Epidemiology and Infection

cambridge.org/hyg

Reconstruction and prediction of viral disease epidemics

M. U. G. Kraemer[1,2,3], D. A. T. Cummings[4,5], S. Funk[6,7], R. C. Reiner Jr.[8], N. R. Faria[3], O. G. Pybus[3] and S. Cauchemez[9,10]

[1]Harvard Medical School, Harvard University, Boston, MA, USA; [2]Computational Epidemiology Lab, Boston Children's Hospital, Boston, MA, USA; [3]Department of Zoology, University of Oxford, Oxford, UK; [4]Department of Biology, University of Florida, Gainesville, Florida, USA; [5]Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA; [6]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK; [7]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK; [8]Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA; [9]Mathematical Modelling of Infectious Diseases and Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France and [10]CNRS UMR2000: Génomique évolutive, modélisation et santé, Paris, France

Review

Cite this article: Kraemer MUG, Cummings DAT, Funk S, Reiner Jr. RC, Faria NR, Pybus OG, Cauchemez S (2019). Reconstruction and prediction of viral disease epidemics. *Epidemiology and Infection* **147**, e34, 1–7. https://doi.org/10.1017/S0950268818002881

Received: 8 May 2018
Revised: 24 August 2018
Accepted: 21 September 2018

Key words:
Geographic spread; real-time prediction; viral epidemics

Author for correspondence:
M. U. G. Kraemer, E-mail: moritz.kraemer@zoo.ox.ac.uk

Abstract

A growing number of infectious pathogens are spreading among geographic regions. Some pathogens that were previously not considered to pose a general threat to human health have emerged at regional and global scales, such as Zika and Ebola Virus Disease. Other pathogens, such as yellow fever virus, were previously thought to be under control but have recently re-emerged, causing new challenges to public health organisations. A wide array of new modelling techniques, aided by increased computing capabilities, novel diagnostic tools, and the increased speed and availability of genomic sequencing allow researchers to identify new pathogens more rapidly, assess the likelihood of geographic spread, and quantify the speed of human-to-human transmission. Despite some initial successes in predicting the spread of acute viral infections, the practicalities and sustainability of such approaches will need to be evaluated in the context of public health responses.

Introduction

Infectious disease outbreaks pose a significant threat to human health. The frequency of such outbreaks is thought to have increased over the past decade. For example, quickly after an epidemic of Ebola virus affected Guinea, Sierra Leone, and Liberia in 2013–2016 [1], chikungunya virus (CHIKV) caused an extensive international epidemic in the Americas and beyond, and was quickly followed by Zika virus (ZIKV) emergence. To date, there have been more than 500 000 confirmed or probable cases of ZIKV but the true number of cases remains unknown [2]. Yellow fever (YF), a vaccine-preventable disease, recently posed major public health problems. In 2015–2016, the largest YF outbreak since the 1980s was observed in Angola and the Democratic Republic of the Congo, causing 962 confirmed cases and 393 deaths [3]. YF also poses an ongoing public health risk to large, urban and under-vaccinated populations in the coastal areas of southern Brazil, a country that successfully eradicated YF in the 1950s and 1960s [3–5]. Examples of other emerging pathogens that have caused international health security concerns include the severe acute respiratory syndrome (SARS) virus and the Middle East Respiratory Syndrome Coronavirus (MERS-Cov) [6–9]. This list extends to other pathogens such as influenza, Nipah and henipaviral diseases, and Lassa fever [10]. These examples show the continued risks that infectious diseases pose and highlight the challenges of large international outbreaks to epidemic planning and response.

During emerging infectious disease outbreaks, empirical information and mathematical modelling techniques are now commonly used to characterise and predict the spatio-temporal dynamics of the spread of pathogens. Such analyses may help policymakers to evaluate the threat to public health, determine the resources required to reduce disease burden, and guide disease surveillance efforts and the deployment of interventions.

In the last decade, our ability to perform such assessments has been improved by advances in a number of disciplines, including digital disease surveillance [11], environmental modelling [12, 13], genomics [14] and mathematical modelling [15]. For example, environmental variables such as rainfall and precipitation [13, 16–22] can be used to better understand the landscape within which the disease may be transmitted, and detailed transmission data from a small sampled population can be extrapolated to larger, un-surveyed areas [23]. Attempts have been made to illustrate the spatial structure of epidemics mainly using human movement data [24–27], to provide mechanistic insights in how the disease may disperse locally [3, 28, 29] or how effective reactive vaccination campaigns may be [30–32]. There are continued efforts to reconstruct epidemic dynamics using information derived from

© The Author(s) 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

pathogen genomic data, which contain unique information about the history of transmission [2, 33–35]. Although each of these disciplines has an established relationship to disease prevention and control, the benefits of integrating them into a unified framework have yet to be fully achieved.

