A systematic review of the data, methods and environmental covariates used to map *Aedes*-borne arbovirus transmission risk

Ah-Young Lim[1,2*], Yalda Jafari[3,4], Jamie M. Caldwell[5], Hannah E. Clapham[6], Katy A. M. Gaythorpe[7], Laith Hussain-Alkhateeb[8,9], Michael A. Johansson[10], Moritz U. G. Kraemer[11], Richard J. Maude[3,4], Clare P. McCormack[7], Jane P. Messina[12,13], Erin A. Mordecai[14], Ingrid B. Rabe[15], Robert C. Reiner Jr[16,17], Sadie J. Ryan[18], Henrik Salje[19], Jan C. Semenza[20], Diana P. Rojas[15] and Oliver J. Brady[1,2]

Abstract

Background *Aedes (Stegomyia)*-borne diseases are an expanding global threat, but gaps in surveillance make comprehensive and comparable risk assessments challenging. Geostatistical models combine data from multiple locations and use links with environmental and socioeconomic factors to make predictive risk maps. Here we systematically review past approaches to map risk for different *Aedes*-borne arboviruses from local to global scales, identifying differences and similarities in the data types, covariates, and modelling approaches used.

Methods We searched on-line databases for predictive risk mapping studies for dengue, Zika, chikungunya, and yellow fever with no geographical or date restrictions. We included studies that needed to parameterise or fit their model to real-world epidemiological data and make predictions to new spatial locations of some measure of population-level risk of viral transmission (e.g. incidence, occurrence, suitability, etc.).

Results We found a growing number of arbovirus risk mapping studies across all endemic regions and arboviral diseases, with a total of 176 papers published 2002–2022 with the largest increases shortly following major epidemics. Three dominant use cases emerged: (i) global maps to identify limits of transmission, estimate burden and assess impacts of future global change, (ii) regional models used to predict the spread of major epidemics between countries and (iii) national and sub-national models that use local datasets to better understand transmission dynamics to improve outbreak detection and response. Temperature and rainfall were the most popular choice of covariates (included in 50% and 40% of studies respectively) but variables such as human mobility are increasingly being included. Surprisingly, few studies (22%, 31/144) robustly tested combinations of covariates from different domains (e.g. climatic, sociodemographic, ecological, etc.) and only 49% of studies assessed predictive performance via out-of-sample validation procedures.

Conclusions Here we show that approaches to map risk for different arboviruses have diversified in response to changing use cases, epidemiology and data availability. We identify key differences in mapping approaches between

*Correspondence:
Ah-Young Lim
Ahyoung.Lim@lshtm.ac.uk

Background

Arboviruses, commonly referred to as arthropod-borne viruses, are a wide range of viral pathogens transmitted through the bite of arthropods such as mosquitoes and ticks. The term arbovirus does not refer to a distinct taxonomic group, but the viruses have similar transmission mechanisms, which makes information gained from one virus potentially useful in understanding and preventing the spread of other viruses [1]. In this paper, we focus on *Aedes* (*Stegomyia*)-borne arboviruses, including dengue, Zika, chikungunya, and yellow fever, which are of particular concern due to their high disease burden and life-threatening health consequences [2]. The geographical spread and burden of this group of arboviruses have been rapidly increasing in recent years. It has been estimated that 100–400 million dengue infections occur each year worldwide, mainly in South America and South-East Asia (SE Asia), with the disease threatening to spread to new regions including Europe [3–5]. Zika and chikungunya viruses were first identified in Africa and Asia, but emerged and rapidly spread throughout the Americas between 2013 and 2015, likely due to a combination of suitable climatic factors, increasing international air travel and possible immunological drivers [6, 7]. The Zika outbreak received global attention due to its link to congenital and neurological complications, resulting in the declaration of a Public Health Emergency of International Concern by the World Health Organization (WHO) in 2016 [7]. Chikungunya is frequently accompanied by joint pain and rheumatic manifestations that can persist for a long time and have a significant impact on the quality of life of affected individuals [4]. Yellow fever is endemic in tropical and subtropical countries of South America and Africa, with an estimated number of 109,000 severe infections and 51,000 deaths in 2018 [8]. A sylvatic cycle between non-human primate reservoirs and mosquitoes is the most common source of yellow fever virus infection; however, humans can also become infected through the urban cycle, which can potentially lead to large outbreaks, as recently seen in Angola, Nigeria and the Democratic Republic of the Congo [8, 9]. As these *Aedes*-borne arboviruses share a common mechanism of transmission, the WHO launched the Global Arbovirus Initiative in 2022, which includes the aim of developing a comprehensive risk monitoring and early detection tool that will allow countries to assess global risk of different *Aedes*-borne viruses, strengthen vector control, and develop global systems and strategies to monitor and reduce the risk in the local, regional, and national levels. This initiative identified reviewing the drivers of spatial arbovirus risk at global and regional levels as a key priority.

