

Scoping future outbreaks: a scoping review on the outbreak prediction of the WHO Blueprint list of priority diseases

Nils Jonkmans,¹ Valérie D'Acremont,^{1,2} Antoine Flahault³

To cite: Jonkmans N, D'Acremont V, Flahault A. Scoping future outbreaks: a scoping review on the outbreak prediction of the WHO Blueprint list of priority diseases. *BMJ Global Health* 2021;**6**:e006623. doi:10.1136/bmjgh-2021-006623

Handling editor Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2021-006623>).

Received 15 June 2021

Accepted 1 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

²Swiss Tropical and Public Health Institute, Basel, Switzerland

³Institute of Global Health, Faculty of Medicine, Université de Genève, Geneva, Switzerland

Correspondence to

Mr Nils Jonkmans;
nils.jonkmans@unil.ch

ABSTRACT

Background The WHO's Research and Development Blueprint priority list designates emerging diseases with the potential to generate public health emergencies for which insufficient preventive solutions exist. The list aims to reduce the time to the availability of resources that can avert public health crises. The current SARS-CoV-2 pandemic illustrates that an effective method of mitigating such crises is the pre-emptive prediction of outbreaks. This scoping review thus aimed to map and identify the evidence available to predict future outbreaks of the Blueprint diseases.

Methods We conducted a scoping review of PubMed, Embase and Web of Science related to the evidence predicting future outbreaks of Ebola and Marburg virus, Zika virus, Lassa fever, Nipah and Henipaviral disease, Rift Valley fever, Crimean-Congo haemorrhagic fever, Severe acute respiratory syndrome, Middle East respiratory syndrome and Disease X. Prediction methods, outbreak features predicted and implementation of predictions were evaluated. We conducted a narrative and quantitative evidence synthesis to highlight prediction methods that could be further investigated for the prevention of Blueprint diseases and COVID-19 outbreaks.

Results Out of 3959 articles identified, we included 58 articles based on inclusion criteria. 5 major prediction methods emerged; the most frequent being spatio-temporal risk maps predicting outbreak risk periods and locations through vector and climate data. Stochastic models were predominant. Rift Valley fever was the most predicted disease. Diseases with complex sociocultural factors such as Ebola were often predicted through multifactorial risk-based estimations. 10% of models were implemented by health authorities. No article predicted Disease X outbreaks.

Conclusions Spatiotemporal models for diseases with strong climatic and vectorial components, as in River Valley fever prediction, may currently best reduce the time to the availability of resources. A wide literature gap exists in the prediction of zoonoses with complex sociocultural and ecological dynamics such as Ebola, COVID-19 and especially Disease X.

INTRODUCTION

In 2015, the member states of the WHO produced a Research and Development Blueprint list of priority diseases, diseases

Key questions

What is already known?

- The Blueprint list denotes diseases with the potential to cause severe public health emergencies for which there is an urgent need for accelerated research and development.
- Outbreak prediction has previously been applied with success to diseases such as Rift Valley fever and influenza, for prevention and the pre-emptive implementation of health measures.
- A systematic review of the outbreak prediction methods of the Blueprint diseases does not exist.

What are the new findings?

- Explicit predictions of timing and locations of future outbreaks are most often carried out for diseases with strong climatic components such as Rift Valley fever.
- The current literature in outbreak prediction of Blueprint diseases can be categorised into five domains: spatiotemporal modelling and risk mapping, time series forecasting and regression analysis, internet-based computing and phone-based predictions, narrative and qualitative models, and quantitative probabilistic models.
- No articles in this review predicted an outbreak of a novel Disease X in 2019, including coronaviruses.

What do the new findings imply?

- Prediction models of diseases with strong climatic components such as for Rift Valley fever may currently be most appropriate for policymakers in making explicit prediction statements.
- Predictions based on outbreak receptivity and risk maps for diseases such as Ebola, Lassa and Crimean-Congo haemorrhagic fever identify outbreak risk factors targetable by policymakers.
- A pressing need exists for research investment in outbreak prediction of Blueprint zoonoses and especially unknown future pathogens such as Disease X in order to tackle new pathogens such as SARS-CoV-2.

for which, 'given their potential to cause a public health emergency and the absence of efficacious drugs and/or vaccines, there is an urgent need for accelerated research and

Box 1 2018 WHO Blueprint list of priority diseases

- Ebola virus disease: the Ebola virus outbreak in West Africa in 2014–2016 primarily affected Sierra Leone, Guinea and Liberia. This outbreak caused 28 639 deaths, a loss of 2.2 billion dollars, a substantial loss in private and public sector growth, agriculture production, food security concerns and restrictions of movement, goods and services.^{70 71}
- Zika virus disease: Zika virus has infected millions in the Americas since 2014 and has caused an increase in medical sequelae in some populations (congenital disease, Guillain-Barré) as well as socioeconomic disparities.^{72 73}
- River Valley fever (RVF): RVF outbreaks in South and Eastern Africa impacted the economy and public health of multiple countries, leaving behind long-term societal consequences.⁷⁴
- Lassa fever: Lassa fever, a severe haemorrhagic fever transmitted through rats, is estimated to infect millions in West Africa each year, with 20% of those patients experiencing severe multisystemic disease.⁷⁵
- Nipah and henipaviral disease: Nipah virus, discovered in 1998, is isolated to Malaysia, Singapore, Bangladesh and India. However, Nipah continues to pose a significant threat as a bat-based zoonosis with yearly spillover events, significant morbidity and case fatality rates of up to 100%.⁷⁶
- Crimean-Congo haemorrhagic fever: CCHF, a widely distributed tick-borne zoonosis, infects domestic and wild vertebrates as hosts, resulting in severe human disease and high mortality, and poses a continued threat to central Africa, South-Western Russia and central Asia.⁷⁷
- Severe acute respiratory syndrome (SARS) due to SARS-CoV-1: a coronavirus causing SARS emerged in 2003 and rapidly spread from Southeast Asia to Canada. SARS causes severe atypical pneumonia, leading to high morbidity, mortality and major economic consequences.⁷⁸
- Middle East respiratory syndrome (MERS): another coronavirus, MERS-CoV, emerged in the Arabian Peninsula in 2012 and has been found in Europe, North America and Asian countries with a mortality of about 30%.⁷⁹
- Disease X: disease X is a term used to 'enable cross-cutting R&D preparedness that is relevant for currently unknown diseases'.¹¹ Disease X was included to stimulate research into emerging pathogens, before their formal discovery.

development'.¹ The list, updated in 2018, includes eight emerging pathogens (box 1). These diseases are RNA viruses with moderate to severe case fatality rates and the potential for outbreaks of severe economic and health consequences.^{2–8} The BP diseases are zoonotic pathogens, which constitute about 60% of emerging infections in humans.⁹ Emerging infectious disease outbreaks are incited by an acceleration of urban development, socioeconomic disparities and an encroachment of the natural reservoir, and the number of such outbreaks is predicted to increase.¹⁰

While the Blueprint (BP) list 'does not aim to predict the next epidemic',¹¹ its aim is to further the ability 'to reduce the time lag between the identification of a nascent outbreak and approval of the most advanced products that can be used to save lives and stop larger crises'.¹ The most effective method to reduce the time

lag between an outbreak and implementation of public health measures may be to anticipate future outbreaks. Analysis of viral outbreaks is complex, frequently relying on a mixture of ground-level epidemiological data, advanced mathematical and statistical analysis and the inherent stochasticity of disease processes.¹² While the complexity and randomness in epidemics render prediction challenging, the inherent entropy barrier to prediction is 'beyond the timescale of single outbreaks, implying [outbreak] prediction is likely to succeed'.¹³ Furthermore, challenges intrinsic to outbreak prediction should not discourage us from attempting to pre-empt outbreaks. Preventative efforts would involve public health measures such as vaccination, pesticide use and awareness sensitisation.¹⁴ COVID-19, similar in certain respects to Middle East respiratory syndrome (MERS) and Severe acute respiratory syndrome (SARS), has brought with it the consequences of a severe global pandemic and demonstrates the need for establishing a research and development pipeline to predict the next Disease X.^{15 16} As the BP diseases lack a dedicated research pipeline, we are ill-equipped to deal with outbreaks of these priority pathogens.¹⁷ The COVID-19 pandemic underlines that pre-emptive implementation of preventive public health measures in areas at risk of future outbreak is of social, health and economic importance. Influenza also demonstrates this point with a successful history of surveillance, forecasting and epidemic modelling enabling prediction of future epidemics, which has guided pre-emptive vaccination efforts and other preparedness measures.^{18 19}

We therefore undertook a scoping review to identify and characterise studies attempting to predict future outbreaks of the BP diseases. A scoping review was selected as the preferred method, as it enables a broader search strategy and research question, which is useful in an ill-defined field of research.²⁰ Furthermore, it allows integration of new findings and research developments during the review process. A frequent goal of scoping reviews is also to map the existing literature, without drawing far-reaching conclusions. Our objective was thus twofold. First, to map the literature concerning outbreak prediction: what aspects of future outbreaks were predicted and what methods and data sources were used to make these predictions. Second, we attempted to synthesise the data in a manner that would be useful for policymakers, in order to inform and highlight prediction methods that may warrant future exploration and implementation.