Here we describe the common applications and models used to predict acute viral diseases and discuss the current challenges and limitations. We then outline the advantages of integrating disparate data sources to advance our understanding of epidemic spread. We discuss how such research has been used in recent outbreaks and outline shortcomings that may be addressed in the future.

Reconstruction of transmission pathways using genomic data

Phylogenetic and phylodynamic tools are increasingly being used to infer a range of outbreak properties [36]. Common spatio-temporal analyses of pathogen genomes focus on mapping and predicting virus lineage exchange among locations, with the underlying aim of reconstructing the pathways of disease introduction and spread, albeit at a coarse spatial resolution, and often retrospectively [2, 8, 33, 35, 37, 38]. An additional feature that can be inferred from genomic data is the timing of individual *founder introductions* [39]. Blue arrows in Fig. 1 indicate the time when the first report was published inferring the likely geographic origin of four major international infectious disease outbreaks. Phylogenetic tools can help to characterise the number of introductions that lead to disease transmission in a new location [41], quantify the risk of cross-species transmission [42], and infer ecological drivers of transmission [43, 44]. Genome-derived estimates have been compared qualitatively to those from epidemiological data, but formal model-based integration of both data sources are rare [45, 46]. In principle, pairing genomic information with epidemiological inference should enable us to quantify the number of cases missed in each location and help to estimate parameters such as the basic reproductive number and doubling time of the epidemic, as done for ZIKV at the tail end of the epidemic (Fig. 1a) [46–48]. A common limitation when genetic data are used is the absence of a rigorous and formal sampling scheme. In many instances, genomic sampling is affected by convenience and expedience and may not reflect underlying incidence, although this can be improved post-hoc in large data sets via sub-sampling using, for example continuous phylogenetic inference [49–51]. Strong sampling biases may affect estimates of the arrival time of a pathogen and its pathways of dissemination among locations [33].

Prediction of disease spread using spatial information

Static disease mapping is a powerful tool to visualise and defines the landscape within which transmission occurs, based on ecological drivers of transmission [17, 18, 22]. When combined with global data on human travel and mobility, it can be used to understand the global dynamic risk surface of infectious disease, especially when there are strong ecological determinants of transmission, as there are for the vector-borne diseases Zika, dengue, chikungunya and YF [27, 52]. Publication of reports that estimate geographic spread for the diseases in Fig. 1 are indicated by green arrows. The global epidemic history of Zika, for example, remains poorly understood. The challenge to accurately reconstruct the epidemic pathway of the virus is further complicated by its relatively unspecific clinical presentation. This may explain why the initial studies that aimed to understand the geographic origin of the Zika epidemic in the Americas were published relatively late into the epidemic (>1 year, Fig. 1a). For the other major outbreaks highlighted in Fig. 1, estimates of the geographic origin were documented between 6 and 8 months after the first reports of human cases (Fig. 1b–d; Table 1). However, given the underlying ecological determinants of transmission that restrict the reproduction of the virus in the mosquito vector species, large areas can be excluded from the risk of local virus transmission. When overlaying information on the reported presence of Zika cases *vs.* the underlying ecological risk map, surveillance gaps may be identified [19, 27]. Areas where there is a mismatch in the predicted presence and reported presence (i.e. cases detected) should be targeted for active surveillance.

The spatial spread process of new pathogens, however, is not only determined by the underlying ecological determinants in each location but also by the dynamic nature of importation, often driven by human movements [61]. Spatial models that take into account the patterns of human spread and mobility may, therefore, improve our ability to characterise and anticipate spatial expansion. Different models have been proposed to predict the geographic spread of epidemics but rarely have they been used in real time during the course of an epidemic [3, 62] (Fig. 1). For example, during the YF outbreak in Angola and the Democratic Republic of the Congo, estimates of geographic spread to provinces outside Luanda, the capital of Angola, were published >6 months after the last cases were reported (Fig. 1c). Such information could guide public health institutions to decide where and when to implement surveillance and control programs [27]. More work, however, is needed to dynamically map the spread of infectious diseases and to extract meaningful and interpretable quantities for public health practitioners.

