and how this information can feed into response and mitigation.

The new frontier of vaccination for Lassa virus presents a big opportunity for improving public health in West Africa (Hallam *et al.* 2018) with increased resourcing and funding accelerating vaccine research in recent years (IAVI 2018; CEPI *et al.* 2021). However, several gaps need to be addressed in order for vaccination to be efficient, effective, and measurable. Population-level coverage of vaccination should be determined based on risk of infection by and complications from Lassa fever, which as previously discussed is unknown. On a broad scale, delivery of vaccines to rural communities in known Lassa virus endemic regions would

be an efficient use of resources, but would be most valuable alongside renewed spatial surveillance efforts to identify missed regions of endemicity. Measuring the effect of vaccine deployment will also require improved estimates of disease incidence and burden, so that vaccinated and unvaccinated groups can be compared to assess the value of vaccination. For example, it would be valuable to assess the extent of Lassa-related chronic deafness (Cummins *et al.* 1990)and to monitor this following vaccination programmes. Many vaccines are contraindicated in pregnancy (Röbl-Mathieu *et al.* 2021), while pregnant women are among those at greatest risk of complications due to Lassa fever, with an estimated risk ratio of death of 2.86 (95% CI [1.77, 4.63]) compared to non-pregnant women with Lassa fever (Kayem *et al.* 2020). Given the importance of maternal health in alleviating poverty (Filippi *et al.* 2006), the association of Lassa fever with low-income and rural communities (Richmond & Baglole 2003), and that poverty and rural location are also associated with reduced access to healthcare (Appendix A), targeting vaccination programmes at women of childbearing age and potentially already-pregnant women in rural communities could be an enhanced opportunity to reduce disease burden and fatality.

While the focus of this thesis and existing research has been on *M. natalensis* as the primary reservoir host of Lassa virus, the role of other hosts must be assessed now and in the future. It would be short-sighted to assume that Lassa fever is restricted by the distribution of *M. natalensis* given that other rodent species have been found to host Lassa virus. Lassa virus has been detected in *Hylomyscus pamfi* in Nigeria and *M. erythroleucus* in Nigeria and Guinea (Olayemi *et al.* 2016a), and in *Mus baoulei* in Benin (Yadouleton *et al.* 2019). It will therefore be important to continue to screen potential hosts, especially if large-scale mitigation strategies centred around *M. natalensis* are put into place, such as rodent control. It should also not be assumed that the Lassa virus disease system will remain static, since the virus is already believed to have spread across West Africa on the timescale of centuries (Andersen *et al.* 2015).

5.5 Opportunities and challenges for zoonotic disease forecasting

The ideal of a fully mechanistic model informed by the processes in the zoonotic transmission pathway requires models of every process in this pathway, in turn requiring suitable data (Hassell *et al.* 2021). These pathways, and relevant types of data, vary across diseases, making a universal mechanistic model of zoonotic disease unfeasible. Instead, process-based models to forecast zoonotic disease incidence and risk factors may have to be

tailored to each zoonotic disease system. A universal method to generate a model for disease based on different and diverse types of data could be possible using pattern-based algorithmic approaches such as machine learning, however, such an approach may hinder the inclusion of biological realism and identification of processes driving disease risk. I therefore believe that a unified model of all zoonotic disease is unrealistic and should not be a research focus. However, there are time-saving opportunities when developing bespoke process-based disease models. By promoting open science behaviour, in which methods are clearly explained and code for new computational methods shared, models can borrow from one another. Such modular models could prevent unnecessary development of methods, or "reinventing the wheel", and allow researchers to update models when applicable innovations are developed. Importantly, combining different models together in this way could improve the integration of methods which are typically used in different disciplines, enabling models to move beyond still-present disciplinary silos (Manlove *et al.* 2016) and integrate cross-disciplinary expertise and experience into modelling frameworks (Scoones *et al.* 2017).

Within any given data type, the sources of data can be numerous; for example, trapping studies of *M. natalensis* have been conducted by multiple groups of researchers but encompass different regions and habitat types (Massawe *et al.* 2005; Makundi *et al.* 2007; Kronmann *et al.* 2013; Fichet-Calvet *et al.* 2014; Mulungu *et al.* 2016; Olayemi *et al.* 2018; Mariën *et al.* 2020), therefore it is necessary to use these data together. This presents the practical challenge of identifying and collating data from different sources. In the absence of a central, universal database, compiling data relies on dedicated systematic reviews which are time consuming, subject to human error, and are outdated as soon as another data collection exercise is performed. Online, freely available databases such as the Global Biodiversity Information Facility (GBIF) are helpful and an important step forwards, however without confidence that a database contains enough data of a given type, collating such data for a model will still require a full literature search. Additionally, current databases are siloed in terms of the data types which they collect; for instance, GBIF data include species presence data and not other measures such as absence or abundance (GBIF 2022).

Integrating data from different sources typically leads to a variety of, and sometimes absence of, metadata. For example, different types of serological assays have different levels of sensitivity and specificity when testing for antibodies related to a given pathogen (Turgeon

2020). Therefore, without information on which assay was used, a serosurvey has limited use when in combination with other serosurveys, since observed patterns cannot be controlled for the effect of assay type. Even when metadata are available, it is sometimes unknown how these variables may impact the variable of interest. For instance, date of rodent trapping study is likely to affect estimated abundance of a rodent species since rodents are highly seasonal, but these population dynamics may not be well-understood. While studies of metadata may appear to be a less constructive use of resources than studies of the variables of interest, without controlling for metadata, models are likely to be biased. To address this challenge, at the study design stage researchers should consider whether their study can be used to measure the impact of metadata on variables of interest, rather than prioritising collecting as much data as possible. Studying the effects of metadata variables, and publishing metadata alongside studies whenever collected, will enable consolidation of existing and future data since the impacts of extraneous factors will be better known. Addressing knowledge gaps around metadata should be considered a valuable research exercise, because it enables forecasting models to make the best use of the diverse data available and reduce biases in these models (Lievesley 2001).

5.6 Future of zoonotic disease management

When vaccines are developed, it will be important for those developing and distributing vaccines to be mindful of the context of zoonotic disease and how the purpose of vaccination differs as compared with disease sustained by human-to-human transmission. While diseases affecting economically valuable animals such as livestock may attract commercial funding, vaccination of zoonotic diseases transmitted to humans mainly from wildlife is rarely commercially attractive and therefore relies on government and charitable funding (Monath 2013). Vaccine development is more costly in terms of time and resources for humans than animals, but – as previously discussed – the vaccination of wild animals is challenging and, additionally, is sometimes limited by the lifespan of the animal in question, potentially requiring frequent vaccination programmes until the disease is potentially eradicated. Additionally, human vaccination alone cannot eradicate zoonotic disease, so a human vaccination strategy requires regular vaccination deployment on the long term. Cross-disciplinary collaborations, such as CEPI and IAVI, and advances in vaccine development (CEPI 2020; Piszczatoski & Gums 2020; CEPI *et al.* 2021) present promising opportunities for effective vaccination strategies for zoonotic disease in the future.

The lack of understanding surrounding relative risks from different routes of transmission for Lassa virus exemplifies a problem common to zoonotic diseases, especially those which originate in wildlife. While for some diseases the individual-level risk is well-qualified, such as rabies which is well-known to be transmitted through bites from an infected, rabid host (Wunner & Briggs 2010), for others, like Lassa fever, the risk factors of spillover are less clear. In Nipah virus outbreaks in Malaysia, it is not known through what mode the pathogen was shed from the intermediate host – domestic pigs, with aerosol spread and infected urine both being possible but neither quantified, making it challenging to identify how humans can modify their behaviour to reduce risk of exposure (Kulkarni *et al.* 2013). An additional challenge is that different zoonoses, and sometimes even the same zoonotic disease in different locations, have idiosyncratic routes of transmission. In Bangladesh and India, Nipah virus is thought to primarily be transmitted from wild fruit bats to humans through the consumption of date palm sap from collection containers which have been contaminated with infected bat fruit bat saliva or urine, as contrasted with the human risk from infected swine in Malaysia, where no instances of bat-to-human infection have been recorded (Kulkarni *et al.* 2013; Sharma *et al.* 2019). To a greater extent than for diseases propagated primarily by human-to-human infection, zoonotic diseases have a diversity of reservoirs, hosts, transmission routes, and human behaviours leading to exposure. Zoonotic disease management is, and will continue to be, hampered by this heterogeneity, with large-scale monitoring needed to choose appropriate mitigation strategies for each disease and each location.

Current evidence suggests that habitat disturbance broadly causes an increased risk of zoonotic spillover due to the predominance of species suited to hosting zoonotic pathogens in disturbed environments (Gibb *et al.* 2020b; Johnson *et al.* 2020), presenting a strong case for biodiversity protection and restoration as a mitigation strategy for zoonotic disease (Keesing & Ostfeld 2021b). For specific diseases, exploring land management strategies to limit the increased prevalence of zoonotic reservoirs and hosts may provide important directions for long term disease management. More broadly, it is possible that funding organisations, including governments, will increase resourcing for maintenance of biodiversity in order to safeguard human health, motivated also by findings across other One Health research areas (Marselle *et al.* 2019; Rojas-Rueda *et al.* 2019). While this mitigation

route may be less attractive since it is indirect and, perhaps, harder to measure, this could be a valuable win-win for human and wildlife wellbeing.

The emergence of a zoonosis as a human-to-human transmissible disease with a basic reproduction number greater than one (meaning that a given infection or disease case is expected to result in more than one secondary infection or disease case in a fully susceptible population) presents an evident epidemic risk with potentially global health implications (e.g., COVID-19; Cucinotta & Vanelli 2020). While risk factors and spatial patterns of zoonotic emergence have been identified (Taylor *et al.* 2001; Jones *et al.* 2008; Smith *et al.* 2014), the stochastic and rare nature of emergence has so far precluded the quantification of disease-specific risk of emergence or identification of measures which will reduce this risk for most zoonoses. The global spread and substantial public health impacts of SARS-CoV-2 (Cucinotta & Vanelli 2020; Wang *et al.* 2022) has stimulated renewed debate into the best way to predict, prevent, or prepare for the next emerging zoonosis.

To predict and prevent a known zoonotic disease from emerging requires an understanding of every step of the zoonotic spillover pathway; a data-intensive activity demanding cross-disciplinary collaboration (Plowright & Hudson 2021). Given that many zoonoses and potential zoonoses are likely unknown however, it is unfeasible to study and monitor every pathogen with emergent potential (Carlson *et al.* 2021). Therefore, many instead suggest strengthening healthcare and governance systems and set up emergency protocols which enable rapid access to research and resources (Carlson *et al.* 2021; Frieden *et al.* 2021; Haldane *et al.* 2021). The common themes in solutions for zoonotic disease management are open science, cross-disciplinary collaboration, and equitable access to resources; by taking positive action in these areas, our global research and healthcare system may be better equipped to mitigate the negative effects of both endemic and emerging zoonotic diseases.

References

Abdulhamid, A., Hussaini, N., Musa, S.S. & He, D. (2022). Mathematical analysis of Lassa fever epidemic with effects of environmental transmission. *Results Phys*, 35, 105335.

Aber, J.D. (1997). Why don't we believe the models? *Bull Ecol Soc Am*, 78(3), 232–233.

Abidemi, A., Owolabi, K.M. & Pindza, E. (2022). Modelling the transmission dynamics of Lassa fever with nonlinear incidence rate and vertical transmission. *Physica A*, 597, 127259.

Ajayi, N.A., Nwigwe, C.G., Azuogu, B.N., Onyire, B.N., Nwonwu, E.U., Ogbonnaya, L.U., *et al.* (2013). Containing a Lassa fever epidemic in a resource-limited setting: outbreak description and lessons learned from Abakaliki, Nigeria (January-March 2012). *Int J Infect Dis*, 17(11), e1011–e1016.

Akhmetzhanov, A.R., Asai, Y. & Nishiura, H. (2019). Quantifying the seasonal drivers of transmission for Lassa fever in Nigeria. *Philos Trans R Soc B*, 374(1775), 20180268.

Akinwumi, A.A., Ademola, A.O., Oladimeji, A.E., Abiodun, O.C., Oghenevo, A.G., Adebola, A.O., *et al.* (2016). Knowledge of Lassa fever among students of a college education: Call for inclusion in curriculum. *Br J Med Med Res*, 16(9), 1–8.

Alexander, K.A., Lewis, B.L., Marathe, M., Eubank, S. & Blackburn, J.K. (2012). Modeling of wildlife-associated zoonoses: Applications and caveats. *Vector Borne Zoonotic Dis*, 12(12), 1005–1018.

Allen, T., Murray, K.A., Zambrana-Torrelio, C., Morse, S.S., Rondinini, C., Di Marco, M., *et al.* (2017). Global hotspots and correlates of emerging zoonotic diseases. *Nat Commun*, 8(1), 1124.

Amorosa, V., MacNeil, A., McConnell, R., Patel, A., Dillon, K.E., Hamilton, K., *et al.* (2010). Imported Lassa fever, Pennsylvania, USA, 2010. *Emerg Infect Dis*, 16(10), 1598–1600.

Andersen, K.G., Shapiro, B.J., Matranga, C.B., Sealfon, R., Lin, A.E., Moses, L.M., *et al.* (2015). Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell*, 162(4), 738.

Andreassen, H.P., Sundell, J., Ecke, F., Halle, S., Haapakoski, M., Henttonen, H., *et al.* (2021). Population cycles and outbreaks of small rodents: ten essential questions we still need to solve. *Oecologia*, 195(3), 601.

Anyamba, A., Chretien, J.P., Small, J., Tucker, C.J., Formenty, P.B., Richardson, J.H., *et al.* (2009). Prediction of a Rift Valley fever outbreak. *PNAS*, 106(3), 955–959.

Arruda, L.B., Haider, N., Olayemi, A., Simons, D., Ehichioya, D., Yinka-Ogunleye, A., *et al.* (2021). The niche of One Health approaches in Lassa fever surveillance and control. *Ann Clin Microbiol Antimicrob*, 20(1), 29.

Asogun, D.A., Adomeh, D.I., Ehimuan, J., Odia, I., Hass, M., Gabriel, M., *et al.* (2012). Molecular diagnostics for Lassa fever at Irrua Specialist Teaching Hospital, Nigeria: Lessons learnt from two years of laboratory operation. *PLOS Negl Trop Dis*, 6(9), e1839.

Awosanya, E.J. (2018). Post-epidemic awareness and knowledge of Lassa fever among residents in affected community in Ibadan, Oyo State, Nigeria. *Vet World*, 11(8), 1059–1063.

Bardosh, K.L., Scoones, J.C., Grace, D., Kalema-Zikusoka, G., Jones, K.E., De Balogh, K., *et al.* (2017). Engaging research with policy and action: what are the challenges of responding to zoonotic disease in Africa? *Philos Trans R Soc B*, 372(1725), 20160172.

Barnett, K.M. & Civitello, D.J. (2020). Ecological and evolutionary challenges for wildlife vaccination. *Trends Parasitol*, 36(12), 970–978.

Barua, S., Dénes, A. & Ibrahim, M.A. (2021). A seasonal model to assess intervention strategies for preventing periodic recurrence of Lassa fever. *Heliyon*, 7(8), e07760.

Basinski, A.J., Fichet-Calvet, E., Sjodin, A.R., Varrelman, T.J., Remien, C.H., Layman, N.C., *et al.* (2021). Bridging the gap: Using reservoir ecology and human serosurveys to estimate Lassa virus spillover in West Africa. *PLOS Comput Biol*, 17(3), e1008811.

Bausch, D.G., Demby, A.H., Coulibaly, M., Kanu, J., Goba, A., Bah, A., *et al.* (2004). Lassa fever in Guinea: I. Epidemiology of human disease and clinical observations. *Vector Borne Zoonotic Dis*, 1(4), 269–281.

De Bellocq, J.G., Bryjová, A., Martynov, A.A. & Lavrenchenko, L.A. (2020). Dhati Welel virus, the missing mammarenavirus of the widespread *Mastomys natalensis*. *J Vertebr Ecol*, 69(2), 20018.

Bernstein, A.S., Ando, A.W., Loch-Temzelides, T., Vale, M.M., Li, B. V., Li, H., *et al.* (2022). The costs and benefits of primary prevention of zoonotic pandemics. *Sci Adv*, 8(5), eabl4183.

Blackburn, J.K., McNyset, K.M., Curtis, A. & Hugh-Jones, M.E. (2007). Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecological [corrected] niche modeling. *Am J Trop Med Hyg*, 77(6), 1103–1110.