Surveillance of arboviral diseases varies among countries, by clinical manifestations, and over time, but three main data types are used most commonly for risk mapping: disease occurrence, case incidence, and seroprevalence data. Occurrence data represent a specific location where one or more cases of a disease has occurred [10] (e.g. an outbreak report) and is often available even in otherwise data-sparse regions, but conveys limited information about the magnitude of risk. Case incidence, as measured by traditional, largely passive disease surveillance systems, provides more information on magnitude due to being denominator-based (e.g. cases per 1,000 residents), but often underestimates the incidence of infection and is often not directly comparable between countries due to differing case definitions, health seeking patterns, health care and laboratory capacity, immunological landscape and surveillance systems. Age-specific community-representative seroprevalence survey data, when combined with models, can be used to estimate force of infection. This provides a less biased measure of long-term transmission risk, but is the least abundant data type and is subject to the limitations of serology in the context of cross-reactive flavivirus infections [11].

The geographic distribution and intensity of *Aedes*-borne arbovirus transmission have been attributed to a combination of pathogen, environmental, demographic and socioeconomic factors such as climate change, urbanisation and local and international travel. Temperature, in particular, is a frequently cited determinant of arbovirus transmission, as temperature drives all important metabolic traits for vector mosquitoes to transmit the virus to humans [12]. Rapid unplanned urbanisation increases human population density, can create urban heat islands and can lead to inadequate water provision and solid waste disposal which favour the proliferation of both vectors and virus transmission [13]. Increasing trade has facilitated expansion of *Aedes* vectors while increasing travel of humans has spread new viruses and virus sub-types into previously naive populations [14]. Finally, the level of local immunity also helps determine arboviral transmission patterns. Immunity is driven by both demography and past pathogen circulation patterns and can vary substantially between populations. The inherent spatial and temporal patterns of arbovirus transmission

are therefore the result of the complex interactions of multiple factors, likely differing between arbovirus, location and spatial scale.

A wide range of spatial modelling techniques has been developed to account for complexities in investigating the variations in geographic spread of *Aedes*-borne arbovirus infections. Broadly, these can be categorised into (i) data-driven approaches where flexible statistical models aim to recreate observed patterns with fewer built-in mechanistic assumptions about how variables are related to risk or (ii) process-driven approaches where assumptions about drivers and how they affect transmission are encoded in a mechanistic (mathematical) model, which is then fit to observational data. Due to data scarcity in many risk mapping applications, implementing statistical and mathematical models in Bayesian frameworks has become increasingly popular due to incorporating prior information and better representing uncertainty in their predictions.

Previous systematic reviews have been conducted to identify and characterise dengue transmission models focused on predicting trends over time (hindcasting with the goal of developing forecasting systems) as opposed to spatially explicit prediction (risk mapping) [15–17]. Some of these systematic reviews included risk mapping studies but they have been limited to just a single arbovirus, usually dengue [7, 18–20]. Although arbovirus risk mapping studies have become more diverse and advanced, to our knowledge, no review has systematically assessed these studies to identify best practices in the field, characterise similarities and differences when mapping different arboviral diseases, identify knowledge gaps and develop recommendations for future arbovirus mapping studies. Therefore, this systematic review aims to address these unmet needs through systematically evaluating the epidemiological data, covariates, modelling frameworks and methods of evaluation used by arbovirus mapping studies.

Methods

This review employed a search strategy and inclusion and exclusion criteria based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [21].

Search strategy

Four online bibliographic databases were searched: Embase, Global Health, Medline, and Web of Science. The final search was conducted on 15 June 2022 using institutional access from Oxford University. The search strategy included keywords and Medical Subject Headings (MesH) related to different arboviral diseases (namely dengue, Zika, chikungunya, and yellow fever) and those related to prediction. Search terms included "(Dengue OR DENV OR Zika OR ZIKV OR Chikungunya OR CHIKV OR Yellow fever OR YFV) AND (predict* OR forecast* OR map* OR driver*)". Additionally, we manually searched the reference lists of articles and contacted experts in the field of arbovirus modelling to identify any studies not identified through the database search.