Characterising disease dynamics and transmission clusters

In parallel to these efforts to model the spread of pathogens at a meta-population level (e.g. between cities, regions, countries or continents), we also need to better understand transmission dynamics at a much more granular level and assess the characteristics of the inter-human transmission. While historically, the potential for inter-human transmission has often been summarised with a single statistic; the reproduction number R (i.e. the average number of secondary infections generated by a case). However, it has long been recognised that it is also essential to assess heterogeneities in individual R values, since the presence of super-spreading events may have a major impact on the risk of emergence and our ability to control outbreaks [63]. This was exemplified in a large MERS-CoV outbreak in South Korea in 2015 in which only a small number of cases were responsible for the majority of infections [64, 65]. Other factors that may drive the spatial differences in the reproductive number are ecological (population density, climatic factors, or others) and can now be readily incorporated in transmission models [66].

Ideally, these assessments should be performed on detailed data documenting chains of transmission, as such data can provide precise quantification of the transmission potential and the impact of targeted interventions in different settings and over time, and allow testing specific hypotheses about the transmission process (e.g. what is the contribution of re-introductions to the overall dynamic?) [67]. However, such data are rarely available

https://doi.org/10.1017/S0950268818002881 Published online by Cambridge University Press

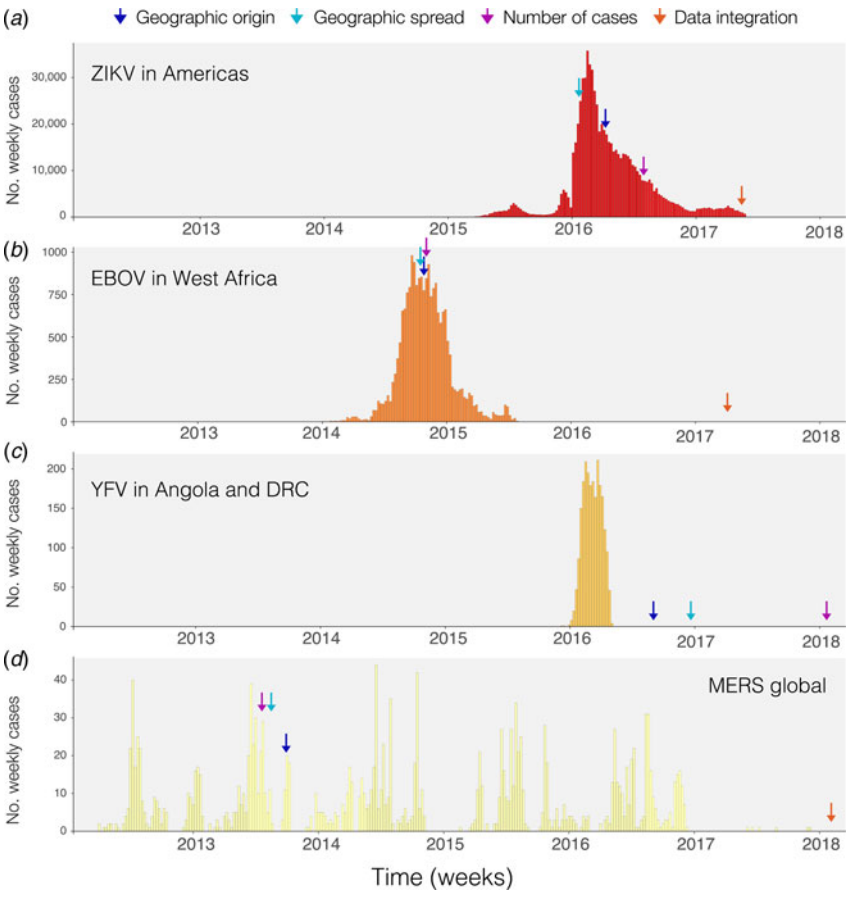


Fig. 1. Timing of publications addressing key questions during outbreaks. Blue shows the first peer-reviewed publication identifying the geographic origin of the outbreak, green shows the date predictions about geographic spread are published, purple shows the date when predictions of numbers of cases are made and red indicates the date when work on the integration of geographic, genomic and epidemiological data was published. (a) Shows weekly cases of the 2014–2017 Zika virus epidemic in the Americas using data from [33, 38] and the Pan American Health Organization (PAHO) available from https://github.com/andersen-lab/Zika-cases-PAHO. (b) Shows weekly cases from the West African Ebola epidemic published by the World Health Organization (WHO). (c) Shows weekly cases of the 2015–2016 Yellow fever epidemic in Angola and the Democratic Republic of Congo, published by WHO [40]. (d) Shows weekly cases from 2012 to 2017 Middle Eastern Respiratory Syndrome outbreak available from https://github.com/rambaut/MERS-Cases.

as it is difficult to identify the source of infection for most pathogens. As a result, sophisticated statistical techniques are often required to reconstruct chains of transmission and estimate transmission parameters from more limited data that may include: (i) in the context of zoonoses, the size of human clusters [68–70] or the proportion of surveillance cases that reported a contact with the natural reservoir [71], (ii) the growth rate in the case count [72–75], (iii) partial data on chains of transmission [76], or (iv) outbreak data where the timing of symptom onsets and location of cases are recorded in small communities such as households [77–80], schools [81] or villages [82]. In cases of high-density sampling, genomic data can help to reconstruct transmission chains [83].