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(2), 241–244.

Bonneaud, C. & Longdon, B. (2020). Emerging pathogen evolution: Using evolutionary theory to understand the fate of novel infectious pathogens. *EMBO Rep*, 21(9), e51374.

Bonner, P.C., Schmidt, W.-P., Belmain, S.R., Oshin, B., Baglole, D. & Borchert, M. (2007). Poor housing quality increases risk of rodent infestation and Lassa fever in refugee camps of Sierra Leone. *Am J Trop Med Hyg*, 77(1), 169–175.

Bonwitt, J., Dawson, M., Kandeh, M., Ansumana, R., Sahr, F., Brown, H., *et al.* (2018). Unintended consequences of the 'bushmeat ban' in West Africa during the 2013–2016 Ebola virus disease epidemic. *Soc Sci Med*, 200, 166–173.

Bonwitt, J., Kelly, A.H., Ansumana, R., Agbla, S., Sahr, F., Saez, A.M., *et al.* (2016). Rat-atouille: A mixed method study to characterize rodent hunting and consumption in the context of Lassa fever. *Ecohealth*, 13(2), 234–247.

Bonwitt, J., Sáez, A.M., Lamin, J., Ansumana, R., Dawson, M., Buanie, J., *et al.* (2017). At home with *Mastomys* and *Rattus*: Human-rodent interactions and potential for primary transmission of Lassa virus in domestic spaces. *Am J Trop Med Hyg*, 96(4), 935–943.

Borremans, B., Sluydts, V., Makundi, R.H., Leirs, H., Borremans, B., Sluydts, V., *et al.* (2014). Evaluation of short-, mid- and long-term effects of toe clipping on a wild rodent. *Wildl*

Res, 42(2), 143–148.

Borremans, B., Vossen, R., Becker-Ziaja, B., Gryseels, S., Hughes, N., Van Gestel, M., *et al.* (2015). Shedding dynamics of Morogoro virus, an African arenavirus closely related to Lassa virus, in its natural reservoir host *Mastomys natalensis*. *Sci Rep*, 5, 10445.

Botti-Lodovico, Y., Nair, P., Nosamiefan, D., Stremlau, M., Schaffner, S., Agignoae, S. V., *et al.* (2021). The origins and future of Sentinel: An early-warning system for pandemic preemption and response. *Viruses*, 13(8), 1605.

Bowen, M.D., Rollin, P.E., Ksiazek, T.G., Hustad, H.L., Bausch, D.G., Demby, A.H., *et al.* (2000). Genetic diversity among Lassa virus strains. *J Virol*, 74(15), 6992–7004.

Bright Ross, J.G., Newman, C., Buesching, C.D. & Macdonald, D.W. (2022). Preserving identity in capture–mark–recapture studies: increasing the accuracy of minimum number alive (MNA) estimates by incorporating inter-census trapping efficiency variation. *Mamm Biol*, 1, 1–14.

Burnham, K.P. & Anderson, D.R. (1987). *Model selection and multimodel inference: A practical information-theoretical approach*. Springer-Verlag, New York, USA.

Burniston, S., Okello, A.L., Khamlome, B., Inthavong, P., Gilbert, J., Blacksell, S.D., *et al.* (2015). Cultural drivers and health-seeking behaviours that impact on the transmission of pig-associated zoonoses in Lao People's Democratic Republic. *Infect Dis Poverty*, 4(1), 11.

Butcher, J.C. (2016). The Euler method. In: *Numerical Methods for Ordinary Differential Equations*. Wiley, New York, USA, pp. 55–69.

Carlson, C.J., Farrell, M.J., Grange, Z., Han, B.A., Mollentze, N., Phelan, A.L., *et al.* (2021). The future of zoonotic risk prediction. *Philos Trans R Soc B*, 376(1837), 20200358.

Carroll, D., Daszak, P., Wolfe, N.D., Gao, G.F., Morel, C.M., Morzaria, S., *et al.* (2018). The Global Virome Project. *Science*, 359(6378), 872–874.

Carver, S., Mills, J.N., Parmenter, C.A., Parmenter, R.R., Richardson, K.S., Harris, R.L., *et al.* (2015). Toward a mechanistic understanding of environmentally forced zoonotic disease emergence: Sin Nombre hantavirus. *Bioscience*, 65(7), 651–666.

Caswell, H. (2006). Matrix population models. In: *Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd, Chichester, United Kingdom.

CEPI. (2020). *CEPI-funded Nipah virus vaccine candidate first to reach Phase 1 clinical trial*. Available at: https://cepi.net/news_cepi/cepi-funded-nipah-virus-vaccine-candidate-first-to-reach-phase-1-clinical-trial/?swcfpc=1. Last accessed 23/03/2022.

CEPI, EDCTP & IAVI. (2021). *EDCTP and CEPI funding moves IAVI's Lassa fever vaccine candidate into advanced clinical development*. Available at: https://cepi.net/news_cepi/edctp-and-cepi-funding-moves-iavis-lassa-fever-vaccine-candidate-into-advanced-clinical-development/. Last accessed 23/03/2022.

Chalghaf, B., Chlif, S., Mayala, B., Ghawar, W., Bettaieb, J., Harrabi, M., *et al.* (2016). Ecological niche modeling for the prediction of the geographic distribution of cutaneous leishmaniasis in Tunisia. *Am J Trop Med Hyg*, 94(4), 844–851.

Christensen, J.T. (1993). The seasonal variation in breeding and growth of *Mastomys natalensis* (Rodentia: Muridae): Evidence for resource limitation. *Afr J Ecol*, 31(1), 1–9.

Ciota, A.T. & Keyel, A.C. (2019). The role of temperature in transmission of zoonotic arboviruses. *Viruses*, 11(11), 1013.

Clark, J., Yakob, L., Douno, M., Lamine, J., Magassouba, N. 'Faly, Fichet-Calvet, E., *et al.* (2021). Domestic risk factors for increased rodent abundance in a Lassa fever endemic region of rural Upper Guinea. *Sci Rep*, 11, 20698.

Clayton, B.A., Middleton, D., Bergfeld, J., Haining, J., Arkinstall, R., Wang, L., *et al.* (2012). Transmission routes for Nipah virus from Malaysia and Bangladesh. *Emerg Infect Dis*, 18(12), 1983.

Clutton-Brock, T. & Sheldon, B.C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends Ecol Evol*, 25(10), 562–573.

Coetzee, C.G. (1975). The biology, behaviour, and ecology of *Mastomys natalensis* in southern Africa. *Bull World Health Organ*, 52(4–6), 637.

Colangelo, P., Verheyen, E., Leirs, H., Tatard, C., Denys, C., Dobigny, G., *et al.* (2013). A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biol J Linn Soc*, 108(4), 901–916.

Connolly, S.R., Keith, S.A., Colwell, R.K. & Rahbek, C. (2017). Process, mechanism, and modeling in macroecology. *Trends Ecol Evol*, 32(11), 835–844.

Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot. *Biodivers Conserv*, 19(3), 883–899.

da Costa, V.G., Moreli, M.L. & Saivish, M.V. (2020). The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. *Arch Virol*, 165(7), 1517.

Cucinotta, D. & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomed*, 91(1), 157–160.

Cummins, D., McCormick, J.B., Bennett, D., Samba, J.A., Farrar, B., Machin, S.J., *et al.* (1990). Acute sensorineural deafness in Lassa fever. *JAMA*, 264(16), 2093–2096.

Cunningham, A.A. (2005). A walk on the wild side — emerging wildlife diseases. *Br Med J (Clin Res Ed)*, 331(7527), 1214–5.

Cunningham, A.A., Daszak, P. & Wood, J.L.N. (2017). One Health, emerging infectious diseases and wildlife: two decades of progress? *Philos Trans R Soc B*, 372(1725), 20160167.

Dan-Nwafor, C.C., Ipadeola, O., Smout, E., Ilori, E., Adeyemo, A., Umeokonkwo, C., *et al.* (2019). A cluster of nosocomial Lassa fever cases in a tertiary health facility in Nigeria: Description and lessons learned. *Int J Infect Dis*, 83, 88–94.

Davison, C.W., Chapman, P.M., Wearn, O.R., Bernard, H. & Ewers, R.M. (2019). Shifts in the demographics and behavior of bearded pigs (*Sus barbatus*) across a land-use gradient. *Biotropica*, 51(6), 938–948.

Deka, R.P., Magnusson, U., Grace, D. & Lindahl, J. (2019). Bovine brucellosis: prevalence, risk factors, economic cost and control options with particular reference to India — a review. *Infect Ecol Epidemiol*, 8(1), 1556548.

Demby, A.H., Inapogui, A., Kargbo, K., Koninga, J., Kourouma, K., Kanu, J., *et al.* (2001). Lassa fever in Guinea: II. Distribution and prevalence of Lassa virus infection in small mammals. *Vector Borne Zoonotic Dis*, 1(4), 283–297.

Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. *Numer Math*, 1(1), 269–271.

Dobschuetz, S. von, Roche, X.C.R., Pittiglio, C., Plee, L., Shadomy, S., Palamara, E., *et al.* (2019). Ebola in animals – our knowledge to date: assessing human exposure risks. *Int J Infect Dis*, 79(Supplement 1), 1–2.

Dobson, A. (2009). Climate variability, global change, immunity, and the dynamics of infectious diseases. *Ecology*, 90(4), 920–927.

Dobson, F.S. (2007). Fast and slow life histories of rodents. In: *Rodent Societies: An ecological and evolutionary perspective* (eds. Wolff, J.O. & Sherman, P.W.). The University of Chicago Press, Chicago, USA.

Dorjee, S., Poljak, Z., Revie, C.W., Bridgland, J., Mcnab, B., Leger, E., *et al.* (2013). A review of simulation modelling approaches used for the spread of zoonotic influenza viruses in animal and human populations. *Zoonoses Public Health*, 60(6), 383–411.

Duignan, B. (2021). Occam's razor. *Encycl Brittanica*.

Easterling, M.R., Ellner, S.P. & Dixon, P.M. (2000). Size-specific sensitivity: Applying a new structured population model. *Ecology*, 81(3), 694–708.

ECDC. (2019). *Cases of Lassa fever in the Netherlands ex Sierra Leone*. Available at: https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-cases-lassa-fever-netherlands-ex-sierra-leone. Last accessed 01/12/2021.

Ehichioya, D.U., Dellicour, S., Pahlmann, M., Rieger, T., Oestereich, L., Becker-Ziaja, B., *et al.* (2019). Phylogeography of Lassa virus in Nigeria. *J Virol*, 93(21).

Ehlkes, L., George, M., Samosny, G., Burckhardt, F., Vogt, M., Bent, S., *et al.* (2017). Management of a Lassa fever outbreak, Rhineland-Palatinate, Germany, 2016. *Euro Surveill*, 22(39), 16–00728.

Ehreth, J. (2003). The value of vaccination: a global perspective. *Vaccine*, 21(27–30), 4105–4117.

Ehrlén, J. & Morris, W.F. (2015). Predicting changes in the distribution and abundance of species under environmental change. *Ecol Lett*, 18(3), 303–314.

Einstein, A. (1933). *On the method of theoretical physics: the Herbert Spencer lecture; delivered at Oxford 10 June 1933*. Clarendon Press, Oxford, United Kingdom.

Elith, J. & Leathwick, J.R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst*, 40, 677–697.

Ellner, S.P. & Rees, M. (2006). Integral projection models for species with complex demography. *Am Nat*, 167(3), 410–428.

Engering, A., Hogerwerf, L. & Slingenbergh, J. (2013). Pathogen-host-environment interplay and disease emergence. *Emerg Microbes Infect*, 2(2), e5.

Estrada-Peña, A., Ostfeld, R.S., Peterson, A.T., Poulin, R. & de la Fuente, J. (2014). Effects of

environmental change on zoonotic disease risk: an ecological primer. *Trends Parasitol*, 30(4), 205–214.

Estrada, A. & Arroyo, B. (2012). Occurrence vs abundance models: Differences between species with varying aggregation patterns. *Biol Conserv*, 152, 37–45.

FAO. (2021). *Evaluation of FAO/USAID Emerging Pandemic Threats programme phase II (EPT-2)*. Rome, Italy.

Feldmann, H., Sprecher, A. & Geisbert, T.W. (2020). Ebola. *N Engl J Med*, 382(19), 1832–1842.

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(9), 665.

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

Fichet-Calvet, E. & Rogers, D.J. (2009). Risk maps of Lassa fever in West Africa. *PLOS Negl Trop Dis*, 3(3), e388.

Filippi, V., Ronsmans, C., Campbell, O.M., Graham, W.J., Mills, A., Borghi, J., *et al.* (2006). Maternal health in poor countries: the broader context and a call for action. *Lancet*, 368(9546), 1535–1541.

Firquet, E., Leirs, H. & Bronner, G. (1996). Germinating grasses and reproductive seasonality of *Mastomys*-species (Rodentia, Muridae). *Mammalia*, 60(4), 775–779.

Fisher-Hoch, S.P., Tomori, O., Nasidi, A., Perez-Oronoz, G.I., Fakile, Y., Hutwagner, L., *et al.* (1995). Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. *Br Med J* , 311(7009), 857.

Fitzpatrick, M.C. & Hargrove, W.W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodivers Conserv*, 18(8), 2255–2261.

Frame, J.D., Baldwin, J.M., Gocke, D.J. & Troup, J.M. (1970). Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings. *Am J Trop Med Hyg*, 19(4), 670–676.

Franklinos, L.H.V., Jones, K.E., Redding, D.W. & Abubakar, I. (2019). The effect of global change on mosquito-borne disease. *Lancet Infect Dis*, 19(9), e302–e312.

Friant, S., Ayambem, W.A., Alobi, A.O., Ifebueme, N.M., Otukpa, O.M., Ogar, D.A., *et al.* (2020). Eating bushmeat improves food security in a biodiversity and infectious disease "hotspot." *Ecohealth*, 17(1), 125–138.

Frieden, T.R., Buissonnière, M. & McClelland, A. (2021). The world must prepare now for the next pandemic. *Br Med J Glob Heal*, 6(3), e005184.

Gaetani, M., Janicot, S., Vrac, M., Famien, A.M. & Sultan, B. (2020). Robust assessment of the time of emergence of precipitation change in West Africa. *Sci Rep*, 10(1), 7670.

GBIF. (2022). *What is GBIF?* Available at: https://www.gbif.org/what-is-gbif. Last accessed 12/05/2022.

Gentles, A.D., Guth, S., Rozins, C. & Brook, C.E. (2020). A review of mechanistic models of viral

dynamics in bat reservoirs for zoonotic disease. *Pathog Glob Health*, 114(8), 407–425.

George, D.B., Webb, C.T., Farnsworth, M.L., O'Shea, T.J., Bowen, R.A., Smith, D.L., *et al.* (2011). Host and viral ecology determine bat rabies seasonality and maintenance. *PNAS*, 108(25), 10208–10213.

Gibb, R., Franklinos, L.H. V, Redding, D.W. & Jones, K.E. (2020a). Ecosystem perspectives are needed to manage zoonotic risks in a changing climate. *Br Med J* , 371, m3389.

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(6), 276–288.

Gibb, R., Redding, D.W., Chin, K.Q., Donnelly, C.A., Blackburn, T.M., Newbold, T., *et al.* (2020b). Zoonotic host diversity increases in human-dominated ecosystems. *Nature*, 584(7821), 398–402.

Gibb, R.J. (2020). Understanding and predicting effects of global environmental change on zoonotic disease. PhD thesis. UCL (University College London).

Gimenez, O., Buckland, S.T., Morgan, B.J.T., Bez, N., Bertrand, S., Choquet, R., *et al.* (2014). Statistical ecology comes of age. *Biol Lett*, 10, 20140698.

Glidden, C.K., Nova, N., Kain, M.P., Lagerstrom, K.M., Skinner, E.B., Mandle, L., *et al.* (2021). Human-mediated impacts on biodiversity and the consequences for zoonotic disease spillover. *Curr Biol*, 31(19), R1342–R1361.

Goossens, S., Wybouw, N., Van Leeuwen, T. & Bonte, D. (2020). The physiology of movement. *Mov Ecol*, 8(1), 5.

Gostin, L.O., Hodge, J.G.J., Bloom, B.R., El-Mohandes, A., Fielding, J., Hotez, P., *et al.* (2020). The public health crisis of underimmunisation: a global plan of action. *Lancet Infect Dis*, 20(1), e11–e16.