Aims and research question

The aim of this scoping review was to evaluate and map the scientific literature and identify research gaps concerning outbreak prediction methods of the BP diseases. The question this scoping review will attempt to answer is: what methodologies, model types and data are used to predict future outbreaks? Furthermore, what scholarly consensus exists on predictive models for future outbreaks and how can these be leveraged by current

health actors in pre-empting BP diseases, including COVID-19?

METHODS

The Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist for scoping reviews was adopted.²¹ A scoping review is a form of ‘knowledge synthesis, [that] follow[s] a systematic approach to map evidence on a topic and identify main concepts, theories, sources, and knowledge gaps’.²¹ The main author developed the scoping review protocol, the eligibility criteria and the summative data extraction tables.

Literature search strategy

The search strategy was developed by the main author with the help of research staff at the Lausanne University Hospital Library in Switzerland and included a broad range of terms related to outbreak prediction and the blueprint diseases through a combination of free text and Medical Subject Headings (MeSH) terms. The search terms were designed to identify literature that related the prediction of future outbreaks to the blueprint diseases while minimising the identification of literature related to the modelling of current epidemics. Prediction as we define is the attempt to foresee outbreaks of a disease in a location or timeframe in the future during a non-epidemic epidemiological situation. Disease-related search terms were also identified using MeSH terms and their catalogued synonyms from the National Library of Medicine database.²² Disease, forecasting and prediction related search terms were then combined and run in advanced search settings in the respective databases. The full search strategy is detailed in the supplementary material.

Databases

Three databases were used to identify relevant literature: PubMed, Web of Science and Embase. Grey literature was not searched. No hand searching was performed. No date limit was applied. Articles were searched up to and including the 4 July 2019. This temporal limitation was applied as we sought to focus on the state of research imminently preceding COVID-19 and to exclude COVID-19’s influence on outbreak modelling of the Blueprint diseases.

Screening, study selection, and inclusion and exclusion criteria

Relevant literature identified through our search strategy was extracted from the aforementioned databases, then transferred to EndNote X9 (Clarivate Analytics) for deduplication. Following deduplication, preliminary article selection was carried out through a modified two-reviewer screening system using Rayyan, an online-based literature screening platform. The main author screened titles and abstracts of all articles identified according to the eligibility criteria below and labelled articles as

‘included’, ‘excluded’ or ‘maybe’. During screening, full-text articles were occasionally retrieved and reviewed if the title or abstract did not provide enough information to decide whether it ought to be included or excluded. A second researcher aided with reviewing articles labelled as ‘included’ and ‘maybe’ and provided expertise in the analysis of statistical/mathematical models. Where conflict between article selection existed, both reviewers debated the articles according to inclusion/exclusion criteria until consensus was reached.

Inclusion criteria

The review considered any studies predicting or forecasting future outbreaks through prediction of outbreak timing and/or location, outbreak risk maps, qualitative or quantitative outbreak risk, other risk assessments and other future epidemiological outbreak phenomena. Prediction in the chosen articles was defined either as an explicit statement by the authors stating ‘we believe at X time and/or Y location in the future, an outbreak of Z disease will occur’ or models containing either a quantitative (eg, percentage likelihood and numerical scale) or qualitative (eg, highly likely vs unlikely) risk of outbreak during a timeframe and/or location in the future. We required studies to denote some form of outbreak prediction in the abstract and title. We also included studies that predicted a future outbreak without mentioning a specific date/size of the outbreak/phenomena. We did not set a threshold degree of certainty for the predictions included. The review considered original quantitative and qualitative studies. There were no restrictions with regard to geographic location, population or study design.

Exclusion criteria

Articles that were excluded contained one or more of the following criteria:

- ▶ Reviews, editorials, viewpoints and letters, duplicate studies and literature with a strong veterinarian focus not linked to public health.
- ▶ Studies solely modelling current outbreaks of Blueprint diseases at the time of publishing, without predicting future phenomena.
- ▶ Studies solely predicting outbreak risk factors.
- ▶ In vivo and in vitro basic science models (eg, vaccine trials and animal models).
- ▶ Purely descriptive epidemiological and ecological publications (eg, serological studies and risk factors) without prediction of future epidemiological changes.
- ▶ Models that only examined causality of Blueprint diseases, rather than estimating risk or burden.
- ▶ Languages other than English, Spanish, French or German.
- ▶ Portable Document Format (PDF) not accessible.

Data extraction, synthesis and abstraction

After initial screening and study selection using Rayyan, the selected list of articles (n=123) was transferred to

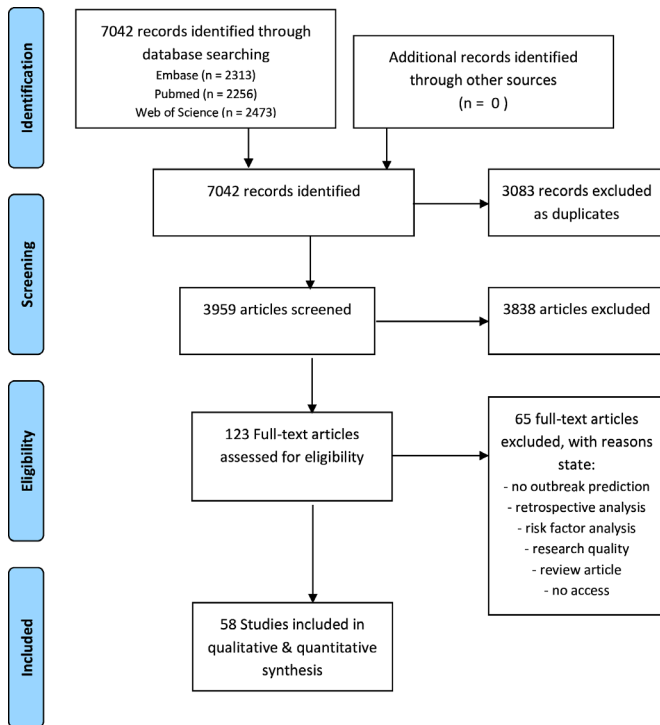


Figure 1 PRISMA flow diagram of search strategy. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Papers for literature management, full-text retrieval and data extraction. A primary readthrough of articles was conducted, and data were extracted into a descriptive summative table synthesising study information (online supplemental material 1). Study information included purpose of study, prediction method and key findings answering to scoping question, among other variables (online supplemental material 1). During this full-text analysis, a further 65 articles were removed according to the exclusion criteria. The final list consisted of 58 articles (figure 1). A second round of article readthroughs was then undertaken in order to synthesise and categorise data quantitatively into a numeric table (online supplemental material 1). Definitions of the various variables extracted and the complete quantitative and narrative analyses for each article are presented in the supplementary material.

A summative table was created highlighting the principal findings (table 1).

RESULTS

The database search identified 7042 articles. A total of 3083 articles were excluded as duplicate studies. Titles and abstracts of the remaining 3959 articles were screened for inclusion. One hundred and twenty-three abstracts met the inclusion criteria. Sixty-five articles were further excluded on full-text analysis. In total, 58 articles were retained (figure 1).