Disease outbreak modelling

Mechanistic models of infectious disease dynamics can be used to make predictions about the future course of an outbreak within a given location [84]. Increasingly, such models are being used in real time, such that predictions are updated every time a new data point becomes available [85, 86]. Some other applications track pathogen evolution over time as data become available [87]. However, the perceived ability of such models to successfully or unsuccessfully make 'correct' predictions can generate considerable controversy [88, 89]. There are few studies that systematically investigate forecasting accuracy and its relationship to the length of time that is being predicted and to the quantity and quality of data available [90, 91]. Other examples are forecasting challenges for ongoing epidemics such as CHIKV in the Americas (https://www.darpa.mil/news-events/2014-08-15), EVD in West Africa [92] and seasonal influenza [85, 93], designed and initiated by funding agencies and public health governments. This is an important area for future research.

Integrating data sources to better characterise infectious disease dynamics

There are clear benefits to combining information from different data sources in order to better predict viral epidemic spread. Previous work most commonly presents estimates from different sources side-by-side, for example, estimates of the epidemic reproductive number derived from genomic *vs.* epidemiological data [46]. Such comparisons are important to assess the consistency of data sources and may help to derive new hypotheses. Spurred by technological innovation such as portable sequencing using the MinION device (Oxford Nanopore Technologies, Oxford, UK) [94] and by interdisciplinary collaborations during disease outbreaks, researchers have started to work to combine three types of transmission data: spatial, genomic and epidemiological which have now been published for three of the four major outbreaks we considered here (Fig. 1, red arrows) [33, 38, 41].

For example, such interdisciplinary work helped to identify the introduction of Zika into the Americas [2], investigated the main drivers of transmission of ZIKV through climatic suitability of its mosquito vectors [25] and tried to extrapolate how many people had been infected with the virus [23, 92, 93]. In the context of phylogenetic analyses, environmental and other spatial data may be helpful in reconstructing the drivers of transmission

https://doi.org/10.1017/S0950268818002881 Published online by Cambridge University Press

Table 1. Key dates and publications describing the geographic origin and spread of four major international outbreaks prediction of the expected number of cases, and integration of geographical, epidemiological and genetic data

Characteristic of outbreak	Date published (online)	Citation
Zika virus outbreak in the Americas (2014–2017)		
Geographic origin	24 March 2016	Faria *et al.*, 2016 [2]
Geographic spread	14 January 2016	Bogoch *et al.*, 2016 [52]
Number of cases	25 July 2016	Perkins *et al.*, 2016 [23]
Integration of genomic/epidemiological/geographical data	17 May 2016	Grubaugh *et al.*, 2017 [33] and Faria *et al.*, 2017[38]
Ebola virus disease outbreak in West Africa (2013–2016)		
Geographic origin	12 September 2016	Gire *et al.*, 2014 [53]
Geographic spread	2 September 2014	Gomes *et al.*, 2014 [54]
Number of cases	8 September 2016	Fisman *et al.*, 2014 [55]
Integration of genomic/epidemiological/geographical data	12 April 2017	Dudas *et al.*, 2017 [35]
Yellow fever outbreak in Central Africa (2015–2016)		
Geographic origin	20 September 2016	Grobbelaar *et al.*, 2016 [56]
Geographic spread	22 December 2016	Kraemer *et al.*, 2016 [3]
Number of cases	16 January 2018	Zhao *et al.*, 2018 [57]
Integration of genomic/epidemiological/geographical data	NA	NA
Middle East Respiratory Syndrome outbreak (2012–2017)		
Geographic origin	20 September 2013	Cotten *et al.*, 2013 [58]
Geographic spread	17 July 2013	Khan *et al.*, 2013 [59]
Number of cases	5 July 2013	Breban *et al.*, 2013 [60]
Integration of genomic/epidemiological/geographical data	16 January 2018	Dudas *et al.*, 2018 [42]