Selection process

Results from database searches were combined and stored using Zotero referencing software; duplicates were removed using R (version 4.2.2) [22] by comparing the Digital Object Identifier (DOI) numbers of each study. Titles and abstracts were screened independently by two team members. All identified papers were included in full-text review and irrelevant articles were excluded. Full-text review was completed and disagreements on inclusion were resolved by consensus.

Inclusion/exclusion criteria

Articles must be peer-reviewed, published in English and contain a spatial model that investigates the transmission of the arboviruses to humans. Spatial models were defined as models that included geographically realistic and explicit representations of more than one spatial location. While our primary focus was to review spatial models, spatiotemporal models were also included. There were no geographical or publishing date restrictions applied. We only included models that made predictions of some measure of the population-level virus infection risk, including but not limited to occurrence, incidence, prevalence, and derivatives of these measures (e.g. reproduction number). Studies where the model was developed and/or validated in a previous paper were also included.

Articles were excluded if they did not fit to or validate predictions using autochthonous human case data. Studies that only predicted suitability for the mosquito vectors or non-human hosts were excluded. Articles that predicted mosquito-based virus transmission metrics, such as vectorial capacity, were excluded, unless they were combined with other variables in a model fit to or validated with autochthonous human case data. Studies were excluded if they had only descriptive mapping of incidence using geographic information systems. Simulation-based and theoretical modelling studies were excluded unless their predictions of *Aedes*-borne disease transmission risk (as opposed to model parameters) were validated using data from real-world settings. Conference and workshop proceedings were excluded, as were review articles. This systematic review is registered on PROSPERO (reference: CRD42022343032).

Data extraction

The following variables were extracted from eligible articles:

– study identification (title, author names, year of publication, study area, disease studied);
– model characteristics (type of model used, covariates included, covariates tested and not included, spatiotemporal resolution, assessment of collinearity);
– model validation (validation methods, metrics used to assess the model performance).

Analysis of the data and visualisations were carried out using R (version 4.2.2) [22]. The complete list of all included studies and data extracted from each study are available in Additional file 1.

Quality assessment

A quality assessment tool was developed using the EPI-FORGE checklist (Additional file 2), a guideline for standardised reporting of epidemic forecasting and prediction research, to assess the reporting quality of included studies [23]. This guideline assesses whether studies report on the following domains: study goals, data sources, model characteristics and assumptions, model evaluation, and study generalisability. The nine criteria were equally weighted, each with a score of 0 (poor) to 2 (good), for a maximum of 18 points. On the basis of the overall score, each paper was rated 'low' (<10), 'medium' (10–12), 'high' (13–15) or 'very high' (>15).