Most publications concerned RVF (36%), Zika (22%) and CCHF (14%). No publications were produced on

Disease X. Ninety-six per cent of articles were published between 2000 and 2019. The most studied region was the African continent (48%). Outbreak prediction and forecasting strategies were mapped into five categories (table 2): (1) spatiotemporal modelling (43%) and risk mapping (45%), (2) time series forecasting (40%) and regression analysis (36%), (3) internet-based computing and phone-based systems (10%), (4) qualitative models (12%) and (5) other quantitative models (16%). Certain articles fell into multiple categories when prediction strategies were combined. Most model types were stochastic (60%) in nature. The most common data types used in predictions were case count (81%), climate/meteorological (67%), vector (53%) and sociodemographic data (41%). Future outbreaks were most commonly predicted by evaluating outbreak risk (62%), spatial (76%) and/or temporal predictions (67%). Future case numbers were predicted in 36% of studies, and 64% of articles concomitantly evaluated outbreak risk factors. A significant portion of articles studied environmental suitability to future outbreaks (34%). Of note, few articles evaluated climate change effects on outbreaks (5%). A synthetic table of the outbreak prediction and forecasting methods highlights our principal findings (table 1).

Spatiotemporal modelling and risk mapping

The most common outbreak prediction methods were risk mapping (45%) and spatiotemporal modelling (43%). Most often, disease–environment relationships predicted future outbreak location through climatic colayers and vector–host data.²³ Many articles (n=16) applied machine learning algorithms (maximum entropy and boosted regression trees) integrating climatic, socioeconomic, ecological and transportation data into niche models.²⁴ Qualitative models such as analytical hierarchy process were also applied synthesising scientific literature to create RVF risk maps.²⁵ Overall, aspects frequently predicted were outbreak timing and location, environmental transmission risk/susceptibility, outbreak hotspots, predictability of outbreaks, future outbreak risk and risk factors. Data types most often integrated were outbreak and case count, climatic, ecological and vector/host data. Risk mapping was frequently used for diseases such as RVF and Zika. Models of this domain were often proposed or even implemented as early warning systems based on climatic anomaly surveillance.

Statistical analysis: time series and regression models

Statistical analysis in prediction and forecasting most often used time series (40%) and regression analysis (36%). Most often, models input past case count and risk factors and output a numeric value representing prospective case count. Analytical tools such as ARIMA, Generalized Additive Mixed and Markov switching models were used in predictions relating seasonality to incidence. For example, one approach coupled 12

Table 1 Principal findings of outbreak prediction articles, by disease

| | All diseases | Rift Valley fever | Zika disease | CCHF | Ebola and Marburg disease | Lassa fever | MERS | Nipah and Henipa virus | SARS |
|------------------------------|--------------|-------------------|--------------|--------|---------------------------|-------------|-------|------------------------|-------|
| Number of articles (%) | 58 (100) | 21 (36) | 13 (22) | 8 (14) | 6 (10) | 5 (9) | 4 (7) | 2 (3) | 2 (3) |
| Date range | | | | | | | | | |
| 2010–2019 | 47 (81) | 15 | 13 | 6 | 5 | 4 | 4 | 2 | |
| 2000–2009 | 9 (16) | 4 | | 2 | 1 | 1 | | | 2 |
| 1990–1999 | 2 (3) | 2 | | | | | | | |
| Region of study | | | | | | | | | |
| African continent | 28 (48) | 18 | | 1 | 6 | 5 | | | |
| Asia-Pacific | 4 (7) | | 3 | | | | | 1 | |
| Europe | 3 (5) | 1 | 1 | 1 | | | | 1 | |
| Middle East | 9 (16) | 2 | | 6 | | | 1 | | |
| North America | 2 (3) | | 1 | | | | 1 | | |
| Latin America and Caribbean* | 5 (9) | | 5 | | | | | | |
| Global | 8 (14) | 1 | 3 | | | | 2 | | 2 |
| Prediction methodology | | | | | | | | | |
| Risk mapping | 26 (45) | 14 | 6 | 1 | 3 | 2 | | 1 | 1 |
| Regression model | 21 (36) | 7 | 5 | 4 | 3 | 1 | 1 | | |
| Time series forecasting | 23 (40) | 9 | 5 | 4 | 2 | 1 | 2 | | |
| Qualitative | 7 (12) | 4 | | 2 | 2 | 1 | | 1 | |
| Other quantitative | 9 (16) | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 |
| Species niche model | 15 (26) | 3 | 4 | 1 | 5 | 3 | | 1 | |
| Machine learning | 16 (28) | 3 | 6 | 2 | 2 | 2 | | 1 | |
| Spatiotemporal model | 25 (43) | 13 | 4 | 4 | 2 | 2 | | | |
| Internet/phone/computer† | 6 (10) | | 5 | | | | 1 | | |
| Early warning system** | 17 (29) | 7 | 5 | 3 | 1 | | 1 | | |
| Incidence modelling | 11 (19) | | 5 | 5 | | | 1 | | |
| Model type | | | | | | | | | |
| Deterministic | 6 (10) | 2 | 1 | | | 1 | 1 | 1 | |
| Stochastic | 35 (60) | 11 | 9 | 6 | 3 | 3 | 2 | 1 | |
| Mixed | 6 (10) | 1 | 1 | 1 | 2 | 1 | | | 2 |
| Not applicable/not stated | 11 (19) | 7 | 2 | 1 | 1 | | 1 | | |
| Data sources | | | | | | | | | |
| Case data | 47 (81) | 12 | 12 | 7 | 6 | 5 | 3 | 2 | 2 |
| Other patient health data | 13 (22) | | | | | | | | |
| Meteorological/climate | 39 (67) | 19 | 7 | 5 | 4 | 3 | | 1 | |
| Vector/host | 31 (53) | 13 | 7 | 4 | 5 | 3 | | 1 | |
| Sociodemographic | 24 (41) | 7 | 7 | 3 | 3 | 2 | 2 | 1 | 1 |
| Behaviour (way of infection) | 8 (14) | 2 | | 1 | 1 | 1 | 2 | 1 | |

Continued

Table 1 Continued

| | All diseases | Rift Valley fever | Zika disease | CCHF | Ebola and Marburg disease | Lassa fever | MERS | Nipah and Henipa virus | SARS |
|---|--------------|-------------------|--------------|------|---------------------------|-------------|------|------------------------|------|
| Healthcare | 5 (9) | 1 | 2 | 1 | 1 | 1 | 1 | | |
| Transportation | 12 (21) | 2 | 4 | 1 | 2 | 2 | | 1 | 2 |
| Internet† | 7 (12) | 1 | 5 | | | | 1 | | |
| Geographical | 32 (55) | 15 | 13 | 6 | 5 | 4 | 4 | 2 | |
| Economic | 9 (16) | 2 | 4 | 1 | 1 | 2 | | 1 | |
| Ecological | 18 (31) | 9 | 3 | 2 | 4 | 2 | | | |
| Expert opinion | 5 (9) | 4 | | 1 | | 1 | | | |
| Other‡ | 6 (10) | 1 | 1 | 1 | 3 | 2 | | | |
| Prediction outcome | | | | | | | | | |
| Future cases | 21 (36) | 1 | 9 | 5 | | 1 | 3 | | 2 |
| Outbreak risk factors | 37 (64) | 19 | 4 | 7 | 6 | 3 | 1 | | |
| Immunity parameters§ | 7 (12) | 2 | 1 | | 1 | | 1 | | 2 |
| Risk maps | 29 (50) | 4 | | 2 | 1 | 1 | | | 2 |
| Spatial prediction | 44 (76) | 19 | 10 | 5 | 4 | 4 | 1 | 1 | 2 |
| Temporal prediction | 39 (67) | 15 | 7 | 6 | 4 | 2 | 3 | 1 | 1 |
| Outbreak risk | 36 (62) | 14 | 9 | 3 | 5 | 3 | 2 | 2 | 1 |
| Spillover events | 6 (10) | 2 | | 2 | 3 | 2 | | | |
| Bio-Env-Econ consequences¶ | 4 (7) | 3 | 1 | 1 | | | | | |
| Env transmission suitability | 20 (34) | 10 | 4 | 2 | 4 | 2 | | 1 | |
| Population at risk | 8 (14) | | 3 | 2 | 2 | 3 | | | |
| Introduction risk | 5 (9) | 1 | 2 | 2 | 1 | 1 | | 1 | |
| Effect of climate change | 3 (5) | 2 | | | | | | | |
| Epidemic dynamics | 17 (29) | 3 | 4 | 4 | 2 | 1 | 2 | | 1 |
| Implementation of prediction/methods by decision makers | | | | | | | | | |
| Yes | 6 (10) | 4 | 1 | | 1 | | | | |
| Suggested | 30 (52) | 10 | 4 | 6 | 4 | 4 | 3 | 1 | 1 |
| No | 22 (38) | 7 | 8 | 2 | 1 | 1 | 1 | 1 | 1 |
| Predictions validated against future outbreak data | | | | | | | | | |
| Yes | 24 (41) | 10 | 3 | 4 | 3 | 1 | | 1 | 2 |
| No | 34 (59) | 11 | 10 | 4 | 3 | 4 | 4 | 1 | |

For detailed definitions, see online supplementary material.