Grace, D., Gilbert, J., Randolph, T. & Kang'ethe, E. (2012). The multiple burdens of zoonotic disease and an ecohealth approach to their assessment. *Trop Anim Health Prod*, 44(1), 67–73.

Graipel, M.E., Hernández, M.I.M. & Salvador, C. (2014). Evaluation of abundance indexes in open population studies: a comparison in populations of small mammals in southern Brazil. *Brazilian J Biol*, 74(3), 553–559.

Grant, C., Lo Iacono, G., Dzingirai, V., Bett, B., Winnebah, T.R.A. & Atkinson, P.M. (2016). Moving interdisciplinary science forward: integrating participatory modelling with mathematical modelling of zoonotic disease in Africa. *Infect Dis Poverty*, 5(1), 17.

Gray, C.L., Hill, S.L.L., Newbold, T., Hudson, L.N., Boïrger, L., Contu, S., *et al.* (2016). Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nat Commun*, 7(1), 12306.

Grindrod, P. (1996). *The theory and applications of reaction-diffusion equations: Patterns and waves*. 2nd edn. Clarendon Press, Oxford, United Kingdom.

Gryseels, S., Baird, S.J.E., Borremans, B., Makundi, R., Leirs, H. & Goüy de Bellocq, J. (2017). When viruses don't go viral: The importance of host phylogeographic structure in the spatial spread of arenaviruses. *PLOS Pathog*, 13(1), e1006073.

Guerrant, R.L., Walker, D.H. & Weller, P.F. (2011). *Tropical infectious diseases: Principles, pathogens and practice e-Book*. 3rd edn. Elsevier Health Sciences.

Guth, S., Mollentze, N., Renault, K., Streicker, D.G., Visher, E., Boots, M., *et al.* (2022). Bats host the most virulent — but not the most dangerous — zoonotic viruses. *PNAS*, 119(14), e2113628119.

Haldane, V., Jung, A.S., De Foo, C., Bonk, M., Jamieson, M., Wu, S., *et al.* (2021). Strengthening the basics: public health responses to prevent the next pandemic. *Br Med J* , 375.

Hallam, H.J., Hallam, S., Rodriguez, S.E., Barrett, A.D.T., Beasley, D.W.C., Chua, A., *et al.* (2018). Baseline mapping of Lassa fever virology, epidemiology and vaccine research and development. *npj Vaccines*, 3(1), 11.

Halliday, J.E.B., Allan, K.J., Ekwem, D., Cleaveland, S., Kazwala, R.R. & Crump, J.A. (2015). Endemic zoonoses in the tropics: a public health problem hiding in plain sight. *Vet Rec*, 176(9), 220–225.

Hancock, P.A., Thomas, M.B. & Godfray, H.C.J. (2009). An age-structured model to evaluate the potential of novel malaria-control interventions: A case study of fungal biopesticide sprays. *Proc R Soc B*, 276(1654), 71–80.

Happi, A.N., Happi, C.T. & Schoepp, R.J. (2019). Lassa fever diagnostics: Past, present, and future. *Curr Opin Virol*, 37, 132–138.

Hartnett, J.N., Boisen, M.L., Oottamasathien, D., Jones, A.B., Millett, M.M., Nelson, D.S., *et al.* (2015). Current and emerging strategies for the diagnosis, prevention and treatment of Lassa fever. *Future Virol*, 10(5), 559–584.

Hassell, J.M., Newbold, T., Dobson, A.P., Linton, Y.M., Franklinos, L.H.V., Zimmerman, D., *et al.* (2021). Towards an ecosystem model of infectious disease. *Nat Ecol Evol*, 5(7), 907–918.

Hayman, D.T.S., Bowen, R.A., Cryan, P.M., Mccracken, G.F., O'Shea, T.J., Peel, A.J., *et al.* (2013). Ecology of zoonotic infectious diseases in bats: Current knowledge and future directions. *Zoonoses Public Health*, 60(1), 2.

Helmick, C.G., Scribner, C.L., Webb, P.A., Krebs, J.W. & McCormick, J.B. (1986). No evidence for increased risk of Lassa fever infection in hospital staff. *Lancet*, 328(8517), 1202–1205.

Hofbauer, J. & Sigmund, K. (1998). Dynamical systems and Lotka-Volterra equations. In: *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, United Kingdom, pp. 1–54.

Holmes, E.C., Rambaut, A. & Andersen, K.G. (2018). Pandemics: spend on surveillance, not prediction. *Nature*, 558(7709), 180–182.

Howard, C.R. & Fletcher, N.F. (2012). Emerging virus diseases: Can we ever expect the unexpected? *Emerg Microbes Infect*, 1(12), e46.

Lo Iacono, G., Cunningham, A.A., Fichet-Calvet, E., Garry, R.F., Grant, D.S., Khan, S.H., *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(1), e3398.

IAVI. (2018). *IAVI and the Public Health Agency of Canada enter into license agreement for technology to enable Lassa fever vaccine development*. *IAVI*. Available at: https://www.iavi.org/news-resources/press-releases/2018/iavi-to-apply-expertise-in-

viral-vectors-to-lassa-fever. Last accessed 09/05/2022.

Ibrahim, M.A. & Dénes, A. (2021). A mathematical model for Lassa fever transmission dynamics in a seasonal environment with a view to the 2017–20 epidemic in Nigeria. *Nonlinear Anal Real World Appl*, 60, 103310.

Ilesanmi Olayinka, S., Omotoso, B., Alele Faith, O. & Adewuyi, P. (2015). Awareness of Lassa fever in a rural community In south west Nigeria. *J Community Heal Res*, 4(1), 1–10.

Ilori, E.A., Frank, C., Dan-Nwafor, C.C., Ipadeola, O., Krings, A., Ukponu, W., *et al.* (2019). Increase in Lassa fever cases in Nigeria, January–March 2018. *Emerg Infect Dis*, 25(5), 1026.

Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., *et al.* (2009). Ecological networks – beyond food webs. *J Anim Ecol*, 78(1), 253–269.

Innocent, E.O., Uche, O.A. & Uche, I.B. (2014). Building a solid health care system in Nigeria: challenges and prospects. *Acad J Interdiscip Stud*, 3(6), 501.

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES secretariat, Bonn, Germany.

Irwin, N.R., Bayerlová, M., Missa, O. & MartÍnková, N. (2012). Complex patterns of host switching in New World arenaviruses. *Mol Ecol*, 21(16), 4137–4150.

Jackson, T.P. & van Aarde, R.J. (2003). Sex- and species-specific growth patterns in cryptic African rodents, *Mastomys natalensis* and *M. coucha*. *J Mammal*, 84(3), 851–860.

Jemberu, W.T., Molla, W., Dagnew, T., Rushton, J. & Hogeveen, H. (2020). Farmers' willingness to pay for foot and mouth disease vaccine in different cattle production systems in Amhara region of Ethiopia. *PLOS One*, 15(10), e0239829.

Jerolmack, C. (2013). Who's worried about turkeys? How 'organisational silos' impede zoonotic disease surveillance. *Sociol Health Illn*, 35(2), 200–212.

Johnson, C.K., Hitchens, P.L., Pandit, P.S., Rushmore, J., Evans, T.S., Young, C.C.W., *et al.* (2020). Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proc R Soc B*, 287(1924).

Johnson, I., Hansen, A. & Bi, P. (2018). The challenges of implementing an integrated One Health surveillance system in Australia. *Zoonoses Public Health*, 65(1), e229–e236.

Johnson, N.P.A.S. & Mueller, J. (2002). Updating the accounts: Global mortality of the 1918 "Spanish" influenza pandemic. *Bull Hist Med*, 76(1), 105–115.

Johnston, H.L. & Oliff, W.D. (1954). The oestrous cycle of female *Rattus (Mastomys) natalensis* (Smith) as observed in the laboratory. *Proc Zool Soc London*, 124(3), 605–613.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., *et al.* (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990–993.

Kallio, E.R., Begon, M., Henttonen, H., Koskela, E., Mappes, T., Vaheri, A., *et al.* (2009). Cyclic hantavirus epidemics in humans — Predicted by rodent host dynamics. *Epidemics*, 1(2), 101–107.

Karesh, W.B., Dobson, A., Lloyd-Smith, J.O., Lubroth, J., Dixon, M.A., Bennett, M., *et al.* (2012).

Ecology of zoonoses: Natural and unnatural histories. *Lancet*, 380(9857), 1936–1945.

Kausrud, K.L., Viljugrein, H., Frigessi, A., Begon, M., Davis, S., Leirs, H., *et al.* (2007). Climatically driven synchrony of gerbil populations allows large-scale plague outbreaks. *Proc R Soc B*, 274(1621), 1963–1969.

Kayem, N.D., Benson, C., Aye, C.Y.L., Barker, S., Tome, M., Kennedy, S., *et al.* (2020). Lassa fever in pregnancy: a systematic review and meta-analysis. *Trans R Soc Trop Med Hyg*, 114(5), 385–396.

Kearney, M. & Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol Lett*, 12(4), 334–350.

Keesing, F. & Ostfeld, R.S. (2021a). Dilution effects in disease ecology. *Ecol Lett*, 24(11), 2490–2505.

Keesing, F. & Ostfeld, R.S. (2021b). Impacts of biodiversity and biodiversity loss on zoonotic diseases. *PNAS*, 118(17).

Kermack, W.O. & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proc R Soc A*, 115(772), 700–721.

Keusch, G.T., Pappaioanou, M., Gonzalez, M.C., Scott, K.A., Tsai, P. & Editors. (2009). Governance challenges for zoonotic disease surveillance, reporting, and response. In: *Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases*. National Academies Press, Washington (DC), USA.

Khan, S.H., Goba, A., Chu, M., Roth, C., Healing, T., Marx, A., *et al.* (2008). New opportunities for field research on the pathogenesis and treatment of Lassa fever. *Antiviral Res*, 78(1), 103–115.

Kreuder Johnson, C., Hitchens, P.L., Smiley Evans, T., Goldstein, T., Thomas, K., Clements, A., *et al.* (2015). Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci Rep*, 5(1), 14830.

Kronmann, K.C., Nimo-Paintsil, S., Guirguis, F., Kronmann, L.C., Bonney, K., Obiri-Danso, K., *et al.* (2013). Two novel arenaviruses detected in pygmy mice, Ghana. *Emerg Infect Dis*, 19(11), 1832–1835.

Kshirsagar, D.P., Savalia, C. V., Kalyani, I.H., Kumar, R. & Nayak, D.N. (2013). Disease alerts and forecasting of zoonotic diseases: An overview. *Vet World*.

Kulkarni, D.D., Tosh, C., Venkatesh, G. & Senthil Kumar, D. (2013). Nipah virus infection: Current scenario. *Indian J Virol*, 24(3), 398–408.

Kümpel, N.F., Milner-Gulland, E.J., Cowlishaw, G. & Rowcliffe, J.M. (2010). Incentives for hunting: The role of bushmeat in the household economy in rural Equatorial Guinea. *Hum Ecol*, 38(2), 251–264.

Laake, J.L., Johnson, D.S. & Conn, P.B. (2013). marked: an R package for maximum likelihood and Markov chain Monte Carlo analysis of capture–recapture data. *Methods Ecol Evol*, 4(9), 885–890.

Lal, A., Hales, S., French, N. & Baker, M.G. (2012). Seasonality in human zoonotic enteric diseases: A systematic review. *PLOS One*, 7(4), e31883.

Larrieu, E., Gavidia, C.M. & Lightowlers, M.W. (2019). Control of cystic echinococcosis: Background and prospects. *Zoonoses Public Health*, 66(8), 889–899.

Layan, M., Dellicour, S., Baele, G., Cauchemez, S. & Bourhy, H. (2021). Mathematical modelling and phylodynamics for the study of dog rabies dynamics and control: A scoping review. *PLOS Negl Trop Dis*, 15(5), e0009449.

Leirs, H., Stuyck, J., Verhagen, R. & Verheyen, W. (1990). Seasonal variation in growth of *Mastomys natalensis* (Rodentia: Muridae) in Morogoro, Tanzania. *Afr J Ecol*, 28(4), 298–306.

Leirs, H., Verhagen, R. & Verheyen, W. (1993). Productivity of different generations in a population of *Mastomys natalensis* rats in Tanzania. *Oikos*, 68(1), 53.

Leirs, H., Verhagen, R. & Verheyen, W. (1994). The basis of reproductive seasonality in *Mastomys* rats (Rodentia: Muridae) in Tanzania. *J Trop Ecol*, 10(1), 55–66.

Leirs, H., Verheyen, W., Michiels, M., Verhagen, R. & Stuyck, J. (1989). The relation between rainfall and the breeding season of *Mastomys natalensis* (Smith, 1834) in Morogoro, Tanzania. *Ann la société R Zool Belgique*, 119, 59–64.

Leroy, E.M., Rouquet, P., Formenty, P., Souquière, S., Kilbourne, A., Froment, J.M., *et al.* (2004). Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science*, 303(5656), 387–390.

Levinson, J., Bogich, T.L., Olival, K.J., Epstein, J.H., Johnson, C.K., Karesh, W., *et al.* (2013). Targeting surveillance for zoonotic virus discovery. *Emerg Infect Dis*, 19(5), 743.

Lievesley, D. (2001). Making a difference: A role for the responsible international statistician? *J R Stat Soc Ser D Stat*, 50(4), 367–406.

Luis, A.D., Douglass, R.J., Mills, J.N. & Bjørnstad, O.N. (2010). The effect of seasonality, density and climate on the population dynamics of Montana deer mice, important reservoir hosts for Sin Nombre hantavirus. *J Anim Ecol*, 79(2), 462–470.

Lushasi, K., Hayes, S., Ferguson, E.A., Changalucha, J., Cleaveland, S., Govella, N.J., *et al.* (2021). Reservoir dynamics of rabies in south-east Tanzania and the roles of cross-species transmission and domestic dog vaccination. *J Appl Ecol*, 58(11), 2673–2685.

Mableson, H.E., Okello, A., Picozzi, K. & Welburn, S.C. (2014). Neglected zoonotic diseases — the long and winding road to advocacy. *PLOS Negl Trop Dis*, 8(6), e2800.

Mahase, E. (2020). Covid-19: UK starts social distancing after new model points to 260 000 potential deaths. *Br Med J* , 368, m1089.

Makundi, R.H., Massawe, A.W. & Mulungu, L.S. (2007). Reproduction and population dynamics of *Mastomys natalensis* Smith, 1834 in an agricultural landscape in the Western Usambara Mountains, Tanzania. *Integr Zool*, 2(4), 233–238.

Manlove, K.R., Walker, J.G., Craft, M.E., Huyvaert, K.P., Joseph, M.B., Miller, R.S., *et al.* (2016). "One Health" or three? Publication silos among the One Health disciplines. *PLOS Biol*, 14(4), e1002448.

Manning, J.T., Forrester, N. & Paessler, S. (2015). Lassa virus isolates from Mali and the Ivory Coast represent an emerging fifth lineage. *Front Microbiol*, 6, 1037.

Mari Saez, A., Cherif Haidara, M., Camara, A., Kourouma, F., Sage, M., Magassouba, N., *et al.* (2018). Rodent control to fight Lassa fever: Evaluation and lessons learned from a 4-year study in Upper Guinea. *PLOS Negl Trop Dis*, 12(11), e0006829.

Mariën, J., Borremans, B., Gryseels, S., Soropogui, B., De Bruyn, L., Bongo, G.N., *et al.* (2017). No measurable adverse effects of Lassa, Morogoro and Gairo arenaviruses on their rodent reservoir host in natural conditions. *Parasit Vectors*, 10(1), 210.

Mariën, J., Borremans, B., Kourouma, F., Baforday, J., Rieger, T., Günther, S., *et al.* (2019). Evaluation of rodent control to fight Lassa fever based on field data and mathematical modelling. *Emerg Microbes Infect*, 8(1), 640–649.

Mariën, J., Borremans, B., Verhaeren, C., Kirkpatrick, L., Gryseels, S., Goüy de Bellocq, J., *et al.* (2020). Density dependence and persistence of Morogoro arenavirus transmission in a fluctuating population of its reservoir host. *J Anim Ecol*, 89(2), 506–518.

Marselle, M.R., Martens, D., Dallimer, M. & Irvine, K.N. (2019). Review of the mental health and well-being benefits of biodiversity. In: *Biodiversity and health in the face of climate change* (eds. Marselle, M.R., Stadler, J., Korn, H., Irvine, K.N. & Bonn, A.). Springer Open, Cham, Switzerland, pp. 175–212.