and spread using, for example, information on the reservoir or host movements [35, 95]. In turn, phylogenetic information may complement epidemiological analysis by providing more evidence on the transmission routes that are common in an outbreak [96]. This may be particularly useful for diseases that have a highly structured transmission dynamic, such as MERS or SARS, where a small number of people are responsible for the majority of secondary cases [63, 97], transmission from the animal reservoir is frequent, or importation drives locally observed epidemics [33]. One common assumption in many epidemiological models is that it is equally likely for people to meet and infect others living in the same location and that population immunity is proportional to the demographic structure [98]. Hence, observed cases are often assumed to arise from other cases that are reported locally as long as they are consistent with the generation time of the disease. However, a well-connected location can, in principle, accrue a large number of incident cases through recurring introductions from elsewhere, rather than via local transmission [33]. These results can have large implications for surveillance and control, as different competing strategies (e.g. limiting importations or eradicating the disease locally) may be considered. While analytical approaches of various degrees of complexity have been proposed to probabilistically reconstruct transmission trees from incomplete outbreak data [73, 81, 97], contact tracing, which can be very labour intensive [67], remains a gold standard information source. This may allow us to is to determine the true distribution of cluster sizes (i.e. the number of subsequent cases resulting from each introduction) but is often only available for a small number of locations. However, using genomic data can help refine the understanding of heterogeneity in transmission but such framework does not yet allow to exactly quantify the fraction of observed cases that are attributable to local transmission versus introduction from elsewhere, or to determine how many importations are responsible for the local incidence, despite its crucial importance for eradication campaigns [42, 100, 101]. In the context of the Zika outbreak in Florida, combining genomic data from the outbreak with epidemiological analysis revealed that the outbreak was driven by a large number of introductions rather than by persistent local transmission. In the recent yellow fever outbreak in southern Brazil, linking epidemiological, spatial and genomic data and techniques could provide insights into the transmission potential and risk of urban transmission [102]. One dataset and analysis alone would have not been strong enough to make such conclusions [33].

Inferences about epidemic processes made using mathematical models rely on a number of assumptions. Geographic modelling approaches, mostly informed by spatial ecology, attempt to fill gaps where no data has been observed, hence inferences may be uncertain, as the underlying ecological process may be poorly understood and dynamical aspects of the invasion process are ignored. These deficiencies can be ameliorated, in part, by adding virus genome data that contain information about past transmission and invasion patterns [103]. However, due to incomplete and poor sampling (as discussed above), genomic data alone may provide an incomplete picture of the timing of viral introduction and spread among locations. This, in turn, can be supported by the addition of epidemiological time series of reported cases

https://doi.org/10.1017/S0950268818002881 Published online by Cambridge University Press

and serological information about population immunity [104, 105]. Despite this, building a joint inference framework that combines all available data sources and which characterises observation and sampling processes correctly is a daunting task. However, we are entering a period where the data for this task are becoming available in a timely fashion but need to ensure that results are communicated as soon as they are generated in order to avoid delays. Initial successes have already led to important advancements in epidemic control and should progress to a tool-kit for guiding public health, hopefully available in real time for future epidemics.

Author ORCIDs. M. U. G. Kraemer http://orcid.org/0000-0001-8838-7147

Acknowledgements. MUGK is supported by The Branco Weiss Fellowship – Society in Science, administered by the ETH Zurich and acknowledges funding from a Training Grant from the National Institute of Child Health and Human Development (T32HD040128) and the National Library of Medicine of the National Institutes of Health (R01LM010812, R01LM011965). OGP acknowledges support from the Oxford Martin School. SC acknowledges financial support from the Investissement d'Avenir program, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (Grant ANR-10-LABX-62-IBEID), the Models of Infectious Disease Agent Study of the National Institute of General Medical Sciences, the AXA Research Fund. NRF is supported by a Sir Henry Dale Fellowship (204311/Z/16/Z), internal GCRF grant 005073 and John Fell Fund grant 005166. SF was supported by a Wellcome Trust Senior Research Fellowship in Basic Biomedical Science (210758/Z/18/Z).