*Includes South and Middle America.

†Internet and phone-based system/app/computer programme.

‡Non-categorisable data types.

§Reproduction number (R value).

¶Biological, environmental or economic consequences.

**Or proposed Early Warning System.

CCHF, Crimean-Congo haemorrhagic fever; MERS, Middle East respiratory syndrome; SARS, severe acute respiratory syndrome.

years of dengue census data as Zika case surrogate data onto climate and demographic colayers in order to estimate future Zika incidence.²⁶ CCHF and Zika predictions often applied the aforementioned methodologies. Time series of climatic risk factors, coupled to

incidence, were also used to forecast future outbreak location and timing, case count and epidemic dynamics. For example, a distributed lag non-linear model was used to associate meteorological factors to outbreaks and predict outbreaks with a time lag of 20 weeks.²⁷

Table 2 Main outbreak prediction model themes

| | References |
|---|---|
| <i>Spatiotemporal modelling</i> : analysis of data collected over space and over time. ⁸⁰ | 23–27 29 31 38 42–49 51–57 61 81–94 |
| <i>Risk mapping</i> : a visual representation of risk variation of particular disease processes over a set of spatial units. | |
| <i>Regression analysis</i> : the effect of one or several explanatory variables (eg, exposures, subject characteristics, risk factors) on a response variable such as mortality or cancer. ⁹⁵ | 25–27 31 42–46 48 49 54 57–61 64 65 84 86–89 91–94 98–101 |
| <i>Time series analysis</i> : an analysis on the basis of the fact that ‘data points taken over time may have an internal structure (autocorrelation, trend or seasonal variation) that should be accounted for’. ⁹⁶ Prediction of future events by analysing the trends of the past, on the assumption that future trends will hold similar to historical trends. ⁹⁷ | |
| <i>Other probabilistic/quantitative models</i> : modelling relationships between parameters through equations and/or numerical data. | 28–30 35 45 50 51 62 90 |
| <i>Computing and internet systems</i> : models using computing systems (phone, computer-based) coupled to retrieval of internet data. | 36 37 56 59 64 65 |
| <i>Other qualitative outbreak prediction</i> : models using descriptive, non-quantitative methods (eg, expert opinion) to infer relationships and/or causality between two parameters. | 32–34 42 50 51 54 81 91 |

Quantitative outbreak models

Quantitative risk models (16%) were conducted through a variety of methodologies. A probabilistic model based on worldwide incidence, transportation data, probability of arrival of infected travellers and entomological field data was used to estimate future outbreak likelihood in a large European city.²⁸ A SARS metapopulation model assessed worldwide transportation networks to establish a global, between-country quantitative outbreak likelihood scale.²⁹ Another SARS study coupled transportation data onto a Susceptible-Exposed-Infectious-Recovered (SEIR) framework to estimate future incidence.³⁰ A Zika model applied an ecological study design, using socioeconomic data as a surrogate for unprotected sex, to establish locations of future outbreaks should a Zika introduction event occur.³¹ Quantitative aspects that were predicted included risk of outbreak, total population at risk of disease and projected epidemic size.

Qualitative outbreak models

A few articles evaluated future disease outbreaks through various qualitative model types (12%). One article employed a qualitative risk assessment using Delphi technique to elicit expert opinion as to the EU outbreak risk of CCHF and RVF.³² A study in Saudi Arabia analysed serological data to assume immunity level, exposure risk and descriptively infer RVF outbreak risk.³³ Another article used field epidemiological methods (carcass detection) to set up an ‘Outbreak Alarm Network’ to warn of impending outbreaks. Specifically, Ebola positivity of simian carcasses was communicated to local health centres, predicting outbreaks and improving preparedness.³⁴ Another article evaluated binary Lassa fever risk to populations through a machine learning model by weighing different predictor variables.³⁵ In this category, a frequent data type employed was expert opinion. Furthermore, ground level epidemiological data (carcasses and host immunity) was also used to understand the environmental susceptibility to outbreaks and thus estimate future outbreak risk.

Computing and internet systems

A relatively small number of articles (10%) proposed computer or phone-based internet systems designed to alert application users to new outbreaks and signal health authorities to case build-ups pre outbreak. All models were applied to Zika disease except one, which was applied to MERS. Data integrated into phone applications consisted of user autoreported personal health data (eg, symptoms), patient data from health institutions and internet-based geospatial information (eg, location of nearby mosquito breeding sites, infected persons or risk factors).³⁶ The system would then communicate directly to the user/patient and the healthcare actors the outbreak risks or the patient’s own likely infection status. Other early warning systems used Google trend search term time series as surrogates for epidemiological data and coupled this to hospital and public health records in order to predict outbreak location and time.³⁷ App-based models were based on theoretical exercises and synthetic data, while Google trend studies mostly used real data. Aspects predicted often included infected user location and movement in real time, and case numbers.

Prediction validation and implementation

Forty-one per cent of articles validated their predictions against real data. Seventy-four per cent of articles cited challenges or limitations to their studies. Only 10% of prediction methods or forecasts were implemented in real public health or outbreak settings by decision makers or policymakers.

DISCUSSION

This scoping review aimed to map the available evidence concerning the prediction of future outbreaks of the Blueprint diseases. The most frequent prediction method identified was spatiotemporal modelling and risk mapping. Furthermore, most models predicted spatial or temporal aspects of future outbreaks. Most frequently, outbreak risk in a specific time or location was

predicted, qualitatively or quantitatively. While multiple models predicted future outbreak locations (eg, ‘a RVF outbreak is predicted in central Sudan’) and expected case numbers (eg, ‘12.3million Zika cases could be expected’) with relatively high granularity, few articles besides RVF studies made temporally precise predictions (eg, ‘we predict an outbreak from June to July of 2021’).²⁶ Specifically in the case of RVF models, habitat flooding and vector niche colayers enabled precise spatiotemporal predictions of outbreaks with time lags months in advance.³⁸

RVF and Zika were the most studied diseases, presumably in part due to their strong reliance on a set of predictable and measurable climatic and vector/host factors.³⁹ Zika’s recent multicontinental impact may have led to its elevated level of research.⁴⁰ Comparatively fewer articles concerning Ebola were published on outbreak prediction. However, during screening a significant number of articles were excluded, which modelled current Ebola epidemics.

Spatiotemporal modelling and climatic predictions

Risk mapping was the most widely used method of predicting and forecasting future outbreaks. Many RVF studies used the well-studied relationship between El Niño/Southern Oscillation (ENSO) phenomena, rainfall and cyclical patterns of outbreaks.³⁹ ENSO phenomena refers to the coupling of increased sea surface temperature (SST) and specific wind and rain patterns in the central and eastern tropical Pacific.⁴¹ In East Africa, ENSO results in above average rainfall, in turn flooding dambos of the RVF host *Aedes* mosquito needed for RVF outbreaks. SST and satellite-measured vegetation index (normalised difference vegetation index) are then used as surrogate variables to monitor RVF outbreak conditions.⁴² Many articles were able to make predictions based on this relationship together with historical outbreak and entomological niche data. This enabled the production of within-country, region-specific risk maps predicting outbreaks in East Africa. These maps guided the implementation of public health measures 2–4 months in advance of outbreaks.³⁸ Furthermore, climate-based RVF risk maps were either incorporated into or proposed as early warning systems.^{38 43–45} For example, the US armed forces established a multidisciplinary early warning system that reduced the economic and health consequences of the 2006–2007 Eastern African RVF outbreak, compared with the 1997 RVF outbreak.⁴⁶ We produced a case study illustrating the general RVF prediction methodology (figure 2).