Massawe, A.W., Mulungu, L.S., Makundi, R.H., Dlamini, N., Eiseb, S.J., Kirsten, F., *et al.* (2015). Spatial and temporal population dynamics of rodents in three geographically different regions in Africa: Implication for ecologically-based rodent management. *Afr Zool*, 46(2), 393–405.

Massawe, A.W., Rwamugira, W., Leirs, H., Makundi, R.H. & Mulungu, L.S. (2005). Influence of land preparation methods and vegetation cover on population abundance of *Mastomys natalensis* in Morogoro, Tanzania. *Belgian J Zool*, 135(Supplement), 187–190.

Mathers, C., Stevens, G., Ho, J., Ma Fat, D., Retno Mahanani, W., Andreev, K., *et al.* (2013). *Global health estimates technical paper WHO/HIS*. Available at: http://www.who.int/gho/mortality_burden_disease/en/index.html. Last accessed .

Maudlin, I., Eisler, M.C. & Welburn, S.C. (2009). Neglected and endemic zoonoses. *Philos Trans R Soc B*, 364(1530), 2777–2787.

Mayamba, A., Byamungu, R.M., Leirs, H., Moses, I., Makundi, R.H., Kimaro, D.N., *et al.* (2021). Population and breeding patterns of the pest rodent: *Mastomys natalensis* in a maize dominated agroecosystem in Lake Victoria Crescent Zone, Eastern Uganda. *Afr Zool*, 56(1), 76–84.

McCallen, E., Knott, J., Nunez-Mir, G., Taylor, B., Jo, I. & Fei, S. (2019). Trends in ecology: shifts in ecological research themes over the past four decades. *Front Ecol Environ*, 17(2), 109–116.

McCormick, J.B., King, I.J., Webb, P.A., Scribner, C.L., Craven, R.B., Johnson, K.M., *et al.* (1986). Lassa fever. Effective therapy with ribavirin. *N Engl J Med*, 314(1), 20–26.

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(3), 437–444.

Mertens, P.E., Patton, R., Baum, J.J. & Monath, T.P. (1973). Clinical presentation of Lassa fever cases during the hospital epidemic at Zorzor, Liberia, March—April 1972. *Am J Trop Med Hyg*, 22(6).

Metcalf, C.J.E., Graham, A.L., Martinez-Bakker, M. & Childs, D.Z. (2016). Opportunities and challenges of integral projection models for modelling host-parasite dynamics. *J Anim Ecol*, 85(2), 343–355.

Ter Meulen, J., Lukashevich, I., Sidibe, K., Inapogui, A., Marx, M., Dorlemann, A., *et al.* (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(6), 661–666.

Monath, T.P. (2013). Vaccines against diseases transmitted from animals to humans: A one health paradigm. *Vaccine*, 31(46), 5321.

Morand, S., Bordes, F., Blasdell, K., Pilosof, S., Cornu, J.F., Chaisiri, K., *et al.* (2015). Assessing the distribution of disease-bearing rodents in human-modified tropical landscapes. *J Appl Ecol*, 52(3), 784–794.

Moreno, E. (2014). Retrospective and prospective perspectives on zoonotic brucellosis. *Front Microbiol*, 5, 213.

Morse, S.S., Mazet, J.K., Woolhouse, M., Parrish, C.R., Carroll, D., Karesh, W.B., *et al.* (2012). Prediction and prevention of the next pandemic zoonosis. *Lancet*, 380(9857), 1956–1965.

Mulungu, L.S., Ngowo, V., Mdangi, M.E., Katakweba, A.S., Tesha, P., Mrosso, F.P., *et al.* (2016). Survival and recruitment of the multimammate mouse, *Mastomys natalensis* (Smith 1834), in a rice agro-ecosystem. *Mammalia*, 80(2), 205–210.

Musa, S.S., Zhao, S., Gao, D., Lin, Q., Chowell, G. & He, D. (2020). Mechanistic modelling of the large-scale Lassa fever epidemics in Nigeria from 2016 to 2019. *J Theor Biol*, 493, 110209.

Mylne, A.Q.N., Pigott, D.M., Longbottom, J., Shearer, F., Duda, K.A., Messina, J.P., *et al.* (2015). Mapping the zoonotic niche of Lassa fever in Africa. *Trans R Soc Trop Med Hyg*, 109(8), 483–492.

Nandi, A. & Allen, L.J.S. (2021). Probability of a zoonotic spillover with seasonal variation. *Infect Dis Model*, 6, 514–531.

NCDC. (2018). *National guidelines for Lassa fever case management*. Abuja, Nigeria.

NCDC. (2021a). *Lassa fever public health advisory*. *NCDC*. Available at: https://ncdc.gov.ng/news/350/lassa-fever-public-health-advisory. Last accessed 09/05/2022.

NCDC. (2021b). *NCDC intensifies activities for Lassa fever surveillance and response following outbreaks of cases in Nigeria*. *NCDC*. Available at: https://ncdc.gov.ng/news/349/ncdc-intensifies-activities-for-lassa-fever-surveillance-and-response-following-outbreaks-of-cases-in-nigeria. Last accessed 09/05/2022.

NCDC. (2022). *Nigeria Centre for Disease Control Lassa fever situation reports*. Available at: https://ncdc.gov.ng/diseases/sitreps/?cat=5&name=An update of Lassa fever outbreak in Nigeria. Last accessed 04/04/2022.

Nepusz, T. (2022). igraph: Network analysis and visualization. *CRAN*.

Nimo-Paintsil, S.C., Fichet-Calvet, E., Borremans, B., Letizia, A.G., Mohareb, E., K Bonney, J.H.,

et al. (2019). Rodent-borne infections in rural Ghanaian farming communities. *PLOS One*, 14(4), e0215224.

NOAA/OAR/ESRL PSD. (2021a). *CPC global daily temperature*. Available at: https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html. Last accessed 13/08/2021.

NOAA/OAR/ESRL PSD. (2021b). *CPC global unified gauge-based analysis of daily precipitation*. Available at: https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html. Last accessed 13/08/2021.

NPC & ICF. (2019). *Nigeria Demographic and Health Survey 2018*. Abuja, Nigeria and Rockville, Maryland, USA.

Nuismer, S.L. & Bull, J.J. (2020). Self-disseminating vaccines to suppress zoonoses. *Nat Ecol Evol*, 4(9), 1168–1173.

Okoli, C., Hajizadeh, M., Rahman, M.M. & Khanam, R. (2020). Geographical and socioeconomic inequalities in the utilization of maternal healthcare services in Nigeria: 2003-2017. *BMC Health Serv Res*, 20(1), 849.

Oladeinde, B.H., Omoregie, R. & Odia, I. (2014). Public awareness of Lassa fever in three rural communities of Nigeria. *Int J Heal Promot Educ*, 53(3), 128–135.

Olayemi, A., Cadar, D., Magassouba, N., Obadare, A., Kourouma, F., Oyeyiola, A., *et al.* (2016a). New hosts of the Lassa virus. *Sci Rep*, 6(1), 25280.

Olayemi, A. & Fichet-Calvet, E. (2020). Systematics, ecology, and host switching: Attributes affecting emergence of the Lassa virus in rodents across Western Africa. *Viruses*, 12(3), 312.

Olayemi, A., Obadare, A., Oyeyiola, A., Fasogbon, S., Igbokwe, J., Igbahenah, F., *et al.* (2018). Small mammal diversity and dynamics within Nigeria, with emphasis on reservoirs of the Lassa virus. *Syst Biodivers*, 16(2), 118–127.

Olayemi, A., Obadare, A., Oyeyiola, A., Igbokwe, J., Fasogbon, A., Igbahenah, F., *et al.* (2016b). Arenavirus diversity and phylogeography of *Mastomys natalensis* rodents, Nigeria. *Emerg Infect Dis*, 22(4), 694–697.

Oli, M.K. (2019). Population cycles in voles and lemmings: state of the science and future directions. *Mamm Rev*, 49(3), 226–239.

Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L. & Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature*, 546(7660), 646–650.

Ostfeld, R.S. (2015). Disease Ecology. In: *Ecology*. Oxford University Press, Oxford, United Kingdom.

Ostfeld, R.S. & Keesing, F. (2000). Biodiversity and disease risk: the case of Lyme disease. *Conserv Biol*, 14(3), 722–728.

Ostfeld, R.S. & Mills, J.N. (2007). Social behaviour, demography, and rodent-borne pathogens. In: *Rodent Societies: An ecological and evolutionary perspective* (eds. Wolff, J.O. & Sherman, P.W.). The University of Chicago Press, Chicago, USA.

Overbosch, F., de Boer, M., Veldkamp, K.E., Ellerbroek, P., Bleeker-Rovers, C.P., Goorhuis, B., *et*

al. (2020). Public health response to two imported, epidemiologically related cases of Lassa fever in the Netherlands (ex Sierra Leone), November 2019. *Eurosurveillance*, 25(15), 6–10.

Peel, A.J., Baker, K.S., Hayman, D.T.S., Broder, C.C., Cunningham, A.A., Fooks, A.R., *et al.* (2018). Support for viral persistence in bats from age-specific serology and models of maternal immunity. *Sci Rep*, 8, 3859.

Peel, A.J., Pulliam, J.R.C., Luis, A.D., Plowright, R.K., O'Shea, T.J., Hayman, D.T.S., *et al.* (2014). The effect of seasonal birth pulses on pathogen persistence in wild mammal populations. *Proc R Soc B*, 281(1786), 20132962.

Pertierra, L.R., Bartlett, J.C., Duffy, G.A., Vega, G.C., Hughes, K.A., Hayward, S.A.L., *et al.* (2020). Combining correlative and mechanistic niche models with human activity data to elucidate the invasive potential of a sub-Antarctic insect. *J Biogeogr*, 47(3), 658–673.

Pigott, D.M., Golding, N., Mylne, A., Huang, Z., Henry, A.J., Weiss, D.J., *et al.* (2014). Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife*, 3, e04395.

Pigott, D.M., Golding, N., Mylne, A., Huang, Z., Weiss, D.J., Brady, O.J., *et al.* (2015). Mapping the zoonotic niche of Marburg virus disease in Africa. *Trans R Soc Trop Med Hyg*, 109(6), 366–378.

Piszczatoski, C.R. & Gums, J.G. (2020). Ervebo (Ebola Zaire Vaccine, Live/rVSVΔG-ZEBOV-GP): The first licensed vaccine for the prevention of Ebola virus disease: *J Pharm Technol*, 36(6), 243–250.

Plowright, R.K. & Hudson, P.J. (2021). From protein to pandemic: The transdisciplinary approach needed to prevent spillover and the next pandemic. *Viruses*, 13(7), 1298.

Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., *et al.* (2017). Pathways to zoonotic spillover. *Nat Rev Microbiol*, 15(8), 502–510.

Pongsiri, M.J., Roman, J., Ezenwa, V.O., Goldberg, T.L., Koren, H.S., Newbold, S.C., *et al.* (2009). Biodiversity loss affects global disease ecology. *Bioscience*, 59(11), 945–954.

Pooley, S., Fa, J.E. & Nasi, R. (2015). No conservation silver lining to Ebola. *Conserv Biol*, 29(3), 965–967.

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., *et al.* (2017). Land-use futures in the shared socio-economic pathways. *Glob Environ Chang*, 42, 331–345.

Portner, H.O., Scholes, R.J., Agard, J., Archer, E., Arneth, A., Bai, X., *et al.* (2021). *IPBES-IPCC co-sponsored workshop report on biodiversity and climate change*. IPBES and IPCC.

R Core Team. (2020). R: A language and environment for statistical computing.

Raabe, V. & Koehler, J. (2017). Laboratory diagnosis of Lassa fever. *J Clin Microbiol*, 55(6), 1629–1637.

Randall, D.A., Williams, S.D., Kuzmin, I. V., Rupprecht, C.E., Tallents, L.A., Tefera, Z., *et al.* (2004). Rabies in endangered Ethiopian wolves. *Emerg Infect Dis*, 10(12), 2214–2217.

Redding, D.W., Atkinson, P.M., Cunningham, A.A., Lo Iacono, G., Moses, L.M., Wood, J.L.N., *et al.* (2019). Impacts of environmental and socio-economic factors on emergence and epidemic potential of Ebola in Africa. *Nat Commun*, 10(1), 4531.

Redding, D.W., Gibb, R., Dan-Nwafor, C.C., Ilori, E.A., Yashe, R.U., Oladele, S.H., *et al.* (2021). Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria. *Nat Commun*, 12(1), 5759.

Redding, D.W., Moses, L., Cunningham, A.A., Wood, J. & Jones, K.E. (2016). Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever. *Methods Ecol Evol*, 7(6), 646–655.

Reil, D., Rosenfeld, U.M., Imholt, C., Schmidt, S., Ulrich, R.G., Eccard, J.A., *et al.* (2017). Puumala hantavirus infections in bank vole populations: Host and virus dynamics in Central Europe. *BMC Ecol*, 17, 9.

Reis, R.B., Ribeiro, G.S., Felzemburgh, R.D.M., Santana, F.S., Mohr, S., Melendez, A.X.T.O., *et al.* (2008). Impact of environment and social gradient on *Leptospira* infection in urban slums. *PLOS Negl Trop Dis*, 2(4), e228.

Restif, O., Hayman, D.T.S., Pulliam, J.R.C., Plowright, R.K., George, D.B., Luis, A.D., *et al.* (2012). Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol Lett*, 15(10), 1083–1094.

Richmond, J.K. & Baglole, D.J. (2003). Lassa fever: epidemiology, clinical features, and social consequences. *Br Med J* , 327(7426), 1271.

Riman, H.B. & Akpan, E.S. (2012). Healthcare financing and health outcomes in Nigeria: A state level study using multivariate analysis. *Int J Humanit Soc Sci*, 2(15), 296–309.

Roberts, L. (2018). Nigeria hit by unprecedented Lassa fever outbreak. *Science*, 359(6381), 1201–1202.

Robinson, A. (2018). Did Einstein really say that? *Nature*, 557(7703), 30.

Röbl-Mathieu, M., Kunstein, A., Liese, J., Mertens, T. & Wojcinski, M. (2021). Vaccination in pregnancy. *Dtsch Arztebl Int*, 118(15), 262.

Rojas-Rueda, D., Nieuwenhuijsen, M.J., Gascon, M., Perez-Leon, D. & Mudu, P. (2019). Green spaces and mortality: a systematic review and meta-analysis of cohort studies. *Lancet Planet Heal*, 3(11), e469–e477.

Roth, F., Zinsstag, J., Orkhon, D., Chimed-Ochir, G., Hutton, G., Cosivi, O., *et al.* (2003). Human health benefits from livestock vaccination for brucellosis: case study. *Bull World Health Organ*, 81(12), 867–876.

Røttingen, J.-A., Gouglas, D., Feinberg, M., Plotkin, S., Raghavan, K. V., Witty, A., *et al.* (2017). New vaccines against epidemic infectious diseases. *N Engl J Med*, 376, 610–613.

Rougier, T., Lassalle, G., Drouineau, H., Dumoulin, N., Faure, T., Deffuant, G., *et al.* (2015). The combined use of correlative and mechanistic species distribution models benefits low conservation status species. *PLOS One*, 10(10), e0139194.

Rushton, J., Bruce, M., Bellet, C., Torgerson, P., Shaw, A., Marsh, T., *et al.* (2018). Initiation of Global Burden of Animal Diseases programme. *Lancet*, 392(10147), 538–540.

Safronetz, D., Lopez, J.E., Sogoba, N., Traore', S.F., Raffel, S.J., Fischer, E.R., *et al.* (2010). Detection of Lassa virus, Mali. *Emerg Infect Dis*, 16(7), 1123–1126.

Saylors, K.E., Mouiche, M.M., Lucas, A., McIver, D.J., Matsida, A., Clary, C., *et al.* (2021). Market

characteristics and zoonotic disease risk perception in Cameroon bushmeat markets. *Soc Sci Med*, 268, 113358.

Schelling, E., Grace, D., Willingham, A.L. & Randolph, T. (2007). Research approaches for improved pro-poor control of zoonoses. *Food Nutr Bull*, 28(2), S345–S356.

Scherer, C., Radchuk, V., Staubach, C., Müller, S., Blaum, N., Thulke, H.H., *et al.* (2019). Seasonal host life-history processes fuel disease dynamics at different spatial scales. *J Anim Ecol*, 88(11), 1812–1824.