References

1. **WHO Ebola Response Team** (2014) Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections. *The New England Journal of Medicine*. Published online: 22 September 2014. doi:10.1056/NEJMoa1411100.
2. **Faria NR *et al.*** (2016) Zika virus in the Americas: early epidemiological and genetic findings. *Science* **352**, 345–349.
3. **Kraemer MUG *et al.*** (2017) Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. *The Lancet Infectious Diseases* **17**, 330–338.
4. **Ministerio da Saude (Brazil)** (2018) http://portalarquivos2.saude.gov.br/images/pdf/2018/janeiro/30/af-informe-febre-amarela-11-30jan18.pdf.
5. **Soper FL** (1963) The elimination of urban yellow fever in the Americas through the eradication of Aedes aegypti. *American Journal of Public Health* **53**, 7–16.
6. **Brockmann D and Helbing D** (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science (New York, N.Y.)* **342**, 1337–1342.
7. **Neumann G, Noda T and Kawaoka Y** (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459**, 931–939.
8. **Lemey P *et al.*** (2014) Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathogens* **10**, e1003932.
9. **Khan K *et al.*** (2009) Spread of a novel influenza A (H1N1) virus via global airline transportation. *The New England Journal of Medicine* **361**, 212–214.
10. **Friedrich MJ** (2018) Who's blueprint list of priority diseases. *JAMA* **319**, 1973.
11. **Brownstein JS, Freifeld C and Madoff LC** (2009) Digital disease detection – harnessing the web for public health surveillance. *The New England Journal of Medicine* **360**, 2153–2157.
12. **Pigott DM *et al.*** (2016) Updates to the zoonotic niche map of Ebola virus disease in Africa. *eLife* **5**, e16412.
13. **Pigott DM *et al.*** (2015) Mapping the zoonotic niche of Marburg virus disease in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**, 366–378.
14. **Gardy JL and Loman NJ** (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews. Genetics* **19**, 9–20.
15. **Heesterbeek H *et al.*** (2015) Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, aaa4339.
16. **Bhatt S *et al.*** (2013) The global distribution and burden of dengue. *Nature* **496**, 504–507.
17. **Kraemer MUG *et al.*** (2016) Progress and challenges in infectious disease cartography. *Trends in Parasitology* **32**, 19–29.
18. **Nsoesie EO *et al.*** (2015) Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. *Eurosurveillance* **21**, 30234.
19. **Messina JP *et al.*** (2016) Mapping global environmental suitability for Zika virus. *eLife* **5**, e15272.
20. **Pigott DM *et al.*** (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**, e04395.
21. **Mylne AQN *et al.*** (2015) Mapping the zoonotic niche of Lassa fever in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**, 483–492.
22. **Shearer FM *et al.*** (2018) Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *The Lancet Global Health* **6**, e270–e278.
23. **Perkins TA *et al.*** (2016) Model-based projections of Zika virus infections in childbearing women in the Americas. *Nature Microbiology* **1**, 16126. doi:10.1038/nmicrobiol.2016.126.
24. **Rocklöv J *et al.*** (2016) Assessing Seasonal Risks for the introduction and Mosquito-borne spread of Zika virus in Europe. *EBioMedicine* **9**, 250–256.
25. **Bogoch II *et al.*** (2016) Potential for Zika virus introduction and transmission in resource limited countries in Africa and Asia-Pacific. *The Lancet Infectious Diseases* **16**, 1237–1245.
26. **Bogoch II *et al.*** (2014) Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 West African outbreak. *The Lancet* **6736**: 1–7.
27. **Kraemer MUG *et al.*** (2018) Zika virus transmission in Angola and the potential for further spread to other African settings. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **111**, 527–529.
28. **Johansson Ma *et al.*** (2014) Nowcasting the spread of chikungunya virus in the Americas. *PLoS ONE* **9**, e104915.
29. **Cauchemez S *et al.*** (2014) Local and regional spread of chikungunya fever in the Americas. *Eurosurveillance* **19**, 20854.
30. **Garske T *et al.*** (2014) Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Medicine* **11**, e1001638.
31. **Jentes ES *et al.*** (2011) The revised global yellow fever risk map and recommendations for vaccination, 2010: consensus of the informal WHO working group on geographic risk for yellow fever. *The Lancet Infectious Diseases* **11**, 622–632.
32. **Shearer FM *et al.*** (2017) Global yellow fever vaccination coverage from 1970 to 2016: an adjusted retrospective analysis. *The Lancet Infectious Diseases* **17**, 1209–1217.
33. **Grubaugh ND *et al.*** (2017) Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546**, 401–405.
34. **Ladner JT *et al.*** (2015) Evolution and spread of Ebola virus in Liberia, 2014–2015. *Cell Host and Microbe* **18**, 659–669.
35. **Dudas G *et al.*** (2017) Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315.
36. **Grenfell BT *et al.*** (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332.
37. **Pybus OG, Tatem AJ and Lemey P** (2015) Virus evolution and transmission in an ever more connected world. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20142878.
38. **Faria NR *et al.*** (2017) Epidemic establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546**, 406–410.
39. **Lemey P *et al.*** (2009) Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5**, e1000520. doi:10.1371/journal.pcbi.1000520.
40. **World Health Organization** (2016) Situation report. Yellow fever. 28 October 2016. World Health Organization, Geneva; 2016 http://apps.who.int/iris/bitstream/10665/250661/1/yellowfeversitrep28Oct16-eng.pdf?ua=1 (Accessed Dec 12, 2016).