Complex risk mapping for the Blueprint diseases

Risk maps were also used to predict diseases with more complex and/or poorly understood outbreak risk factors such as Ebola and Marburg disease, Lassa and CCHF.^{2447–49} A multistage outbreak assessment of Ebola integrated broad data types such as epidemiological, vector, expert opinion, demographic, transportation and geopolitical

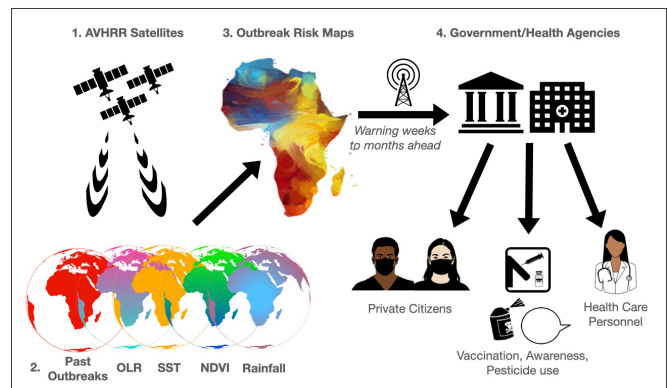


Figure 2 Example case study of Rift Valley fever (RVF) outbreak prediction. Illustration adapted from prediction strategies devised by Anyamba *et al*⁴³: (1) advanced very high resolution radiometers (AVHRR) on satellites measure observations of various global to subregional variables; (2) outgoing longwave radiation (OLR), sea surface temperature (SST), normalised difference vegetation index (NDVI) and rainfall together with coordinates of previous outbreaks are integrated into outbreak risk maps; (3) risk map predictions are associated to persistent anomalies in NDVI over specific locations, for example, predicting RVF outbreaks during future time periods and enabling warnings with time lags weeks to months ahead. Warnings are transmitted as part of an early warning system to different agencies (4), which lead pre-emptive measures: information to private citizens and health personnel, vaccination drives, awareness campaigns and vector control through pesticides.

data.⁵⁰ Pandemic potential from the community to international level could thus be assessed. Outbreak risk was further divided into index case potential, outbreak potential and epidemic potential. This multilayered analysis produced an outbreak receptivity risk map for the entire African continent up to the subregional level, permitting a thorough understanding of the potential of various factors on future outbreak likelihood.⁵⁰ This type of analysis may be useful in providing actionable information on very precise environments for policymakers. Predictions were also made for less well-studied diseases such as Nipah. Previous models had been limited by a paucity of spatial information.⁵¹ While an article used spatial occurrence data from Bangladesh to overcome this issue, limited application of human culture and ecological variables reduced aetiological understanding of this zoonosis.⁵¹ Other risk maps integrated socioeconomic, infrastructural and economic data, identifying societal risk factors and vulnerabilities that could theoretically be addressed pre-emptively by public health actors. Taylor *et al*.⁵² identified social vulnerabilities (eg, income, disease knowledge and phone access) to RVF in East African communities, enabling the production of vulnerability maps. However, in such studies, high model uncertainty remained regarding the integration of social vulnerability parameters in predictions. Lastly, models often mixed stochastic and deterministic models and frequently used machine learning. Models such as random forest algorithms or back propagation neural

networks were employed due to their ability to engage highly complex data sets.^{23 53–56} For example, a gradient tree boosting model integrated transportation, economic, demographic, ecological, case and vector data to create a Zika risk map estimating precise outbreak probability in the Asia Pacific region between 2016 and 2017.⁵⁷

Regression analysis and time series forecasts for CCHF and Lassa fever

The most used statistical outbreak prediction methods were regression analysis and time series forecasts, often predicting future case count. A time-trend model enabled the prediction of Lassa fever cases, a disease with a paucity of information, 5 years in advance.⁴⁸ Solar radiation analyses permitted RVF outbreak predictions 5 years in advance.⁵⁸ Time series predictions employed internet analysis (Google trend) in spatially defined regions as surrogate case data,⁵⁹ or directly obtained case data from governmental health institutions.⁶⁰ In the case of CCHF, case count may be the most readily accessible and reliable data as model input and may explain the application of time series analysis. However, a more layered approach for predicting spatially defined incidence and outbreaks is coupling occurrence time series to risk factors. The first prospective CCHF case prediction tool employed machine learning, 50 spatiotemporal covariates and 14 years of occurrence data, thus facilitating prediction of resources needed (vaccines, ventilators and Intensive Care Unit beds) and enabling preparedness (pesticides and reducing livestock import).⁶¹

Quantitative and probabilistic SARS and Nipah virus predictions

Certain quantitative models enabled multilayered probabilistic approaches to prediction. A mechanistic risk assessment framework gauged Nipah risk and predicted it to be hundreds of years before the introduction of Nipah into the European Union.⁶² This prediction was based on socioeconomic and ecological zoonotic drivers including human travel, trade, live animal movements and illegal bushmeat importation. While the model offered lower spatiotemporal granularity than weather-based risk mapping, such articles enable an analysis of contemporary causes frequently cited as driving zoonotic surges (eg, bushmeat trade).⁶³ Throughout the review, sociopolitical factors such as political instability and social vulnerability were sparsely integrated into models. A single model integrated political instability, conflict and outbreak receptivity.⁵⁰ Transportation was another important data source in prediction. A probabilistic SEIR SARS model evaluated the qualitative and quantitative global risk for spread and predicted infection cases through global transportation networks.³⁰ Case and transportation data enabled authors to predict infected countries on an almost one-to-one basis compared with future case data, and vaccination threshold on epidemic spreading.³⁰

Internet-based predictions of Zika and MERS

Internet-based systems represented a small yet promising domain of outbreak prediction. Analysed studies exclusively researched Zika and MERS. Methods included early warning systems integrating cloud computing and phone application-based risk mapping.^{36 37 56} Other models used Google search term forecasting.^{59 64 65} Cloud computing systems used machine learning to integrate demographic, user and geolocalised internet data to establish a real-time predictive infection and outbreak mapping system. This enabled individualised, context-specific, real-time feedback to application users warning them of imminent outbreak risks such as nearby infected users.⁵⁶ Further forecasting systems coupled Google trend search queries to real-time epidemiological data to enable adequate predictions of outbreak and epidemic onset. Google trend predictions were also easily verifiable through outbreak data weeks later. However, Google search methodology was often limited by its dependence on only two variables, case count and Google trends.⁶⁴ Phone and computer systems were limited by their proof of concept status, as they used synthetic data sets.

Narrative and qualitative outbreak prediction

Lastly, qualitative prediction models used a wide breadth of methodologies. An article predicted RVF outbreak dynamics (onset, plateau and seasonality) by assessing regional outbreak susceptibility through livestock seropositivity and host immunity. Coupled with local climatic and agricultural data, the authors produced a narrative assessment and prediction of a future RVF outbreak in Saudi Arabia, should an introduction event occur.³³ Another qualitative model established a surveillance network monitoring animal mortality to detect animal Ebola outbreaks and thus predict and prevent human outbreaks.³⁴ Specifically, local hunters and epidemiologists identified Ebola positive gorilla and simian carcasses and referred their observations to local healthcare actors. On two occasions, this network was able to warn the authorities in the Republic of Congo and Gabon of an imminent risk for human outbreaks.³⁴ Other authors evaluated the RVF-related knowledge level of locals at risk of RVF contact, cattle farmers, to inform model risk maps and in return produce tailored awareness programmes for such persons.⁴⁵

Stochasticity and determinism in predictions

60% of models reviewed were stochastic in nature. Stochasticity allows for uncertainty in modelling while respecting the inherent randomness of inferred underlying disease processes. For example, in stochastic models, it is possible for outbreaks to die out even if $R > 1$, which is not the case in deterministic modelling.⁶⁶ However, stochastic processes may not always be ideal for future predictions, as the underlying disease processes may change in future environments. In contrast, deterministic models tend to remain accurate in future environments, as changes to host–pathogen dynamics or disease processes are

more easily adapted into the model. Regardless, deterministic models pose a significant challenge in terms of complexity, especially for diseases studied herein that lack a well-established body of literature. This, and the relative simplicity of stochastic models, may explain part of the reliance on stochastic models.

Convergence and divergence in data sources

In general, data tended to diverge rather than converge on common sources. Foremost, and inherent to the research question, we compared 58 prediction models across nine different diseases on a global scale over a timeperiod of 20 years. Different diseases and different analysed environments resulted in diverging data sources. Second, heterogeneous study methods yielded different requirements in terms of data needed. Third, models often integrated authors' own assumptions. Lastly, multiple data sources were literature based and these sources have varied and changed over the 20-year time period studied. Convergence in data sources, when seen, was most often for climate data through the use of NASA and National Oceanic and Atmospheric Administration satellites or climate databases such as WorldClim. The listed data sources by disease can be found in online supplemental material.