Schmidt, K.A. & Ostfeld, R.S. (2001). Biodiversity and the dilution effect in disease ecology. *Ecology*, 83(3), 609–619.

Schuler, A. (2005). Billions for biodefense: Federal agency biodefense budgeting, FY2005–FY2006. *Biosecurity Bioterrorism Biodefense Strateg Pract Sci*, 3(2), 94–101.

Scoones, I., Jones, K.E., Lo Iacono, G., Redding, D.W., Wilkinson, A. & Wood, J.L.N. (2017). Integrative modelling for one health: Pattern, process and participation. *Philos Trans R Soc B*, 372(1725), 20160164.

Seber, G.A.F. & Schofield, M.R. (2019). *Capture-recapture: Parameter estimation for open animal populations*. 1st edn. Springer Nature, Cham, Switzerland.

Seidahmed, O.M.E., Lu, D., Chong, C.S., Ng, L.C. & Eltahir, E.A.B. (2018). Patterns of urban housing shape dengue distribution in Singapore at neighborhood and country scales. *GeoHealth*, 2(1), 54–67.

Sharma, V., Kaushik, S., Kumar, R., Yadav, J.P. & Kaushik, S. (2019). Emerging trends of Nipah virus: A review. *Rev Med Virol*, 29(1), e2010.

Sharp, P.M. & Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*, 1(1), a006841.

Siddle, K.J., Eromon, P., Barnes, K.G., Mehta, S., Oguzie, J.U., Odia, I., *et al.* (2018). Genomic analysis of Lassa virus during an increase in cases in Nigeria in 2018. *N Engl J Med*, 379(18), 1745–1753.

Simons, D. (2022). *Systematic review of rodent trapping data in West Africa [data repository]*. *GitHub*. Available at: https://github.com/DidDrog11/scoping_review. Last accessed 16/03/2022.

Sluydts, V., Crespin, L., Davis, S., Lima, M. & Leirs, H. (2007). Survival and maturation rates of the African rodent, *Mastomys natalensis*: density-dependence and rainfall. *Integr Zool*, 2(4), 220–232.

Sluydts, V., Davis, S., Mercelis, S. & Leirs, H. (2009). Comparison of multimammate mouse (*Mastomys natalensis*) demography in monoculture and mosaic agricultural habitat: Implications for pest management. *Crop Prot*, 28(8), 647–654.

Smith, G.C., Thulke, H.-H., Fooks, A.R., Artois, M., Macdonald, D.W., Eisinger, D., *et al.* (2008). What is the future of wildlife rabies control in Europe? *Dev Biol*, 131, 283.

Smith, K.F., Goldberg, M., Rosenthal, S., Carlson, L., Chen, J., Chen, C., *et al.* (2014). Global rise in human infectious disease outbreaks. *J R Soc Interface*, 11(101), 20140950.

Smither, S.J., Eastaugh, L.S., Findlay, J.S., O'Brien, L.M. & Lever, M.S. (2020). Survival of Lassa

virus in blood and tissue culture media and in a small particle aerosol. *Pathogens*, 9(9), 680.

Sogoba, N., Feldmann, H. & Safronetz, D. (2012). Lassa fever in West Africa: Evidence for an expanded region of endemicity. *Zoonoses Public Health*, 59(s2), 43–47.

Sogoba, N., Rosenke, K., Adjemian, J., Diawara, S.I., Maiga, O., Keita, M., *et al.* (2016). Lassa virus seroprevalence in Sibirilia commune, Bougouni District, southern Mali. *Emerg Infect Dis*, 22(4), 657–663.

Stephens, P.R., Altizer, S., Smith, K.F., Alonso Aguirre, A., Brown, J.H., Budischak, S.A., *et al.* (2016). The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distributions and impacts. *Ecol Lett*, 19(9), 1159–1171.

Suwannarong, K. & Chapman, R.S. (2015). Characteristics associated with contact with rodents in, around, and outside homes in Khon Kaen Province, Thailand. *Am J Trop Med Hyg*, 92(4), 784–790.

Swai, E.S., Schoonman, L. & Daborn, C.J. (2010). Knowledge and attitude towards zoonoses among animal health workers and livestock keepers in Arusha and Tanga, Tanzania. *Tanzan J Health Res*, 12(4), 272–277.

Taylor, L.H., Latham, S.M. & Woolhouse, M.E.J. (2001). Risk factors for human disease emergence. *Philos Trans R Soc B*, 356(1411), 983–989.

The World Bank. (2022a). *World Development Indicators: Birth rate, crude*. Available at: https://data.worldbank.org/indicator/SP.DYN.CBRT.IN. Last accessed 03/04/2022.

The World Bank. (2022b). *World Development Indicators: Death rate, crude*. Available at: https://data.worldbank.org/indicator/SP.DYN.CDRT.IN. Last accessed 03/04/2022.

The World Bank. (2022c). *World Development Indicators: Population, total*. Available at: https://data.worldbank.org/indicator/SP.POP.TOTL. Last accessed 03/04/2022.

The World Bank. (2022d). *World Development Indicators: Urban population*. Available at: https://data.worldbank.org/indicator/SP.URB.TOTL. Last accessed 03/04/2022.

Thompson, G.D., Robertson, M.P., Webber, B.L., Richardson, D.M., Le Roux, J.J. & Wilson, J.R.U. (2011). Predicting the subspecific identity of invasive species using distribution models: *Acacia saligna* as an example. *Divers Distrib*, 17(5), 1001–1014.

Tjaden, N.B., Caminade, C., Beierkuhnlein, C. & Thomas, S.M. (2018). Mosquito-borne diseases: Advances in modelling climate-change impacts. *Trends Parasitol*, 34(3), 227–245.

Tompros, A., Dean, A.D., Fenton, A., Wilber, M.Q., Carter, E.D. & Gray, M.J. (2021). Frequency-dependent transmission of *Batrachochytrium salamandrivorans* in eastern newts. *Transbound Emerg Dis*, 69(2), 731–741.

Toor, J., Echeverria-Londono, S., Li, X., Abbas, K., Carter, E.D., Clapham, H.E., *et al.* (2021). Lives saved with vaccination for 10 pathogens across 112 countries in a pre-COVID-19 world. *eLife*, 10, e67635.

Torgerson, P.R., Rüegg, S., Devleesschauwer, B., Abela-Ridder, B., Havelaar, A.H., Shaw, A.P.M., *et al.* (2018). zDALY: An adjusted indicator to estimate the burden of zoonotic diseases. *One Heal*, 5, 40–45.

Townsend Peterson, 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(8), e100711.

Turgeon, M. (2020). *Immunology and Serology in Laboratory Medicine e-Book*. Elsevier Health Sciences.

UK Health Security Agency. (2022). *Lassa fever cases identified in England, following travel to West Africa*. *GOV.UK*. Available at: https://www.gov.uk/government/news/lassa-fever-cases-identified-in-england-following-travel-to-west-africa-1. Last accessed 31/03/2022.

UNEP & ILRI. (2020). *Preventing the next pandemic: Zoonotic diseases and how to break the chain of transmission*. Nairobi, Kenya.

USAID. (2021). *Emerging pandemic threats [Fact sheet]*. Available at: https://www.usaid.gov/news-information/fact-sheets/emerging-pandemic-threats-program. Last accessed 25/03/2022.

Usuwa, I.S., Akpa, C.O., Umeokonkwo, C.D., Umoke, M., Oguanuo, C.S., Olorukooba, A.A., *et al.* (2020). Knowledge and risk perception towards Lassa fever infection among residents of affected communities in Ebonyi State, Nigeria: implications for risk communication. *BMC Public Health*, 20(1), 217.

Voutilainen, L., Kallio, E.R., Niemimaa, J., Vapalahti, O. & Henttonen, H. (2016). Temporal dynamics of Puumala hantavirus infection in cyclic populations of bank voles. *Sci Rep*, 6, 21323.

Wang, H., Paulson, K.R., Pease, S.A., Watson, S., Comfort, H., Zheng, P., *et al.* (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet*, 399(10334), 1513–1536.

Wang, X., Tang, S. & Cheke, R.A. (2016). A stage structured mosquito model incorporating effects of precipitation and daily temperature fluctuations. *J Theor Biol*, 411, 27–36.

Welburn, S.C., Beange, I., Ducrotoy, M.J. & Okello, A.L. (2015). The neglected zoonoses — the case for integrated control and advocacy. *Clin Microbiol Infect*, 21(5), 433–443.

Whitmer, S.L.M., Strecker, T., Cadar, D., Dienes, H.P., Faber, K., Patel, K., *et al.* (2018). New lineage of Lassa virus, Togo, 2016. *Emerg Infect Dis*, 24(3), 599.

WHO. (2015). *Blueprint for R&D preparedness and response to public health emergencies due to highly infectious pathogens: Workshop on prioritization of pathogens: 8-9 December 2015*. Geneva, Switzerland.

WHO. (2017). *Lassa fever [Fact sheet]*. Available at: https://www.who.int/news-room/fact-sheets/detail/lassa-fever. Last accessed 31/03/2022.

WHO. (2020a). *Vector-borne diseases [Fact sheet]*. Available at: https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases. Last accessed 22/03/2022.

WHO. (2020b). *WHO COVID-19 Dashboard*. Available at: https://covid19.who.int/. Last accessed 08/04/2022.

WHO. (2020c). *World malaria report 2020: 20 years of global progress and challenges*. Geneva, Switzerland.

WHO. (2020d). *Zoonoses [Fact sheet]*. Available at: https://www.who.int/news-room/fact-sheets/detail/zoonoses. Last accessed 22/03/2022.

WHO. (2021a). *Ebola virus disease [Fact sheet]*. Available at: https://www.who.int/en/news-room/fact-sheets/detail/ebola-virus-disease. Last accessed 10/07/2022.

WHO. (2021b). *HIV/AIDS [Fact sheet]*. Available at: https://www.who.int/news-room/fact-sheets/detail/hiv-aids. Last accessed 22/03/2022.

WHO & DFID-AHP. (2005). *The control of neglected zoonotic diseases: a route to poverty: report of a joint WHO/DFID-AHP Meeting with the participation of FAO and OIE. Geneva, 20 and 21 September 2005.* Geneva, Switzerland.

WHO, ICONZ & UK DFID. (2011). *The control of neglected zoonotic diseases: community based interventions for NZDs prevention and control: report of the third conference organized with ICONZ, DFID-RiU, SOS, EU, TDR and FAO with the participation of ILRI and OIE: 23 - 24 November 2010*. Geneva, Switzerland.

WHO Regional Office for Africa. (2019). *Technical guidelines for Integrated Disease Surveillance and Response in the African region: Third edition*. Brazzaville, Republic of the Congo.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A. & Snyder, M.A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. *PNAS*, 106(SUPPL. 2), 19729–19736.

Wilcox, B.A. & Gubler, D.J. (2005). Disease ecology and the global emergence of zoonotic pathogens. *Environ Health Prev Med*, 10(5), 263–272.

Woolhouse, M.E.J. (2002). Population biology of emerging and re-emerging pathogens. *Trends Microbiol*, 10(10), s3–s7.

Woolhouse, M.E.J. & Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerg Infect Dis*, 11(12), 1842.

WorldPop & CIESIN. (2018). *Global High Resolution Population Denominators Project*. Available at: www.worldpop.org/geodata/summary?id=24776. Last accessed 12/12/2021.

Wunner, W.H. & Briggs, D.J. (2010). Rabies in the 21st century. *PLOS Negl Trop Dis*, 4(3), e591.

Yadouleton, A., Agolinou, A., Kourouma, F., Saizonou, R., Pahlmann, M., Bedié, S.K., *et al.* (2019). Lassa virus in pygmy mice, Benin, 2016–2017. *Emerg Infect Dis*, 25(10), 1977–1979.

Yadouleton, A., Picard, C., Rieger, T., Loko, F., Cadar, D., Kouthon, E.C., *et al.* (2020). Lassa fever in Benin: description of the 2014 and 2016 epidemics and genetic characterization of a new Lassa virus. *Emerg Microbes Infect*, 9(1), 1761–1770.

Yun, N.E. & Walker, D.H. (2012). Pathogenesis of Lassa fever. *Viruses*, 4(10), 2031.

Zeimes, C.B., Olsson, G.E., Ahlm, C. & Vanwambeke, S.O. (2012). Modelling zoonotic diseases in humans: comparison of methods for hantavirus in Sweden. *Int J Health Geogr*, 11(1), 39.

Permissions

Figure 1.3: Map of reported distribution of Lassa fever in West Africa.
© World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.
who.int/about/policies/publishing/copyright

Figure 1.4: Photograph of wild *M. natalensis*.
© Joe Blossom, Alamy Stock Photo; 2009. Reproduced under licence.
Licence obtained for non-commercial use (alamy.com/terms/uk.aspx).

Figure 1.5: The geographic distribution of *M. natalensis* lineages and mammarenaviruses which infect them.
© Journal of Vertebrate Ecology; 2020. Licence: CC-BY.
jvertbiol.cz/submission/#copyright

Figure 1.6: Spatial models of Lassa fever cases, Lassa virus presence, and Lassa virus infection.

Figure 1.6A: © Fichet-Calvet, Rogers; 2009. Licence: CC BY.
journals.plos.org/plosone/s/licenses-and-copyright
Figure 1.6B: © Mylne *et al.*; 2015. Published by Oxford University Press on behalf of Royal Society of Tropical Medicine and Hygiene. Licence: CC BY.
academic.oup.com/trstmh/article/109/8/483/1910156
Figure 1.6C: © Redding et al.; 2016. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society. Licence: CC BY.
besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12549
Figure 1.6D: © Basinski et al.; 2021. Licence: CC BY.
journals.plos.org/plosone/s/licenses-and-copyright

Figure 3.1: CMR study location and patterns of weather and seasonal climate at the site.
© Google, CNES / Airbus, Maxar Technologies; 2021.
about.google/brand-resource-center/products-and-services/geo-guidelines/#google-earth

Figure 5.1: Spatial heterogeneity in observations of Lassa fever and Lassa virus infection.

Figure S1.1: Modelled temporal dynamics and drivers of confirmed Lassa fever cases in the south and north of Nigeria.

Figure S2.6: Weekly confirmed COVID-19 cases in Nigeria.

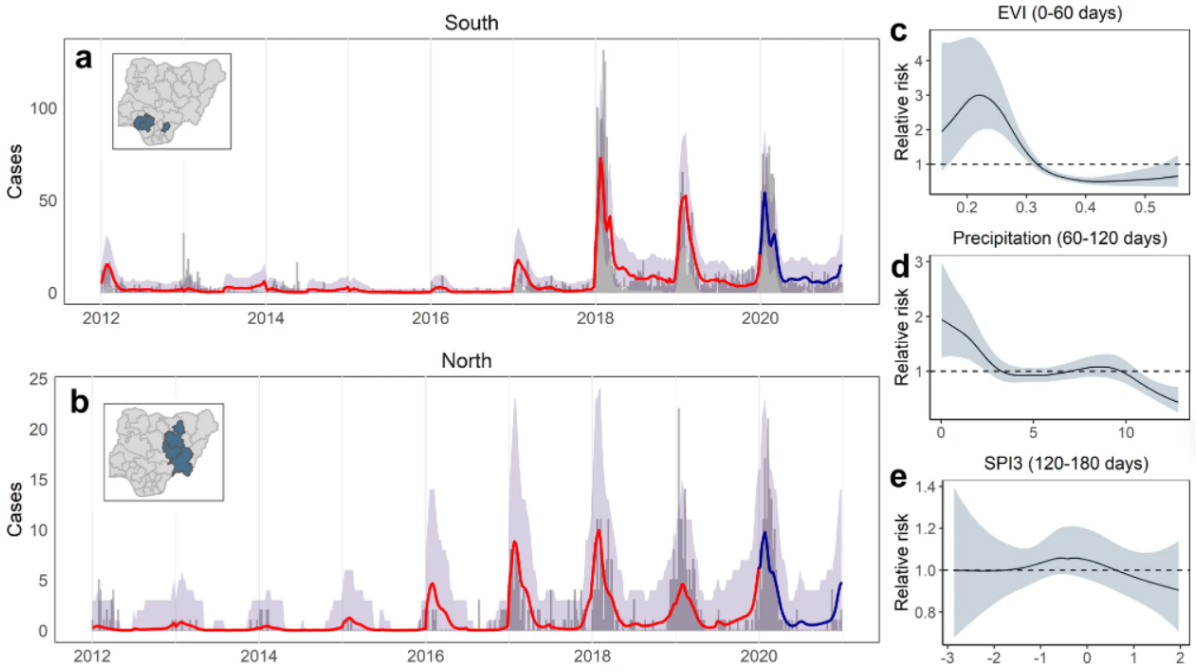
Figure S5.1: Spatiotemporal trends in Lassa fever surveillance.