https://doi.org/10.1017/S0950268818002881 Published online by Cambridge University Press

41. **Faria NR *et al.*** (2017) Genomic and epidemiological characterisation of a dengue virus outbreak among blood donors in Brazil. *Scientific Reports* **7**, 1–12.
42. **Dudas G *et al.*** (2018) MERS-CoV spillover at the camel-human interface. *eLife* **7**, e31257.
43. **Dellicour S, Rose R and Pybus OG** (2016) Explaining the geographic spread of emerging viruses: a new framework for comparing viral genetic information and environmental landscape data. *BMC Bioinformatics* **17**, 82. https://doi.org/10.1186/s12859-016-0924-x.
44. **Dellicour S *et al.*** (2017) Using viral gene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics. *Molecular Biology and Evolution* **34**, 2563–2571.
45. **Cauchemez S *et al.*** (2014) Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *The Lancet Infectious Diseases* **14**, 50–56.
46. **Fraser C *et al.*** (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* **324**, 1557–1562.
47. **Stadler T *et al.*** (2012) Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Science of the USA* **110**, 228–233.
48. **Stadler T *et al.*** (2014) Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. *PLoS Currents Outbreaks* 1–18.
49. **Bromham L *et al.*** (2018) Bayesian molecular dating: opening up the black box. *Biological Reviews* **93**, 1165–1191.
50. **Lemey P *et al.*** (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* **27**, 1877–1885.
51. **Dellicour S *et al.*** (2018) Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. *Nature Communications* **9**, Article number: 2222. Published online: 2018. doi:10.1038/s41467-018-03763-2.
52. **Bogoch II *et al.*** (2016) Anticipating the international spread of Zika virus from Brazil. *The Lancet* **387**, 335–336.
53. **Gire SK *et al.*** (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372.
54. **Gomes MFC *et al.*** (2014) Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLOS Currents Outbreaks* Edition 1. doi:10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5.
55. **Fisman D, Khoo E and Tuite A** (2015) Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model. *PLOS Currents Outbreaks* Edition 1. doi:10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571.
56. **Grobbelaar AA *et al.*** (2016) Resurgence of yellow fever in Angola, 2015–2016. *Emerging Infectious Diseases* **22**, 1854–1855.
57. **Zhao S *et al.*** (2018) Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLoS Neglected Tropical Diseases* **12**, 1–24.
58. **Cotten M *et al.*** (2013) Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *Lancet* **382**, 1993–2002.
59. **Khan K *et al.*** (2013) Potential for the international spread of Middle East respiratory syndrome in association with Mass Gatherings in Saudi Arabia. *PLOS Current Outbreaks*. Published online: 2013. Edition 1. doi:10.1371/currents.outbreaks.a7b70897ac2fa4f79b59f90d24c860b8.
60. **Breban R, Riou J and Fontanet A** (2013) Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *Lancet* **382**, 694–699.
61. **Kraemer MUG, Bisanzio D and Reiner RC** (2018) Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about intra-urban human mobility. *EPJ Data Science* **7**, 16. https://doi.org/10.1140/epjds/s13688-018-0144-x.
62. **Kramer AM *et al.*** (2016) Spatial spread of the West Africa Ebola epidemic. *Royal Society Open Science* **3**, 160294.
63. **Lloyd-Smith JO *et al.*** (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359.
64. **Su S *et al.*** (2015) MERS in South Korea and China: a potential outbreak threat? *The Lancet* **385**, 2349–2350.
65. **Cho SY *et al.*** (2016) MERS-CoV outbreak following a single patient exposure in an emergency room in South Korea: an epidemiological. *The Lancet Elsevier Ltd*, **388**, 994–1001.
66. **Kraemer MUG *et al.*** (2015) Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *Journal of the Royal Society Interface* **12**, 20150468.
67. **Faye O *et al.*** (2015) Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *The Lancet Infectious Diseases* **15**, 320–326.
68. **Ferguson NM *et al.*** (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214.
69. **Blumberg S and Lloyd-Smith JO** (2013) Inference of R0 and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Computational Biology* **9**, 1–17.
70. **Blumberg S, Funk S and Pulliam JRC** (2014) Detecting differential transmissibilities that affect the size of self-limited outbreaks. *PLoS Pathogens* **10**, e1004452. https://doi.org/10.1371/journal.ppat.1004452.
71. **Cauchemez S *et al.*** (2013) Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: application to the emergence of US swine origin influenza A H3N2v virus. *PLoS Medicine* **10**, e1001399. https://doi.org/10.1371/journal.pmed.1001399.
72. **Wallinga J and Lipsitch M** (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of The Royal Society B* **274**, 599–604.
73. **Wallinga J and Teunis P** (2004) Similar impacts of control measures. *American Journal of Epidemiology* **160**, 509–516.
74. **Cori A *et al.*** (2013) Practice of epidemiology a new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* **178**, 1505–1512.
75. **Cauchemez S *et al.*** (2006) Real-time estimates in early detection of SARS. *Emerging Infectious Diseases* **12**, 110–113.
76. **Agua-Agum J *et al.*** (2016) Exposure patterns driving Ebola transmission in West Africa: a retrospective observational study. *PLoS Medicine* **13**, 1–23.
77. **Cauchemez S *et al.*** (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine* **23**, 3469–3487.
78. **Cauchemez S *et al.*** (2009) Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine* **361**, 2619–2627.
79. **Tsang TK *et al.*** (2014) Association between antibody titers and protection against influenza virus infection within households. *Journal of Infectious Diseases* **210**, 684–692.
80. **Tsang TK *et al.*** (2016) Household transmission of influenza virus. *Trends in Microbiology* **24**, 123–133.
81. **Cauchemez S *et al.*** (2011) Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2825–2830.
82. **Salje H *et al.*** (2016) How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. *Proceedings of the National Academy of Science of the USA* **113**, 13420–5.
83. **Poon LLM *et al.*** (2016) Quantifying influenza virus diversity and transmission in humans. *Nature Genetics* **48**, 195–200.
84. **Chretien J, Riley S and George DB** (2016) Mathematical modeling of the West Africa Ebola epidemic. *eLife* **4**, e09186.
85. **Andronico A *et al.*** (2017) Real-time assessment of health-care requirements during the Zika virus epidemic in Martinique. *American Journal of Epidemiology* **186**, 1194–1203.
86. **Camacho A *et al.*** (2015) Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLOS Currents Outbreaks* **7**, 1–28.
87. **Hadfield J *et al.*** (2017) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, https://doi.org/10.1093/bioinformatics/bty407.
88. **Rivers C** (2014) Ebola: the power of behaviour change. *Nature* **515**, 492–492.