Avenues in SARS-CoV-2 outbreak prediction

When taking our results into consideration, this review delineates a variety of avenues worthy of exploration in the prediction of SARS-CoV-2 outbreaks. Compartmental SEIR and time series models may currently be the most readily available models to make predictions for new COVID-19 outbreaks. However, such models must be continuously adapted as fluid governmental measures (eg, lockdowns and changing social norms) upend the underlying assumptions on which many of these models are based. Increasingly, machine learning algorithms such as Bayesian inference can help in evaluating large data sets with fluctuating underlying assumptions and high uncertainties in data value, as is the case with the current influx of COVID-19 datasets.¹⁷ Certain literature shows promise in combining SEIR/SIR models with machine learning algorithms.⁶⁷ Climate-based predictions, while rewarding for RVF, may be less applicable in predicting SARS-CoV-2. The virus is endemic, and the principal vectors are humans. Models would thus be applied on evaluating human vectors, possibly through behavioural and transportation data. Behavioural studies, such as evaluating MERS risk during the Hajj pilgrimage, and transportation analyses for SARS, may thus be of interest in the context of the current pandemic.^{29 30} Furthermore, as current predictions are for the most part on the regional to national scale and context specific, agent-based models could be useful in integrating these transportation and behavioural dynamics in the context of a confined local, regional or even national scale. No articles in this review applied agent-based models in predictions (online supplemental material 1). While various model

types are emerging in the prediction of COVID-19, there is no consensus on optimal models for COVID-19 prediction, and further research is needed towards predicting COVID-19 outbreaks.

Prediction validation

While 41% of articles validated their predictions against real data, there was significant heterogeneity in the demonstration of outbreak prediction validation. Certain articles graphically illustrated their predictions compared with real data, others purely stated their data had matched the predicted outbreak. Furthermore, many articles focused on whether their predicted outbreaks occurred, rather than outbreaks their models had missed. However, certain articles made easily verifiable outbreak predictions. For example, multiple articles predicted the RVF outbreak in 2006–2007.^{38 43} Of note, a single article predicted the Ebola outbreak in 2014 by predicting a peak of infectious bats in the region where the 2014 West African Ebola outbreak occurred.⁴⁹ Finally, 59% of articles did not validate their predictions against future data, often publishing before the timeline of predicted outbreaks. While certainly an avenue for further research, evaluating the accuracy of outbreak predictions exceeded the scope of this review.

Limitations in predictions

Limitations in the articles, when discussed, referenced the inherent complexity of epidemiological processes and the necessary simplification of their models. Another challenge was incomplete, unreliable or scarce data and patchy surveillance networks. This may have caused pseudoabsence scenarios in modelling where locations without recorded outbreaks may still have had instances of transmission occur. Certain articles only used synthetic data to make predictions, hindering validation.

Challenges in reviewing

This review also had methodological challenges and limitations. Separating articles between modelling of current epidemics on the one hand and prediction of future outbreaks on the other hand was a demanding screening challenge requiring thorough full-text assessments. Furthermore, a semantic challenge arose when certain studies equated transmission risk to outbreak risk, while others made a distinction of transmission risk being only a part of future outbreak risk. We thus excluded articles that merely evaluated for transmission risk or spillover risk defined as transmission of disease from one individual to another. Lastly, only the term Disease X was employed for an as-of-yet undiscovered pathogen in order to evaluate the Blueprint list's impact on research. This may have reduced the number of studies identified predicting outbreaks of undiscovered pathogens.

CONCLUSIONS

While there is ample research on modelling existing epidemics, the current review shows a significant literature

gap in prediction and forecasting of future outbreaks of the Blueprint priority diseases. Only few articles attempt true spatiotemporal prediction. The most common scenario, RVF risk mapping through vector, occurrence and climate data, appears to be precise geographically and temporally and has a track record of enabling prospective interventions that mitigated outbreaks. Even so, the warnings were presented when conditions were already rife for RVF outbreaks.⁶⁸ Thus, there may be room for improvement of early warning systems by including local health actors (eg, farmers, forest rangers and hunters) as frontline epidemiological personnel to warn of future outbreaks and implement public health strategies.^{34 68} Furthermore, only few articles sought to predict outbreaks through common zoonotic drivers such as ecological destruction, wildlife trade, conflict or political data and measures of social vulnerability. A significant research gap also concerned the integration of indicators of health system capacity, government effectiveness or health emergency preparedness into models. International Health Regulations factors such as State Party Self-assessment Annual Reports, Joint External Evaluations and Global Health Security Index measures were not integrated into the models reviewed. However, these indicators are an imperfect assessment of health capacities, and there is ongoing debate as how to best measure pandemic preparedness, which may explain the paucity of integration of such data in the analysed models.⁶⁹

This article attempted to displace the focus from outbreak risk factors to prediction, as prediction integrates risk factors into actionable information. While fraught with inherent uncertainty and stochasticity, prediction may be a pragmatic public health and research avenue that warrants support. Prediction of spatiotemporal and epidemiological qualities of future outbreaks enables public health actors to pre-emptively act on explicit spatiotemporal information. As such, a major challenge remains in foreseeing novel zoonoses. No articles studied Disease X outbreak prediction. Ebola, SARS, Zika and COVID-19 were all Disease X, before their initial outbreaks. While the Blueprint list includes SARS and MERS, no articles in this review predicted an outbreak of COVID-19 in 2019. This demonstrates that efforts to predict the future Disease X remain a major gap in the literature.

Twitter Antoine Flahault @FLAHAULT

Acknowledgements The main author would like to thank Alexia Trombert and Cécile Jaques at the Lausanne University Hospital Library for their help with the database searches, Sarah Nicollier, Alexander Jucht and Josh Reisler with proofreading and Amaury Thiabaud for his help in reviewing and selecting articles.

Contributors NJ contributed to the study design, analysed the data and wrote the manuscript. AF initiated the project, contributed to the study design, discussed the results and revised the manuscript. VD'A discussed the results, advised on study content and data presentation and revised the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- 1 WHO. *Annual review of diseases prioritized under the*. Geneva, Switzerland: Research and Development Blueprint, 2018.
- 2 Nasirian H. New aspects about crimean-congo hemorrhagic fever (CCHF) cases and associated fatality trends: a global systematic review and meta-analysis. *Comp Immunol Microbiol Infect Dis* 2020;69:101429.
- 3 Garry RF. Current topics in microbiology and immunology. *Curr Top Microbiol* 2020;1–22 <https://www.springer.com/series/82>
- 4 Liu S, Chan T-C, Chu Y-T, et al. Comparative epidemiology of human infections with middle East respiratory syndrome and severe acute respiratory syndrome coronaviruses among healthcare personnel. *PLoS One* 2016;11:e0149988.
- 5 Dawes BE, Freiberg AN. Henipavirus infection of the central nervous system. *Pathog Dis* 2019;77. doi:10.1093/femspd/ftz023. [Epub ahead of print: 01 03 2019].
- 6 Diop D, Rambe DS. Zika virus disease epidemics. *J Trop Dis* 2016;4.
- 7 Feldmann H, Sprecher A, Geisbert TW. Ebola. *N Engl J Med* 2020;382:1832–42.
- 8 Nanyingi MO, Munyua P, Kiama SG, et al. A systematic review of Rift Valley fever epidemiology 1931–2014. *Infect Ecol Epidemiol* 2015;5:28024.
- 9 Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature* 2008;451:990–3.
- 10 Morens DM, Fauci AS. Emerging pandemic diseases: how we got to COVID-19. *Cell* 2020;182:1077–92.
- 11 Mehand MS, Al-Shorbaji F, Millett P, et al. The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. *Antiviral Res* 2018;159:63–7.
- 12 Polonsky JA, Baidjoe A, Kamvar ZN, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos Trans R Soc Lond B Biol Sci* 2019;374:20180276.
- 13 Scarpino SV, Petri G. On the predictability of infectious disease outbreaks. *Nat Commun* 2019;10:898.
- 14 Morse SS, Mazet JAK, Woolhouse M, et al. Prediction and prevention of the next pandemic zoonosis. *Lancet* 2012;380:1956–65.
- 15 Jiang S, Shi Z-L. The first disease X is caused by a highly transmissible acute respiratory syndrome coronavirus. *Viral Sin* 2020;35:263–5.
- 16 Petrosillo N, Viceconte G, Ergonul O, et al. COVID-19, SARS and MERS: are they closely related? *Clinical Microbiology and Infection* 2020;26:729–34.
- 17 Roda WC, Varughese MB, Han D, et al. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model* 2020;5:271–81.
- 18 Huang QS, Lopez LD, McCallum L. Influenza surveillance and immunisation in New Zealand, 1997–2006. *Influenza Other Resp* 2008;2:139–45.
- 19 Hay AJ, McCauley JW. The WHO global influenza surveillance and response system (GISRS)-a future perspective. *Influenza Other Respir Viruses* 2018;12:551–7.
- 20 Pham MT, Rajić A, Greig JD, et al. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Res Synth Methods* 2014;5:371–85.