Appendices

Appendix contents

Appendix 1: Supplementary materials for Chapter 1

Supplementary Figure S1.1. Modelled temporal dynamics and drivers of confirmed Lassa fever cases in the south and north of Nigeria. Reproduced from Redding *et al.* (2021). A-B: Observed weekly case counts (grey bars) and predicted weekly case counts (red/blue line with 95% credible intervals in grey shading). The blue line shows prospective predicted cases for the year which was not included in model fitting. A: States in the south of Nigeria (Edo and Ondo). B: States in the north of Nigeria (Bauchi, Plateau, and Taraba). C-E: Nonlinear fitted effects of environmental variables on relative disease risk with posterior mean (line) and 95% credible interval (shaded area). C: Enhanced Vegetation Index (EVI). D: Mean daily precipitation. E: 3-month Standardised Precipitation Index (SPI3). © Redding *et al.*; 2021. Licence: CC BY.



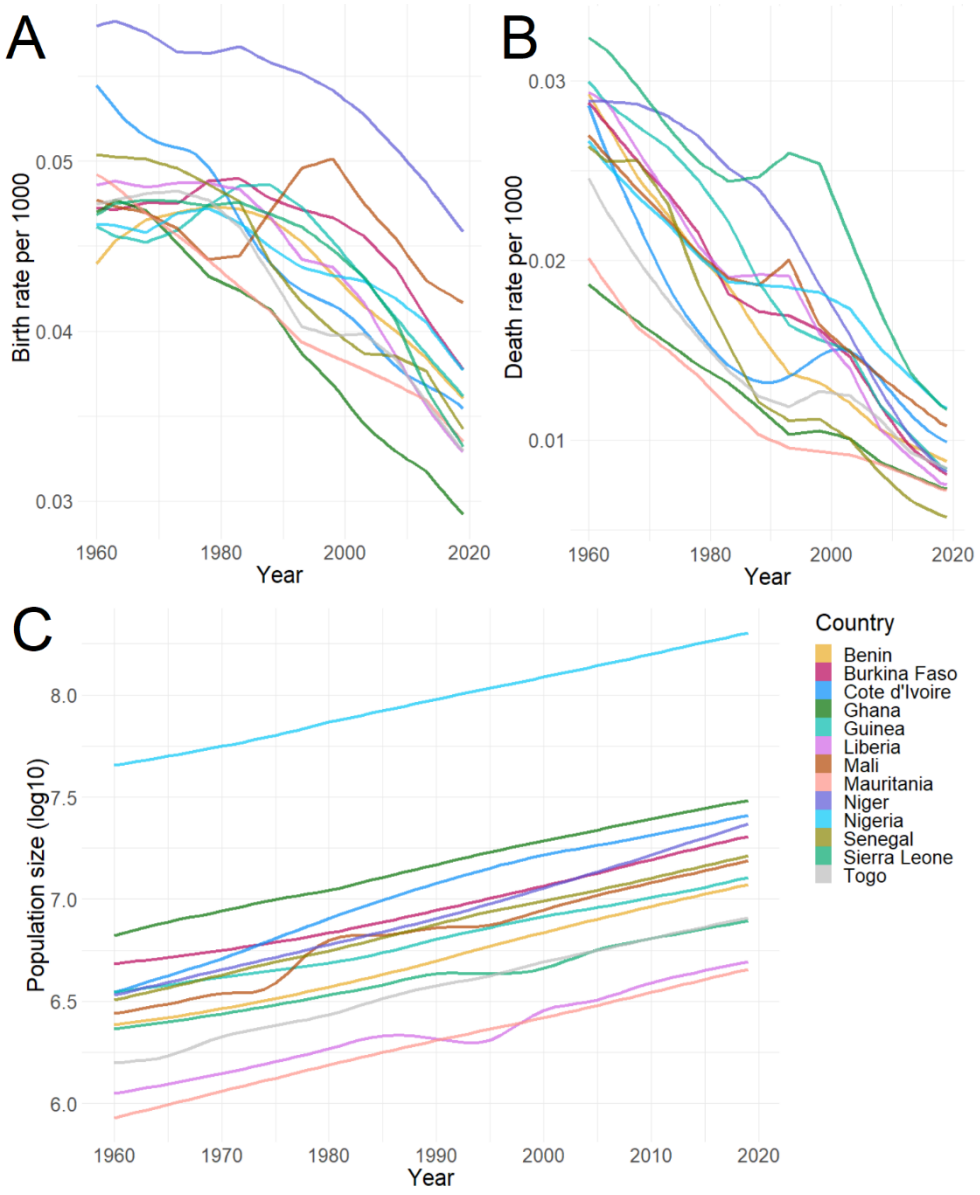
Appendix 2: Supplementary materials for Chapter 2

Supplementary Data S2.1. Extracted data on confirmed cases from NCDC situation reports on Lassa fever from 2018 to 2021 inclusive, by state and year.

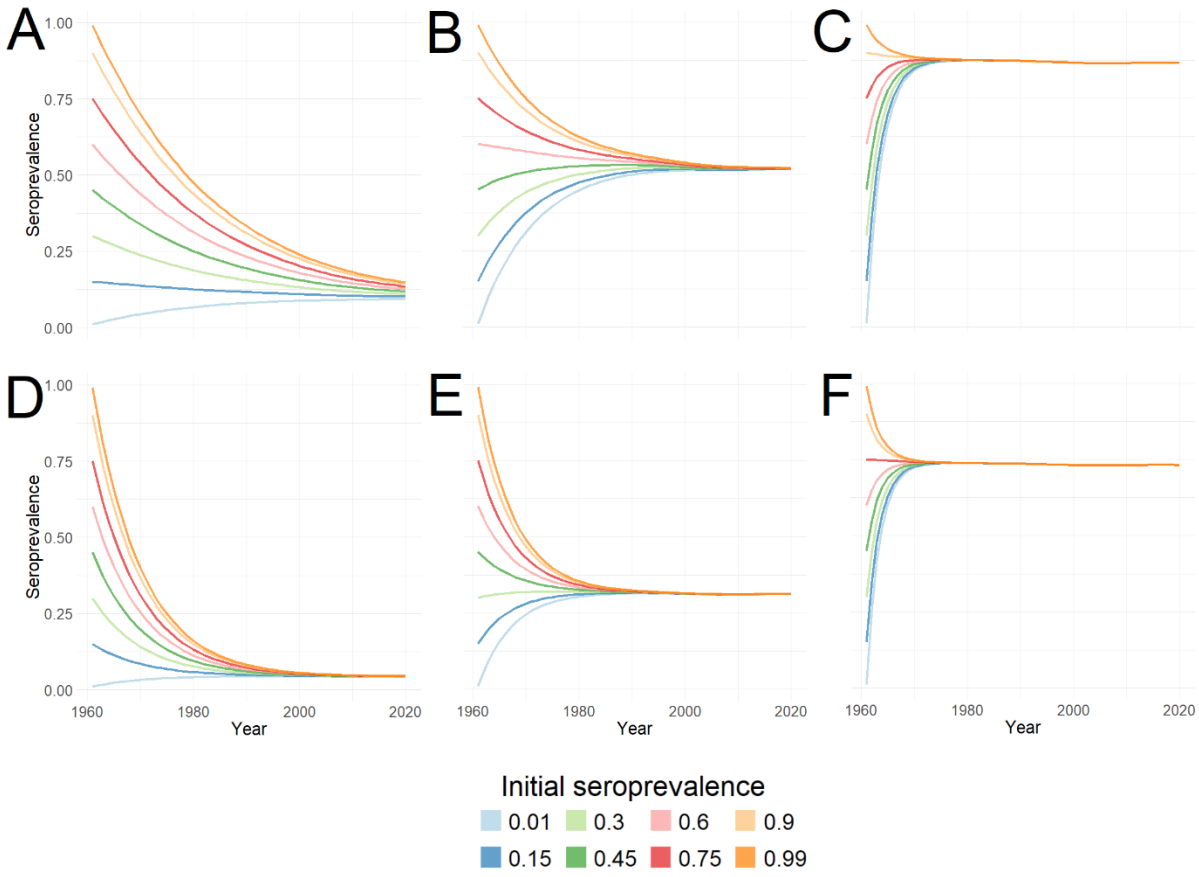
Year	State	Confirmed	Deaths
2018	Taraba	23	7
2018	Rivers	2	1
2018	Nasarawa	4	3
2018	Edo	279	43
2018	Ondo	159	42
2018	Bauchi	24	14
2018	Ebonyi	72	18
2018	Plateau	14	12
2018	Kaduna	3	3
2018	Gombe	4	4
2018	Kano	1	0
2018	Kogi	12	4
2018	Enugu	1	1
2018	Anambra	4	2
2018	Lagos	1	1
2018	Delta	11	6
2018	Imo	5	1
2018	FCT	5	3
2018	Ekiti	2	1
2018	Osun	2	1
2018	Adamawa	3	2
2018	Abia	1	1
2018	Benue	1	1
2019	Taraba	40	16
2019	Rivers	2	1
2019	Nasarawa	7	5
2019	Edo	306	46
2019	Ondo	285	46
2019	Bauchi	56	10
2019	Ebonyi	57	21
2019	Plateau	36	11
2019	Kaduna	5	0
2019	Gombe	4	1
2019	Cross-River	1	1
2019	Kogi	4	3
2019	Enugu	2	1
2019	Lagos	1	0
2019	Kwara	1	0
2019	Delta	3	0
2019	Imo	2	1
2019	Adamawa	1	1
2019	Abia	2	2
2019	Benue	8	5
2019	Kebbi	7	1
2019	Oyo	2	1
2019	Zamfara	1	1

Year	State	Confirmed	Deaths
2020	Ogun	1	0
2020	Taraba	58	22
2020	Rivers	9	3
2020	Nasarawa	10	4
2020	Edo	385	40
2020	Ondo	423	83
2020	Bauchi	53	22
2020	Ebonyi	81	23
2020	Plateau	33	8
2020	Kaduna	7	5
2020	Gombe	9	2
2020	Borno	4	1
2020	Kano	5	1
2020	Kogi	40	8
2020	Enugu	10	2
2020	Anambra	2	1
2020	Lagos	1	0
2020	Delta	18	3
2020	FCT	3	2
2020	Osun	2	0
2020	Adamawa	4	1
2020	Abia	5	2
2020	Benue	10	4
2020	Kebbi	4	2
2020	Oyo	1	0
2020	Katsina	6	2
2020	Sokoto	5	3
2021	Edo	190	14
2021	Ondo	148	44
2021	Taraba	21	12
2021	Ebonyi	17	2
2021	Bauchi	13	3
2021	Benue	8	0
2021	Plateau	8	0
2021	Kaduna	6	4
2021	Enugu	5	1
2021	Nasarawa	3	0
2021	Kogi	3	0
2021	Cross-River	1	0
2021	Imo	1	0
2021	Anambra	1	0
2021	Delta	1	0
2021	Abia	1	0

Supplementary Figure S2.2. Population demography of countries in West Africa has been changing since 1960. Graphs created from crude birth rate, crude death rate, and population size by country from The World Bank development indicators (The World Bank 2022a, b, c).



Supplementary Figure S2.3. Final seroprevalence is not sensitive to initial seroprevalence in mathematical models of seroconversion. Graphs show modelled seroprevalence across time for the example of Nigeria with initial seroprevalence Q_{2019} varied from 0.01 to 0.99 (difference coloured lines). A-C (top row): No seroreversion. D-F (bottom row): Seroreversion implemented with an average seropositive period of 15.6 years. A+D (left column): Force of infection F=0.005 /year. B+E (middle column): $F = 0.05$ /year. C+F (right column): $F = 0.3$ /year.



Supplementary Table S2.4. Bias model parameter estimates, 95% confidence intervals, and p-values for the bias parameter being positive. All reported to 3 significant figures.

Year	a	b	$b > 0$
All years	0.269 [0.214, 0.333]	0.395 [0.337, 0.455]	$p < 0.0001$
2018	0.235 [0.178, 0.302]	0.744 [0.612, 0.866]	$p < 0.0001$
2019	0.228 [0.148, 0.333]	0.485 [0.352, 0.622]	$p < 0.0001$
2020	0.151 [0.127, 0.178]	0.218 [0.148, 0.292]	$p < 0.0001$
2021	0.824 [0.590, 1.17]	0.345, [0.282, 0.416]	$p < 0.0001$

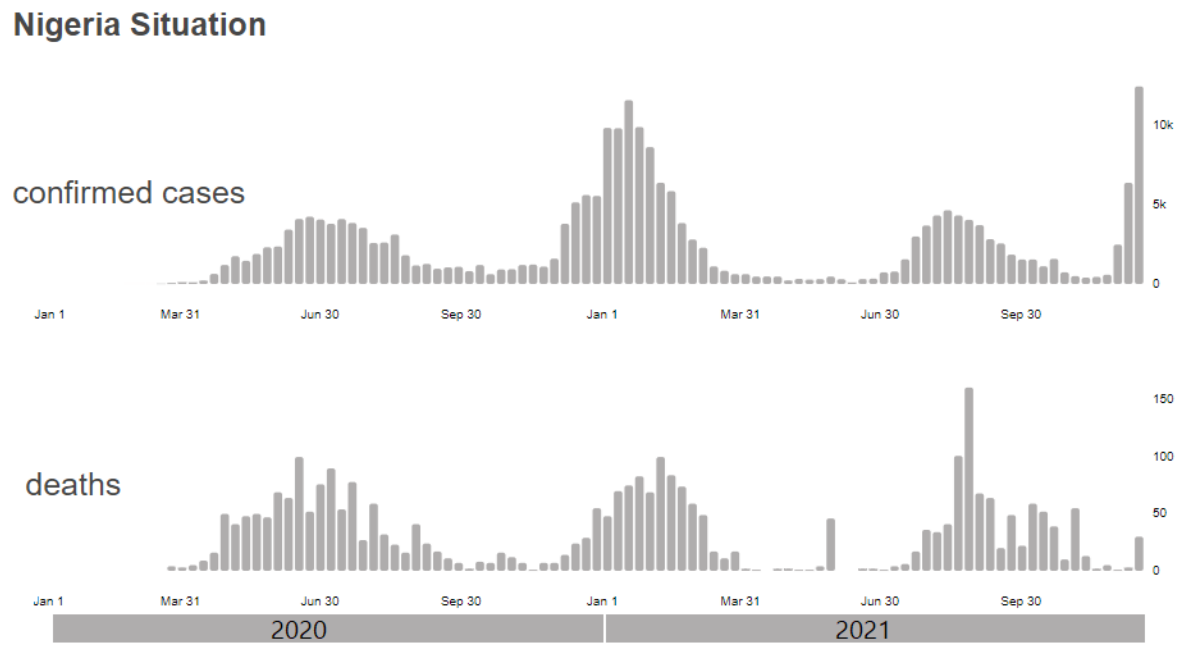
Supplementary Table S2.5. p-values for a decrease in bias between years, reported to 3 significant figures. Comparison years with a significant decrease in bias are in bold face.

Year 1	Year 2	$\Delta b < 0$
2018	2019	$p = 0.138$
2018	**2020**	$\boldsymbol{p = 0.0082}$
2018	**2021**	$\boldsymbol{p = 0.0041}$
2019	2020	$p = 0.108$
2019	2021	$p = 0.169$
2020	2021	$p = 0.727$

Supplementary Table S2.6. MFR and CFR estimates with 95% confidence intervals and probability of CFR being greater than MFR. All reported to 3 significant figures, except for MFR in 2021 which had an estimate and lower confidence bound of less than 10^{-6}.

Year	CFR	MFR	$CFR > MFR$
All years	0.218 [0.206, 0.231]	0.132 [0.113, 0.151]	$p < 0.0001$
2018	0.271 [0.242, 0.300]	0.0862 [0.0177, 0.145]	$p < 0.0001$
2019	0.209 [0.187, 0.233]	0.117 [0.0127, 0.179]	$p < 0.0001$
2020	0.206 [0.187, 0.225]	0.151 [0.127, 0.178]	$p < 0.0001$
2021	0.201 [0.173, 0.231]	<0.0001 (0, 0.106]	$p < 0.0001$

Supplementary Figure S2.6. Weekly confirmed COVID-19 cases in Nigeria. From 01/01/2020 to 02/01/2021, adapted from the WHO COVID-19 dashboard (WHO 2020b). The increase in confirmed cases in January to February 2021 coincides with the usual timing of Lassa fever cases. Available at covid19.who.int/region/afro/country/ng.