https://doi.org/10.1017/S0950268818002881 Published online by Cambridge University Press

89. **Meltzer MI *et al.*** (2014) Estimating the future number of cases in the Ebola epidemic – Liberia and Sierra Leone, 2014–2015. *MMWR supplements* **63**, 1–14.
90. **Held L, Meyer S and Bracher J** (2017) Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture. *Statistics in Medicine* **36**, 3443–3460.
91. **Lowe R *et al.*** (2016) Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. *eLife* **5**, 1–18.
92. **Viboud C *et al.*** (2018) The RAPIDD Ebola forecasting challenge special issue: preface. *Epidemics Elsevier*, **22**, 1–2.
93. **Biggerstaff M *et al.*** (2018) Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics* **24**, 26–33. Published online: 2018. doi:10.1016/j.epidem.2018.02.003.
94. **Quick J *et al.*** (2016) Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232.
95. **Thézé J *et al.*** (2018) Genomic epidemiology reconstructs the introduction and spread of Zika virus in Central America and Mexico. *Cell Host and Microbe* **23**, 855–864.e7.
96. **Baele G *et al.*** (2017) Emerging concepts of data integration in pathogen phylodynamics. *Systematic Biology* **66**, e47–e65.
97. **Cauchemez S *et al.*** (2016) Unraveling the drivers of MERS-CoV transmission. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 9081–9086.
98. **Anderson RM and May RM** (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford, UK: Oxford University Press.
99. **Reiner RC *et al.*** (2015) Mapping residual transmission for malaria elimination. *eLife* **4**, e09520.
100. **Lau MSY *et al.*** (2017) Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proceedings of the National Academy of Sciences* **114**, 2337–2342.
101. **Churcher TS *et al.*** (2014) Measuring the path toward malaria elimination. *Science* **344**, 1230–1232.
102. **Faria NR *et al.*** (2018) Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* **361**, 894–899.
103. **Messina JP *et al.*** (2014) Global spread of dengue virus types: mapping the 70 year history. *Trends in Microbiology* **22**, 138–146.
104. **Metcalf CJE, Ferrari M and Grenfell BT** (2015) Understanding herd immunity. *Trends in Immunology* **36**, 753–755.
105. **Metcalf CJE *et al.*** (2016) Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *The Lancet* **6736**, 5–7.

https://doi.org/10.1017/S0950268818002881 Published online by Cambridge University Press