- 21 Tricco AC, Lillie E, Zarin W, *et al.* PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467.
- 22 Medical subject headings - home page. Available: <https://www.nlm.nih.gov/mesh/meshhome.html> [Accessed 21 Jan 2021].
- 23 Jiang D, Hao M, Ding F, *et al.* Mapping the transmission risk of Zika virus using machine learning models. *Acta Trop* 2018;185:391–9.
- 24 Samy AM, Thomas SM, Wahed AAE, *et al.* Mapping the global geographic potential of Zika virus spread. *Mem Inst Oswaldo Cruz* 2016;111:559–60.
- 25 Williams R, Malherbe J, Weepener H, *et al.* Anomalous high rainfall and soil saturation as combined risk indicator of Rift Valley fever outbreaks, South Africa, 2008–2011. *Emerg Infect Dis* 2016;22:2054–62.
- 26 Colón-González FJ, Peres CA, Steiner São Bernardo C, Bernardo CSS, *et al.* After the epidemic: Zika virus projections for Latin America and the Caribbean. *PLoS Negl Trop Dis* 2017;11:e0006007.
- 27 Chien L-C, Lin R-T, Liao Y, *et al.* Surveillance on the endemic of Zika virus infection by Meteorological factors in Colombia: a population-based spatial and temporal study. *BMC Infect Dis* 2018;18:180.
- 28 Solimini AG, Manica M, Rosà R, *et al.* Estimating the risk of dengue, Chikungunya and Zika outbreaks in a large European City. *Sci Rep* 2018;8:16435.
- 29 Colizza V, Barrat A, Barthélemy M, *et al.* Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC Med* 2007;5:34.
- 30 Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci U S A* 2004;101:15124–9.
- 31 Shacham E, Nelson EJ, Hoft DF, *et al.* Potential high-risk areas for Zika virus transmission in the contiguous United States. *Am J Public Health* 2017;107:724–31.
- 32 Gale P, Brouwer A, Ramnial V, *et al.* Assessing the impact of climate change on vector-borne viruses in the EU through the elicitation of expert opinion. *Epidemiol Infect* 2010;138:214–25.
- 33 Al-Qabati AG, Al-Afaleq AI. Cross-sectional, longitudinal and prospective epidemiological studies of Rift Valley fever in Al-Hasa Oasis, Saudi Arabia. *J Anim Vet Adv* 2010;9:258–65 <https://www.cabdirect.org/cabdirect/abstract/20103053163>
- 34 Rouquet P, Froment J-M, Bermejo M, *et al.* Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001–2003. *Emerg Infect Dis* 2005;11:283–90.
- 35 Mylne AQN, Pigott DM, Longbottom J, *et al.* Mapping the zoonotic niche of Lassa fever in Africa. *Trans R Soc Trop Med Hyg* 2015;109:483–92.
- 36 Sareen S, Gupta SK, Sood SK. An intelligent and secure system for predicting and preventing Zika virus outbreak using fog computing. *Enterp Inf Syst* 2017;4585:1–21.
- 37 Sandhu R, Sood SK, Kaur G. An intelligent system for predicting and preventing MERS-CoV infection outbreak. *J Supercomput* 2016;72:3033–56.
- 38 Anyamba A, Sang R, Nguku PM, Linthicum KJ, Small J, *et al.* Prediction, assessment of the Rift Valley fever activity in East and southern Africa 2006–2008 and possible vector control strategies. *Am J Trop Med Hyg* 2010;83:43–51.
- 39 Anyamba A, Linthicum KJ, Tucker CJ. Climate-disease connections: Rift Valley fever in Kenya. *Cad Saude Publica* 2001;17 Suppl:S133–40.
- 40 Oliveira JFde, Pescarini JM, Rodrigues MdeS, *et al.* The global scientific research response to the public health emergency of Zika virus infection. *PLoS One* 2020;15:e0229790.
- 41 What is the El Niño–Southern oscillation (ENSO) in a nutshell? Available: <https://www.climate.gov/news-features/blogs/enso/what-el-ni%C3%B1o%E2%80%93southern-oscillation-enso-nutshell> [Accessed 19 Jan 2021].
- 42 Linthicum KJ, Bailey CL, Tucker CJ, *et al.* Application of polar-orbiting, Meteorological satellite data to detect flooding of Rift Valley fever virus vector mosquito habitats in Kenya. *Med Vet Entomol* 1990;4:433–8.
- 43 Anyamba A, Chretien J-P, Small J, *et al.* Prediction of a Rift Valley fever outbreak. *Proc Natl Acad Sci U S A* 2009;106:955–9.
- 44 Leedale J, Jones AE, Caminade C. A dynamic, climate-driven model of Rift Valley fever. *Geospatial Health* 2015;11:394.
- 45 Gikungu D, Wakhungu J, Siamba D. Dynamic risk model for Rift Valley fever outbreaks in Kenya based on climate and disease outbreak data. *Geospatial Health* 2015;11:377.
- 46 Witt CJ, Richards AL, Masuoka PM, *et al.* The afhsc-division of GEIS operations predictive surveillance program: a multidisciplinary approach for the early detection and response to disease outbreaks. *BMC Public Health* 2011;11 Suppl 2:S10.
- 47 Nyakarahuka L, Ayebare S, Mosomtai G, *et al.* Ecological niche modeling for filoviruses: a risk map for Ebola and Marburg virus disease outbreaks in Uganda. *PLoS Curr* 2017;9. doi:10.1371/currents.outbreaks.07992a87522e1f229c7cb023270a2af1. [Epub ahead of print: 05 Sep 2017].
- 48 Olugasa BO, Odigie EA, Lawani M, *et al.* Development of a time-trend model for analyzing and predicting case-pattern of Lassa fever epidemics in Liberia, 2013–2017. *Ann Afr Med* 2015;14:89–96.
- 49 Fiorillo G, Bocchini P, Buceta J. A predictive spatial distribution framework for filovirus-infected bats. *Sci Rep* 2018;8:7970.
- 50 Pigott DM, Deshpande A, Letourneau I, *et al.* Local, National, and regional viral haemorrhagic fever pandemic potential in Africa: a multistage analysis. *Lancet* 2017;390:2662–72.
- 51 Peterson AT. Mapping risk of Nipah virus transmission across Asia and across Bangladesh. *Asia Pac J Public Health* 2015;27:NP824–32.
- 52 Taylor D, Hagenlocher M, Jones AE. Environmental change and Rift Valley fever in eastern Africa: projecting beyond healthy futures. *Geospatial Health* 2015;11:387.
- 53 Ak Çiğdem, Ergönül Önder, Şencan İrfan, *et al.* Spatiotemporal prediction of infectious diseases using structured gaussian processes with application to crimean-congo hemorrhagic fever. *PLoS Negl Trop Dis* 2018;12:e0006737.
- 54 Pigott DM, Golding N, Mylne A, *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* 2014;3:e04395.
- 55 Walsh MG, Willem de Smalen A, Mor SM. Wetlands, wild bovidae species richness and sheep density delineate risk of Rift Valley fever outbreaks in the African continent and Arabian Peninsula. *PLoS Negl Trop Dis* 2017;11:e0005756.
- 56 Kostkova P, Grasso F, Castillo C. Zika: a new system to empower health workers and local communities to improve surveillance protocols by e-learning and to forecast Zika virus in real time in Brazil. *Int Conf Digital Heal* 2018:90–4.
- 57 Teng Y, Bi D, Xie G, *et al.* Model-informed risk assessment for Zika virus outbreaks in the Asia-Pacific regions. *J Infect* 2017;74:484–91.
- 58 Stager JC, Ruzmaikin A, Conway D. Sunspots, El Niño, and the levels of Lake Victoria, East Africa. *J Geophys Res Atmospheres* 2007;13.
- 59 Morsy S, Dang TN, Kamel MG, *et al.* Prediction of Zika-confirmed cases in Brazil and Colombia using Google trends. *Epidemiol Infect* 2018;146:1625–7.
- 60 Ansari H, Shahbaz B, Izadi S, *et al.* Crimean-Congo hemorrhagic fever and its relationship with climate factors in Southeast Iran: a 13-year experience. *J Infect Dev Ctries* 2014;8:749–57.
- 61 Ak Ç, Ergönül Ö, Gönen M. A prospective prediction tool for understanding crimean-congo haemorrhagic fever dynamics in turkey. *Clin Microbiol Infect* 2020;26:123.e1–123.e7.
- 62 Simons RRL, Horigan V, Gale P, *et al.* A generic quantitative risk assessment framework for the entry of bat-borne zoonotic viruses into the European Union. *PLoS One* 2016;11:e0165383.
- 63 Kilpatrick AM, Randolph SE, Drivers RSE. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* 2012;380:1946–55.
- 64 Teng Y, Bi D, Xie G, *et al.* Dynamic forecasting of Zika epidemics using google trends. *PLoS One* 2017;12:e0165085.
- 65 Castro JS, Torres J, Oletta J, *et al.* Google trend tool as a predictor of Chikungunya and Zika epidemic in an environment with little epidemiological data, a Venezuelan case. *International Journal of Infectious Diseases* 2016;53:133–4.
- 66 Lloyd AL, Zhang J, Root AM. Stochasticity and heterogeneity in host-vector models. *J R Soc Interface* 2007;4:851–63.
- 67 Ardabili SF, Mosavi A, Ghamisi P, *et al.* COVID-19 outbreak prediction with machine learning. *SSRN Journal* 2020.
- 68 Jost CC, Nzietchueng S, Kihu S, *et al.* Epidemiological assessment of the Rift Valley fever outbreak in Kenya and Tanzania in 2006 and 2007. *Am J Trop Med Hyg* 2010;83:65–72.
- 69 Kentikelenis A, Seabrooke L. Organising knowledge to prevent global health crises: a comparative analysis of pandemic preparedness indicators. *BMJ Glob Health* 2021;6:e006864.
- 70 CDC. Cost of the Ebola epidemic, 2019. Available: <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/cost-of-ebola.html> [Accessed 20 Sep 2020].
- 71 Elston JWT, Cartwright C, Ndumbi P, *et al.* The health impact of the 2014–15 Ebola outbreak. *Public Health* 2017;143:60–70.
- 72 Faria NR, Azevedo RdoSdaS, Kraemer MUG, *et al.* Zika virus in the Americas: early epidemiological and genetic findings. *Science* 2016;352:345–9.