Supplementary Table S2.7. Disease incidence estimates from different models of zoonotic exposure and 95% credible intervals, all reported to 3 significant figures.

Total disease incidence (millions)			
	CS1 (contact scales with population density)	CS2 (contact scales logarithmically with population density)	CS3 (contact only occurs in rural populations)
RH1V1 (occurrence-occurrence model)	13.1 [11.6, 14.8]	8.18 [7.25, 9.23]	5.14 [4.55, 5.79]
RH2V1 (abundance-occurrence model)	15.9 [14.1, 17.9]	14.0 [12.4, 15.8]	7.74 [6.86, 8.73]
RH2V2 (abundance-prevalence model)	28.5 [25.3, 32.2]	19.3 [17.1, 21.8]	13.7 [12.1, 15.5]

Appendix 3: Supplementary materials for Chapter 3

Supplementary Table S3.1. Comparison of lowest AIC growth model with that of next lowest AIC. Coefficients and 95% confidence intervals are presented to 3 significant figures. Coefficients of 0 indicate that the covariate is not included in this model process. Since the difference in AIC was less than 2, the models were compared. Because the processes included were the same and the direction of contribution was similar, the model with the lowest AIC was chosen.

Covariates		Coefficients	
		Lowest AIC (AIC=16305.64)	Next-lowest AIC (AIC=16307.49)
Name	Value	Mean growth ($\boldsymbol{\beta}_G$)	
Intercept	1	-49.5 $[-55.3, -43.7]$	-46.9 $[-52.1, -40.2]$
Body weight	w	0.757 $[0.736, 0.777]$	0.751 $[0.732, 0.776]$
Relative population size	$q(t)$	-0.0156 $[-0.0206, -0.0107]$	-0.0204 $[-0.0311, -0.0124]$
Temperature seasonality	$T_{\text{seas}}(t - L_T)$	2.54 $[2.29, 2.80]$	2.45 $[2.17, 2.77]$
Temperature variability	$T_{\text{var}}(t - L_T)$	0	0
Precipitation seasonality	$P_{\text{seas}}(t - L_P)$	-1.38 $[-1.68, -1.08]$	-1.28 $[-1.75, -1.13]$
Precipitation variability	$P_{\text{var}}(t - L_P)$	0	0
Name	Value	Standard deviation of growth	
Intercept	1	$a_G = 1.07$ $[0.984, 1.16]$	$a_G = 1.08$ $[0.985, 1.18]$
Body weight	w	$b_G = 0.0156$ $[0.0130, 0.0182]$	$b_G = 0.0153$ $[0.0128, 0.0183]$

Supplementary Table S3.2. Estimated coefficients for the chosen demographic models, reported to 3 significant figures, with 95% confidence intervals beneath. Coefficients of 0 indicate that the covariate is not included in this model process.

Covariates		Coefficients		
Name	Value	Survival ($\boldsymbol{\beta}_S$)	Recruitment ($\boldsymbol{\beta}_R$)	Mean growth ($\boldsymbol{\beta}_G$)
Intercept	1	-14.3 $[-15.1, -13.5]$	-42.4 $[-53.3, -32.9]$	-49.5 $[-55.3, -43.7]$
Body weight	w	0.0435 $[0.0395, 0.0476]$	0.0260 $[0.00551, 0.0469]$	0.757 $[0.736, 0.777]$
Relative population size	$q(t)$	0	-0.0387 $[-0.0561, -0.0226]$	-0.0156 $[-0.0206, -0.0107]$
Temperature seasonality	$T_{\text{seas}}(t - L_T)$	0.508 $[0.477, 0.539]$	1.53 $[1.18, 1.94]$	2.54 $[2.29, 2.80]$
Temperature variability	$T_{\text{var}}(t - L_T)$	0	0	0
Precipitation seasonality	$P_{\text{seas}}(t - L_P)$	-0.107 $[-0.146, -0.0687]$	0	-1.38 $[-1.68, -1.08]$
Precipitation variability	$P_{\text{var}}(t - L_P)$	-0.0788 $[-0.113, -0.0445]$	0.354 $[0.181, 0.532]$	0
Name	Value			Standard deviation of growth
Intercept	1			$a_G = 1.07$ $[0.984, 1.16]$
Body weight	w			$b_G = 0.0156$ $[0.0130, 0.0182]$

Supplementary Table S3.3. Model validation by fitting to two additional random samples of 80% of the data. Lags and coefficients are given for the three different samples (S_1, S_2, S_3). Where a validation model fitted to sample S_2 or S_3 includes the same covariate as the base model for S_1, the proportion of that coefficient value relative to S_1 is given in brackets to give a clearer metric of similarity. Bold indicates that the validation model produced a value more than 10% different from the base model. Asterisks (*) indicate that the validation model differs from the base model in terms of which covariates are included.

Parameters		Values		
Name	Value	Survival ($\boldsymbol{\beta}_S$)	Recruitment ($\boldsymbol{\beta}_R$)	Mean growth ($\boldsymbol{\beta}_G$)
Temperature lag (days)	L_T	S_1: 63 S_2: 77 S_3: 70	S_1: 154 S_2: 154 S_3: 154	S_1: 161 S_2: 168 S_3: 168
Precipitation lag (days)	L_P	S_1: 98 S_2: 98 S_3: 98	S_1: 63 S_2: 63 S_3: 63	S_1: 133 S_2: 140 S_3: 140
Covariates		Coefficients		
Intercept	1	S_1: -14.3 S_2: -14.5 (1.01) S_3: -14.3 (1.00)	S_1: -42.4 S_2: -38.7 (0.91) S_3: -41.8 (0.99)	S_1: -49.5 S_2: -45.5 (0.92) S_3: -50.3 (1.02)
Body weight (g)	w	S_1: 0.0435 S_2: 0.0420 (0.97) S_3: 0.0432 (0.99)	S_1: 0.0260 S_2: 0.0292 (**1.12**) S_3: 0.0207 (**0.80**)	S_1: 0.756 S_2: 0.748 (0.99) S_3: 0.747 (0.99)
Relative population size	$q(t)$	S_1: 0 S_2: 0.00158 ***** S_3: 0	S_1: -0.0387 S_2: -0.0331 (**0.85**) S_3: -0.0310 (**0.80**)	S_1: -0.0156 S_2: -0.0185 (**1.18**) S_3: -0.0176 (**1.12**)
Temperature seasonality (°C)	$T_{\text{seas}}(t - L_T)$	S_1: 0.508 S_2: 0.525 (1.03) S_3: 0.517 (1.01)	S_1: 1.53 S_2: 1.39 (0.91) S_3: 1.51 (0.99)	S_1: 2.54 S_2: 2.38 (0.94) S_3: 2.60 (1.02)
Temperature variability (°C)	$T_{\text{var}}(t - L_T)$	S_1: 0 S_2: 0 S_3: 0	S_1: 0 S_2: 0 S_3: 0	S_1: 0 S_2: 0 S_3: 0
Precipitation seasonality (mm)	$P_{\text{seas}}(t - L_P)$	S_1: -0.107 S_2: -0.205 (**1.88**) S_3: -0.173 (**1.62**)	S_1: 0 S_2: 0 S_3: 0	S_1: -1.38 S_2: -1.20 (**0.87**) S_3: -1.47 (1.06)

Precipitation variability (mm)	$P_{\text{var}}(t - L_P)$	S_1: -0.0788 S_2: -0.0980 (**1.24**) S_3: -0.0814 (1.03)	S_1: 0.354 S_2: 0.301 (**0.85**) S_3: 0.333 (0.94)	S_1: 0 S_2: 0 S_3: 0
Name	Value			Standard deviation of growth (a_G, b_G)
Intercept	1			S_1: 1.07 S_2: 1.07 (1.00) S_3: 1.08 (1.01)
Body weight (g)	w			S_1: 0.0156 S_2: 0.0154 (0.99) S_3: 0.0156 (1.00)

Supplementary Table S3.3. Significant covariates in the non-demographic models and comparison of error metrics across the non-demographic and climatic-demographic model, all reported to 3 significant figures.

Model	Significant covariates	Variance of residuals	Median squared error	Root mean squared error
(1) Null $\mu = 61.1$ $\sigma = 63.7$	N/A	2.20	0.949	1.48
(2) Linear	T_{seas} $(p < 0.0001)$ T_{var} $(p < 0.0001)$	0.649	0.171	0.804
(3) Linear growth	P_{seas} $(p = 0.0002)$	2.69	0.989	1.64
(4) Exponential growth	T_{seas} $(p = 0.0134)$	15.6	0.867	3.94
(5) Climatic-demographic	See Supplementary Table S3.2	0.949	0.276	0.973

Appendix 4: Supplementary materials for Chapter 4

Supplementary Table S4.1. Collated data on cases from nosocomial transmission, used to estimate incubation period distribution.

Location	Exposure period	Symptom onset period	Incubation period window (days)	Fatal	Reference
Jos, Plateau State, Nigeria	25/01/1970	03/02/1970	9	Yes, 13/02/1970	(Frame *et al.* 1970)
Jos, Plateau State, Nigeria	25/01/1970 – 13/02/1970	20/02/1970	7 – 26	No	(Frame *et al.* 1970)
Zorzor, Lofa County, Liberia	02/03/1972 – 18/03/1972	22/03/1972	4 – 20	No	(Mertens *et al.* 1973)
Zorzor, Lofa County, Liberia	03/03/1972 – 19/03/1972	22/03/1972	3 – 19	No	(Mertens *et al.* 1973)
Zorzor, Lofa County, Liberia	15/03/1972 – 19/03/1972	20/03/1972	1 – 5	Yes, 04/04/1972	(Mertens *et al.* 1973)
Zorzor, Lofa County, Liberia	02/03/1972 – 19/03/1972	20/03/1972	1 – 18	No	(Mertens *et al.* 1973)
Zorzor, Lofa County, Liberia	01/03/1972 – 10/03/1972	13/03/1972	3 – 12	Unknown	(Mertens *et al.* 1973)
Zorzor, Lofa County, Liberia	01/03/1972 – 08/03/1972	17/03/1972	9 – 16	Unknown	(Mertens *et al.* 1973)
Imo State, Nigeria	23/01/1989	29/01/1989	6	Yes, 15/02/1989	(Fisher-Hoch *et al.* 1995)
Imo State, Nigeria	25/02/1989	03/03/1989	6	Yes, 15/03/1989	(Fisher-Hoch *et al.* 1995)
Imo State, Nigeria	25/02/1989	03/03/1989	6	Yes, 15/03/1989	(Fisher-Hoch *et al.* 1995)
Imo State, Nigeria	25/02/1989	07/03/1989	10	No	(Fisher-Hoch *et al.* 1995)
Imo State, Nigeria	25/02/1989	07/03/1989	10	Unknown	(Fisher-Hoch *et al.* 1995)
Abakaliki, Ebonyi State, Nigeria	03/01/2012	08/01/2012	5	Unknown	(Ajayi *et al.* 2013)
Abakaliki, Ebonyi State, Nigeria	03/01/2012	08/01/2012 – 15/01/2012	5 – 12	Unknown	(Ajayi *et al.* 2013)
Abakaliki, Ebonyi State, Nigeria	03/01/2021	08/01/2012 – 15/01/2012	5 – 12	Unknown	(Ajayi *et al.* 2013)
Abakaliki, Ebonyi State, Nigeria	03/01/2021	08/01/2012 – 15/01/2012	5 – 12	Unknown	(Ajayi *et al.* 2013)
Abakaliki, Ebonyi State, Nigeria	03/01/2021	08/01/2012 – 15/01/2012	5 – 12	Unknown	(Ajayi *et al.* 2013)

Tanguiéta, Atakora Department, Benin	07/10/2014 – 14/10/2014	21/10/2014	7 – 14	Yes, date unknown	(Yadouleton *et al.* 2020)
Tanguiéta, Atakora Department, Benin	07/10/2014 – 14/10/2014	23/10/2014	9 – 16	Yes, date unknown	(Yadouleton *et al.* 2020)
Tchaourou, Borgou Department, Benin	03/01/2016	08/01/2016	5	No	(Yadouleton *et al.* 2020)
Tchaourou, Borgou Department, Benin	03/01/2016	12/01/2016	9	No	(Yadouleton *et al.* 2020)
Tchaourou, Borgou Department, Benin	03/01/2016	13/01/2016	10	Yes, date unknown	(Yadouleton *et al.* 2020)
Tchaourou, Borgou Department, Benin	03/01/2016	13/01/2016	10	No	(Yadouleton *et al.* 2020)
Nigeria	28/12/2017	03/01/2018	6	Yes, 14/01/2018	(Dan-Nwafor *et al.* 2019)
Nigeria	28/12/2017	09/01/2018	12	Yes, 14/01/2018	(Dan-Nwafor *et al.* 2019)
Nigeria	28/12/2017	03/01/2018	6	Yes, 04/01/2018	(Dan-Nwafor *et al.* 2019)
Nigeria	28/12/2017	05/01/2018	8	No	(Dan-Nwafor *et al.* 2019)
Sierra Leone	04/11/2019	11/11/2019	7	Yes, date unknown	(ECDC 2019)
Sierra Leone	04/11/2019	11/11/2019	7	No	(ECDC 2019)

Supplementary Text S4.2. Use of a fourth-order numerical solution to estimate the horizontal rodent-to-rodent transmission rate.

This supplementary text details the use of the Runge-Kutta method (RK4) to improve the estimate of horizontal force of infection $F(t)$ on the susceptible population. Fourth order, rather than a lower or higher order, was chosen as a balance between accuracy and simplicity.

Consider a time step t of the IPM simulation. At this time, let the number of susceptible, horizontally-infected, and vertically-infected individuals be S_t, H_t, and V_t respectively. These can be obtained by integrating over the weight distributions at time step t, for example, $H_t = \int_0^\infty D_H(w,t)\,\mathrm{d}w$. Let $K_t(\tau)$ be the number of horizontal infections which occur between time step t and time $t+\tau$, where τ is no more than 1 (i.e., $t+\tau$ is between time step t and $t+1$). We want to find out $K_t(1)$. Since the model assumes that demographic changes take place every 28 days, we will assume that there are no births or deaths which will impact K_t. We also assume that no individuals leave the horizontally-infected compartment or join the vertically-infected compartment, since these changes are also modelled every 28 days. Therefore, we can simplify the SHVR system to a susceptible-infected (SI) model (equations S4.1–S4.3) to describe how $K_t(\tau)$ changes between $\tau = 0$ and $\tau = 1$.

$$\frac{\mathrm{d}S}{\mathrm{d}\tau} = -\beta SI \tag{S4.1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}\tau} = \beta SI \tag{S4.2}$$

$$\frac{\mathrm{d}K_n}{\mathrm{d}\tau} = \beta SI \tag{S4.3}$$

$$S(\tau = 0) = S_n \tag{S4.4}$$

$$I(\tau = 0) = H_n + V_n \tag{S4.5}$$

$$K_t(\tau = 0) = 0 \tag{S4.6}$$

It remains to estimate K_t at time $\tau = 1$ given the initial conditions. Let h be the size of the time step (in this case, 28 days). Using RK4, we can estimate $K_t(\tau = 1)$ based on initial values and estimates of the slope of K_t between time steps t and $t+1$. These slopes are denoted k_1, k_2, k_3, and k_4.