- 73 Pierson TC, Diamond MS. The emergence of Zika virus and its new clinical syndromes. *Nature* 2018;560:573–81.
- 74 Wright D, Kortekaas J, Bowden TA, *et al.* Rift valley fever: biology and epidemiology. *J Gen Virol* 2019;100:1187–99.
- 75 Richmond JK, Baglole DJ. Lassa fever: epidemiology, clinical features, and social consequences. *BMJ* 2003;327:1271–5.
- 76 Escaffre O, Borisevich V, Rockx B. Pathogenesis of hendra and nipah virus infection in humans. *J Infect Dev Ctries* 2013;7:308–11.
- 77 Messina JP, Pigott DM, Golding N, *et al.* The global distribution of crimean-congo hemorrhagic fever. *Trans R Soc Trop Med Hyg* 2015;109:503–13.
- 78 Hwang SW, Cheung AM, Moineddin R, *et al.* Population mortality during the outbreak of severe acute respiratory syndrome in Toronto. *BMC Public Health* 2007;7:93.
- 79 Carias C, O'Hagan JJ, Jewett A, *et al.* Exportations of symptomatic cases of MERS-CoV infection to countries outside the middle East. *Emerg Infect Dis* 2016;22:723–5.
- 80 Spatiotemporal analysis Columbia public health. Available: <https://www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis> [Accessed 28 Jan 2021].
- 81 Tran A, Trevenne C, Lutwama J, *et al.* Development and assessment of a geographic knowledge-based model for mapping suitable areas for Rift Valley fever transmission in eastern Africa. *PLoS Negl Trop Dis* 2016;10:e0004999.
- 82 Fichet-Calvet E, Rogers DJ. Risk maps of lassa fever in West Africa. *PLoS Negl Trop Dis* 2009;3:e388.
- 83 Redding DW, Moses LM, Cunningham AA, *et al.* Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever. *Methods Ecol Evol* 2016;7:646–55.
- 84 Mosomtai G, Evander M, Sandström P, *et al.* Association of ecological factors with Rift Valley fever occurrence and mapping of risk zones in Kenya. *Int J Infect Dis* 2016;46:49–55.
- 85 Anyamba A, Chretien J-P, Small J, *et al.* Developing global climate anomalies suggest potential disease risks for 2006–2007. *Int J Health Geogr* 2006;5:60.
- 86 Abdelgadir DM, Bashab HMM, Elhadi Mohamed RA, *et al.* Risk factor analysis for outbreak of Rift Valley fever in Khartoum state of Sudan. *J Entomol Sci* 2010;45:239–51.
- 87 Redding DW, Tiedt S, Lo Iacono G, *et al.* Spatial, seasonal and climatic predictive models of Rift Valley fever disease across Africa. *Philos Trans R Soc Lond B Biol Sci* 2017;372:20160165.
- 88 Anyamba A, Linthicum KJ, Small J, *et al.* Remote sensing contributions to prediction and risk assessment of natural disasters caused by large-scale Rift Valley fever outbreaks. *Proc IEEE Inst Electr Electron Eng* 2012;100:2824–34.
- 89 Muniyua PM, Murithi RM, Ithondeka P, *et al.* Predictive factors and risk mapping for Rift Valley fever epidemics in Kenya. *PLoS One* 2016;11:e0144570.
- 90 Siraj AS, Perkins TA. Assessing the population at risk of Zika virus in Asia – is the emergency really over? *BMJ Glob Health* 2017;2:e000309.
- 91 Estrada-Peña A, Vatansever Z, Gargili A, *et al.* An early warning system for crimean-Congo haemorrhagic fever seasonality in turkey based on remote sensing technology. *Geospat Health* 2007;2:127–35.
- 92 Linthicum KJ, Anyamba A, Tucker CJ, *et al.* Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* 1999;285:397–400.
- 93 Mostafavi E, Chinikar S, Bokaei S, *et al.* Temporal modeling of Crimean-Congo hemorrhagic fever in eastern Iran. *Int J Infect Dis* 2013;17:e524–8.
- 94 Drake JM, Hassan AN, Beier JC. A statistical model of Rift Valley fever activity in Egypt. *J Vector Ecol* 2013;38:251–9.
- 95 Bender R. Cancer epidemiology. *Methods Mol Biology* 2009;471:179–95.
- 96 Time series forecasting methods. Available: <https://www.influxdata.com/time-series-forecasting-methods/> [Accessed 28 Jan 2021].
- 97 6.4. Introduction to time series analysis. Available: <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm> [Accessed 28 Jan 2021].
- 98 Ansari H, Mansournia MA, Izadi S, *et al.* Predicting CCHF incidence and its related factors using time-series analysis in the Southeast of Iran: comparison of SARIMA and Markov switching models. *Epidemiol Infect* 2015;143:839–50.
- 99 Wollenberg Valero KC, Isokpehi RD, Douglas NE, *et al.* Plant phenology supports the Multi-emergence hypothesis for Ebola spillover events. *Ecohealth* 2018;15:497–508.
- 100 Da'ar OB, Ahmed AE. Underlying trend, seasonality, prediction, forecasting and the contribution of risk factors: an analysis of globally reported cases of middle East respiratory syndrome coronavirus. *Epidemiol Infect* 2018;146:1343–9.
- 101 Lessler J, Rodriguez-Barraquer I, Cummings DAT, *et al.* Estimating potential incidence of MERS-CoV associated with hajj pilgrims to Saudi Arabia, 2014. *PLoS Curr* 2014;6. doi:10.1371/currents.outbreaks.c5c9c9abd636164a9b6fd4dbda974369. [Epub ahead of print: 24 Nov 2014].