At time step t, the slope of K is equal to the initial value of its derivative, so:

$$k_1 = \beta S_t I_t \tag{S4.7}$$

At the halfway point between time step t and $t+1$ (i.e., $\tau = \frac{1}{2}$), we estimate the new slope based on the change in the force of infection since time step t. Neglecting changes in S and I during this time for any other reason than movement between the two compartments due to disease transmission, at the halfway point we estimate that $I = I_t + \frac{k_1}{2}$ and $S = S_t - \frac{k_1}{2}$, since k_1 is the rate at which infections take place at time t. Therefore, the slope of K_t halfway between time steps t and $t+1$ is estimated as:

$$k_2 = \beta\left(S_t - \frac{k_1}{2}\right)\left(I_t + \frac{k_1}{2}\right) \tag{S4.8}$$

The next slope estimate in RK4 is a second estimate for the slope at the halfway point, this time using k_2 instead of k_1. This is then, similarly:

$$k_3 = \beta\left(S_t - \frac{k_2}{2}\right)\left(I_t + \frac{k_2}{2}\right) \tag{S4.9}$$

Finally, the slope at time step $n+1$ is estimated using k_3. The initial estimates at this point are $I = I_t + k_3$ and $S = S_t + k_3$, so:

$$k_4 = \beta(S_t - k_3)(I_t + k_3) \tag{S4.10}$$

The resulting fourth-order estimate for $K_n(\tau = 1)$, and the estimated number of transmission events between time step t and $t+1$, is

$$E_t = K_t(1) = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{S4.11}$$

In the IPM, not only are the number individuals in different compartments important but also the distribution of weights in these compartments. I therefore implemented horizontal transmission by multiplying the distribution of weights in the susceptible compartment ($D_S(w,t)$) by a force of infection, $F(t)$, assumed to be independent of rodent body weight (equations 4.7a–b). Given that we estimate E_t transmission events to have taken place since the previous time step (equation S4.11), this means that the force of infection on each susceptible individual should be

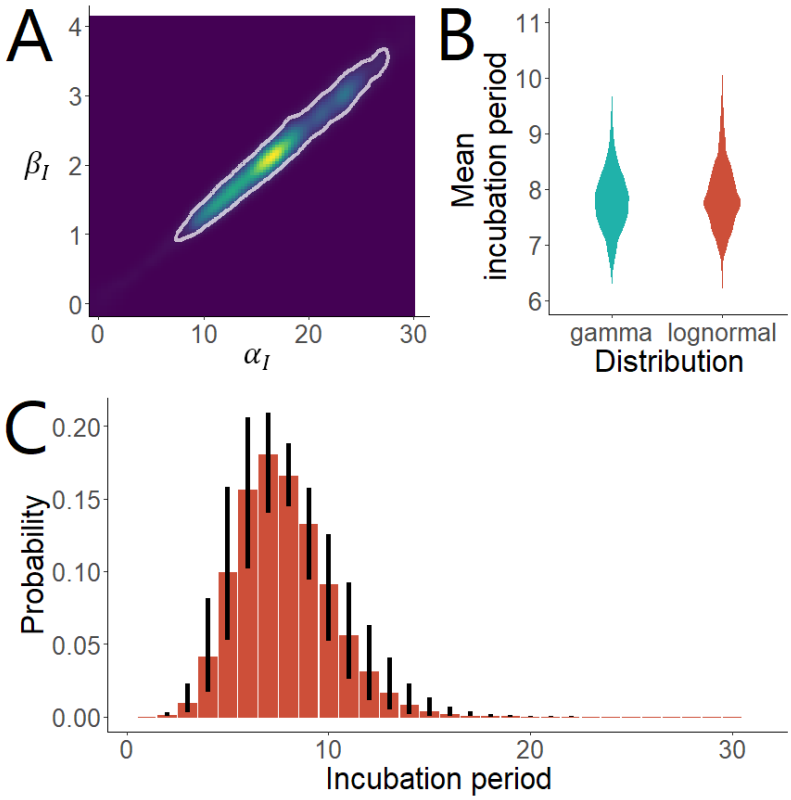
$$F(t) = \frac{E_t}{S_t}, \tag{S4.12}$$

since when multiplied by the distribution of weights in the susceptible compartment, $D_S(w,t)$, and integrated over w, this would result in the correct number of horizontal transmission events.

The procedure for computing the force of infection effective at a given time step t, $F(t)$, is therefore to (1) compute the total infectious and susceptible population size at time t by

integrating (or summing) the relevant distributions over body weight; (2) compute the four slopes in turn (equations S4.7–S4.10); (3) compute the number of transmission events E_t between time step t and $t + 1$ using the RK4 method (equation S4.11); and (4) divide this by the size of the susceptible population at time t, S_t (equation S4.12).

Supplementary Figure S4.3. Credible region for the gamma-distributed incubation period, comparison with the lognormal distribution, and predicted gamma distribution with credible intervals. A: Kernel density estimate for the density in the MCMC chain across parameter space, with lighter colours denoting higher density and 95% credible region enclosed by a white line. B: Kernel density estimate for the mean incubation period obtained by the lognormal and gamma distributions. C: Bar graph showing predicted probability of an incubation period being the given duration. Each line range indicates the 95% credible interval for the probability mass function of that particular duration in isolation.



Supplementary Table S4.4. Estimates for parameter values and derived metrics relating to the lognormal and gamma distributions for incubation period.

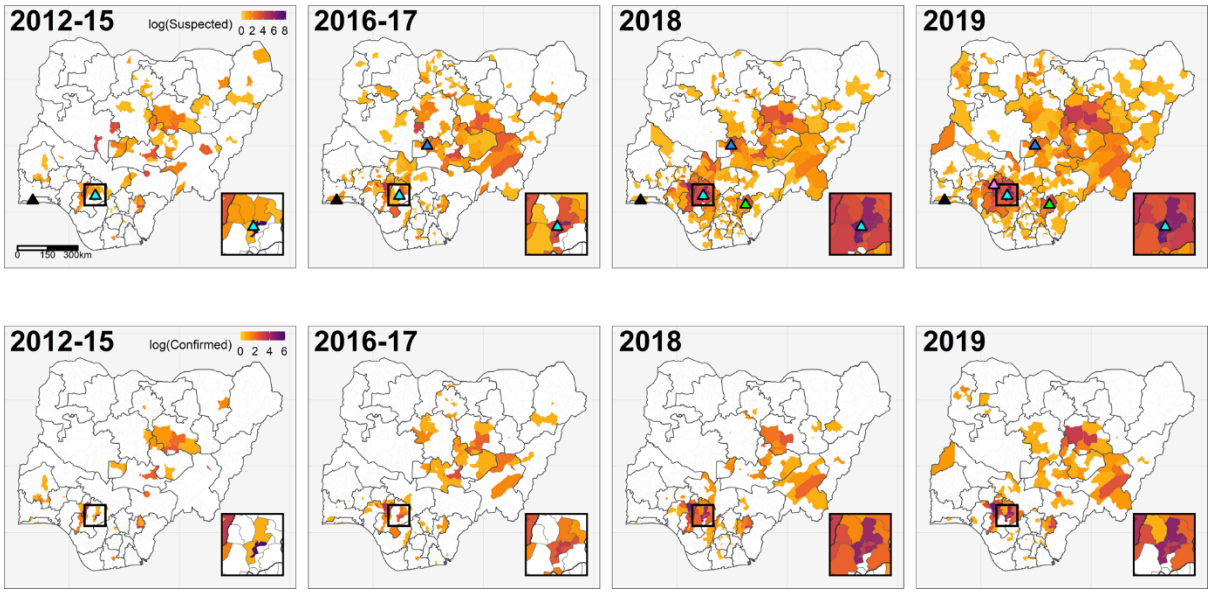
Estimates	**Lognormal distribution**	**Gamma distribution**
Parameter values	$\mu_L = 2.02$ [1.89, 2.15] $\sigma_L = 0.299$ [0.222, 0.420]	$\alpha_I = 16.4$ [9.00, 26.0] $\beta_I = 2.10$ [1.12, 3.39]
Mean incubation period (days)	7.82 [6.95, 9.12]	7.79 [6.84, 8.82]
Median incubation period (days)	7 [6, 8]	7 [6, 8]
Modal incubation period (days)	7 [6, 8]	7 [6, 8]
Central 95% probability range (days)	3 [2, 4] to 13 [11, 18]	3 [3, 4] to 12 [11, 14]
$\mathbb{P}(I > 21 \text{ days})$	0.0248% [0%, 0.856%]	0.00111% [0%, 0.0188%]

Supplementary Table S4.5. Estimates for parameter values and 95% credible intervals in the GLM linking zoonotic hazard with inferred human infections, and the computed p-values for a positive slope. Parameter values are reported to 3 significant figures and credible intervals are reported to 2 significant figures.

State k	a_k	b_k	σ_k	$b_k > 0$
Bauchi	0.679 $[0.43, 0.92]$	0.622 $[0.27, 0.98]$	2.04 $[1.8, 2.3]$	$p = 0.0003$
Ebonyi	-0.145 $[-0.34, 0.051]$	1.16 $[0.87, 1.5]$	1.56 $[1.4, 1.8]$	$p < 0.0001$
Edo	-0.490 $[-1.2, 0.26]$	6.17 $[5.1, 7.3]$	6.10 $[5.2, 6.8]$	$p < 0.0001$
Ondo	-0.0571 $[-0.79, 0.71]$	5.43 $[4.3, 6.5]$	5.96 $[5.1, 6.7]$	$p < 0.0001$
Taraba	0.594 $[0.43, 0.74]$	0.186 $[0.0011, 0.39]$	0.976 $[0.75, 1.1]$	$p = 0.0243$

Appendix 5: Supplementary materials for Chapter 5

Supplementary Figure S5.1. Spatiotemporal trends in Lassa fever surveillance. Reproduced from Redding *et al.* (2021). Maps show total reported Lassa fever cases in each local government authority of Nigeria during the specified year(s) on a logarithmic scale. Top row: Suspected and confirmed cases. Triangles show the locations of laboratories with Lassa fever diagnostic capacity; in addition to two existing diagnostic laboratories in 2012, three additional laboratories became operational in 2017, 2018, and 2019. Bottom row: Laboratory-confirmed cases only. © Redding *et al.*; 2021. Licence: CC BY.



Appendix A: Socioeconomic and geographic factors affecting risk of Lassa fever disease burden in Nigeria

Introduction

Lassa fever is thought to affect some of the poorest communities in Africa due to the presence of the reservoir host in unimproved rural housing (Gibb *et al.* 2017). However, surveillance of Lassa fever is highly spatially heterogeneous, as has been shown for Nigeria (Redding *et al.* 2021; Chapter 2). Meanwhile, there are socioeconomic disparities in access to and engagement with healthcare facilities in Nigeria (Riman & Akpan 2012) which may cause these surveillance heterogeneities, and also decrease likelihood of treatment for Lassa fever. Additionally, pregnancy is a major risk factor for Lassa fever fatality (Kayem *et al.* 2020), and this may also be linked with socioeconomic region or geographic factors. With this analysis I sought to identify whether these factors (pregnancy and access to healthcare) overlap in rural regions which are already predisposed to higher contact rates with the reservoir host.

Methods

I obtained healthcare surveys from 2018 for Nigeria from the Demographic and Health Surveys (DHS) Program (NPC & ICF 2019). Using individual surveys of women aged 15 to 49 years, I sought to answer three questions. Do geopolitical region (north west, north east, north central, south west, south south, and south east) or residence type (urban or rural) impact:

(1) Likelihood of barriers or perceived barriers to healthcare access?
(2) Probability of being pregnant?
(3) Probability of receiving prenatal care?

The interview questions used to inform these risk factors were:

(1) Whether or not the individual interviewed said that each of the following were a "big problem" (as opposed to "not a problem" or "not a big problem") to accessing healthcare; (a) getting permission to go, (b) getting the money needed, (c) distance to a healthcare facility. It was not appropriate to use answers to the question which determined whether or not an individual had accessed healthcare recently, since this would be confounded by individual health.
(2) Whether or not the individual interviewed said that they were currently pregnant.

(3) Out of those who had been pregnant in the last three years, whether or not they received any form of prenatal care during their last pregnancy. Options included traditional care providers outside of formal healthcare facilities.

I removed data relating to women who stated that they were visitors to the area or whom did not complete the full survey, since this might indicate lack of engagement. I modelled all events with generalised linear models (GLMs) with a binomial error structure. I included age bracket (five-year intervals; 15-19, 20-24 etc.) as a covariate to control for demographic effects in addition to region and residence type (Table A1). For question (2) I also fitted additional models with pregnancy status as another covariate, in case pregnant women are more or less likely to feel there are barriers to accessing healthcare.

Table A1. GLMs for healthcare and pregnancy in Nigeria informed by health surveys.

Model number	Outcome variable	Covariates	Sample size
1	Probability of **being pregnant**	Age group Region Residence type	$n = 41{,}615$
2a	Probability of getting **permission** to attend a healthcare facility being a "big problem"	Age group Region Residence type	$n = 41{,}615$
2a*		Age group Region Residence type Current pregnancy status	
2b	Probability of getting the **money** needed to attend a healthcare facility being a "big problem"	Age group Region Residence type	$n = 41{,}615$
2b*		Age group Region Residence type Current pregnancy status	
2c	Probability of the **distance** required to travel to a healthcare facility being a "big problem"	Age group Region Residence type	$n = 41{,}615$
2c*		Age group Region Residence type Current pregnancy status	

3	Probability of having received **no prenatal care** during most recent pregnancy (within the last three years)	Age group Region Residence type	$n = 21{,}682$

Results

Probability of being pregnant, probability of having received no prenatal care, and probability of reporting a "big problem" to healthcare access due to all three of the tested barriers (Figure A1) were significantly increased in rural settings ($p < 0.0001$ for all), and significantly varied between geopolitical regions ($p < 0.0001$ for all). The effect of being in a rural versus urban location was more influential than geopolitical region for likelihood of receiving prenatal care, while these effects were comparable for probability of pregnancy. Being currently pregnant had no significant effect on reporting permission or money being big problems for access to healthcare, however it did have a significant positive on distance to travel being a big problem ($p = 0.047$).

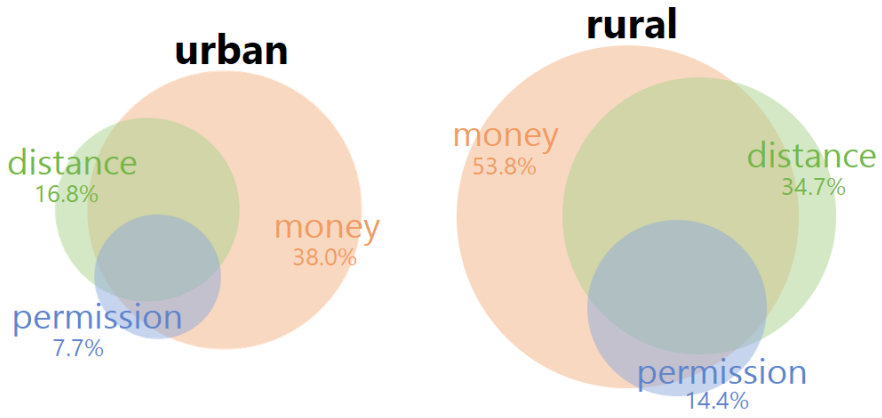


Figure A1. Venn diagrams showing the proportion of women aged 15 to 49 years in urban and rural regions who felt that different factors presented a "big problem" to accessing healthcare.

Conclusions

Exploring reasons behind spatial patterns in access to healthcare across Nigeria is outside the scope of this thesis. However, this analysis indicates that there is geographic variation in healthcare access and pregnancy. Importantly, women in rural areas are both more likely to be pregnant, less likely to feel able to access healthcare, and less likely to receive neonatal care. While this association is important for maternal health generally (Okoli *et al.* 2020), this

is also relevant to the assessment of disease burden of Lassa fever. Given that women in the interviewed age bracket in rural areas are more likely to be pregnant and that pregnancy increases risk of fatality during Lassa fever by a factor of 2.8 (Kayem *et al.* 2020), Lassa fever in rural area versus an urban area may be much more likely to result in fatality. This increased risk of fatality is compounded by the reduced access to healthcare in rural regions and likely increased incidence of Lassa virus infection in rural communities, meaning that the disparity in disease burden of Lassa fever is likely much greater than implied by relative infection incidence alone.

It would be valuable to explore how Lassa fever mitigation measures can be deployed to rural areas; community engagement in continuous rodent control has shown success but only for a limited number of communities (Mari Saez *et al.* 2018). It will be important to ensure Lassa virus vaccines, when developed, are deployed to areas on the basis of high expected disease burden, rather than on the basis of high reported incidence, due to the spatial heterogeneity of surveillance (Chapter 2). However, it is likely a more pressing need to address heterogeneities in access to healthcare in Nigeria, which may be challenging given the political landscape (Innocent *et al.* 2014). It is important to consider that while Lassa fever likely has a large disease burden across West Africa, delivery of improved healthcare, including maternal healthcare, to all rural locations may provide a more effective reduction in overall disease burden than the specific targeting of Lassa fever, since timely provision of healthcare can substantially reduce mortality of Lassa fever (Hallam *et al.* 2018).