Results

A total of 16,625 records were retrieved from the databases and 7,742 titles and abstracts screened after removing duplicates (Fig. 1). A total of 83 records were additionally identified through bibliographic searches and contacts with experts. Of 301 records, a total of 125 studies were excluded. A total of 69 studies did not meet the inclusion criteria because they only included descriptive spatial analyses rather than models (n=48), did not make predictions to more than one spatial location (n=13) or made temporal predictions only (n=8). A total of 49 studies were excluded because they met our exclusion criteria as simulation-only studies (n=14), only predicting the vector distribution (n=10) or not fitting or validating models to autochthonous human case data (n=25). Full text manuscripts were unavailable for 2 studies, a further 2 were published in languages outside the inclusion criteria and 3 studies were deleted
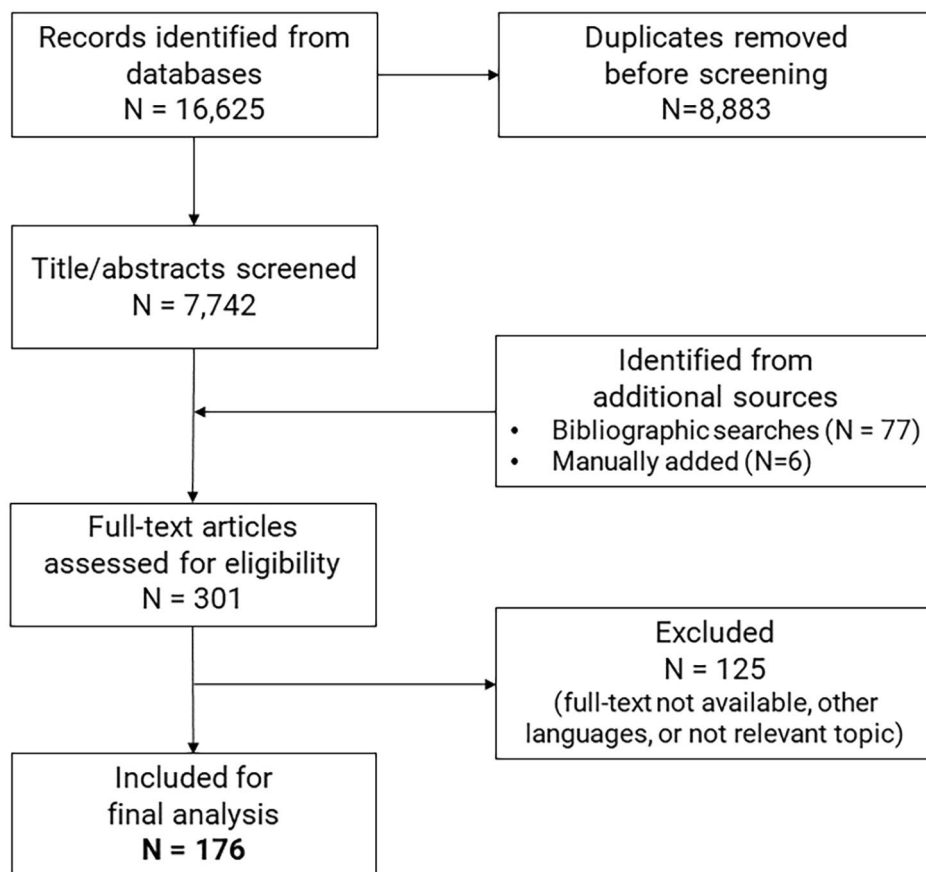

Fig. 1 PRISMA flow chart

because they addressed topics outside the scope of this review. One paper included two different models using different datasets so we counted it as two separate studies [24]. As a result, we identified 176 studies published between 2002 and 2022 that were ultimately included in the review (Fig. 1).

There has been a rapid increase in the number of arboviral spatial modelling studies over the past 20 years, reflecting the growing public health priority of these diseases and increasing accessibility of data and modelling methods. There was an average of 1.7 studies published per year before 2008, 4.6 studies per year between 2008 and 2014 and 18.4 per year between 2015 and 2021 (Fig. 2).

The distribution of risk mapping studies over geography and by disease closely follow the abundance and availability of data. Using WHO Regions, a total of 41.5% (n=73) of the studies were conducted in the Americas, followed by 21.0% (n=37) in SE Asia and 17.3% (n=30) in the Western Pacific region with a wide geographic diversity of studies over the past five years. Brazil (n=35) was the most frequently studied country, followed by Colombia (n=15) and Indonesia (n=13). The diversity of regions studied has also increased: until 2014 studies tended to focus primarily on the Americas and Western Pacific whereas since 2015 studies focusing on SE Asia and the global scale have been increasingly prevalent (Fig. 2). More than 70% (n=125) of the studies modelled dengue transmission, 19 (10.8%) modelled Zika, 15 (8.5%) modelled yellow fever and seven (4.0%) chikungunya. There were six (3.4%) studies that modelled the risk of dengue, Zika, and chikungunya together, while also modelling the diseases individually; two modelled dengue and Zika together and two modelled Zika and chikungunya together.

Purpose of maps

The main groups of purposes or goals of risk maps vary depending on the specific disease and context, but can generally be grouped into four categories: (1) providing a broad overview of the spatial distribution of risk over long-term averages and suggesting how it might change under different scenarios of global changes in climate, economics, and demographics (e.g., [28, 29]); (2) predicting the spread of outbreaks and gaining a better understanding of major drivers of geographical spread (e.g., [30, 31]); (3) evaluating and planning vaccination programs by estimating disease burden and identifying high-risk areas at the continental or country-level scale (e.g., [32, 33]); and (4) informing planning and outbreak response by increasing the precision of risk estimates and
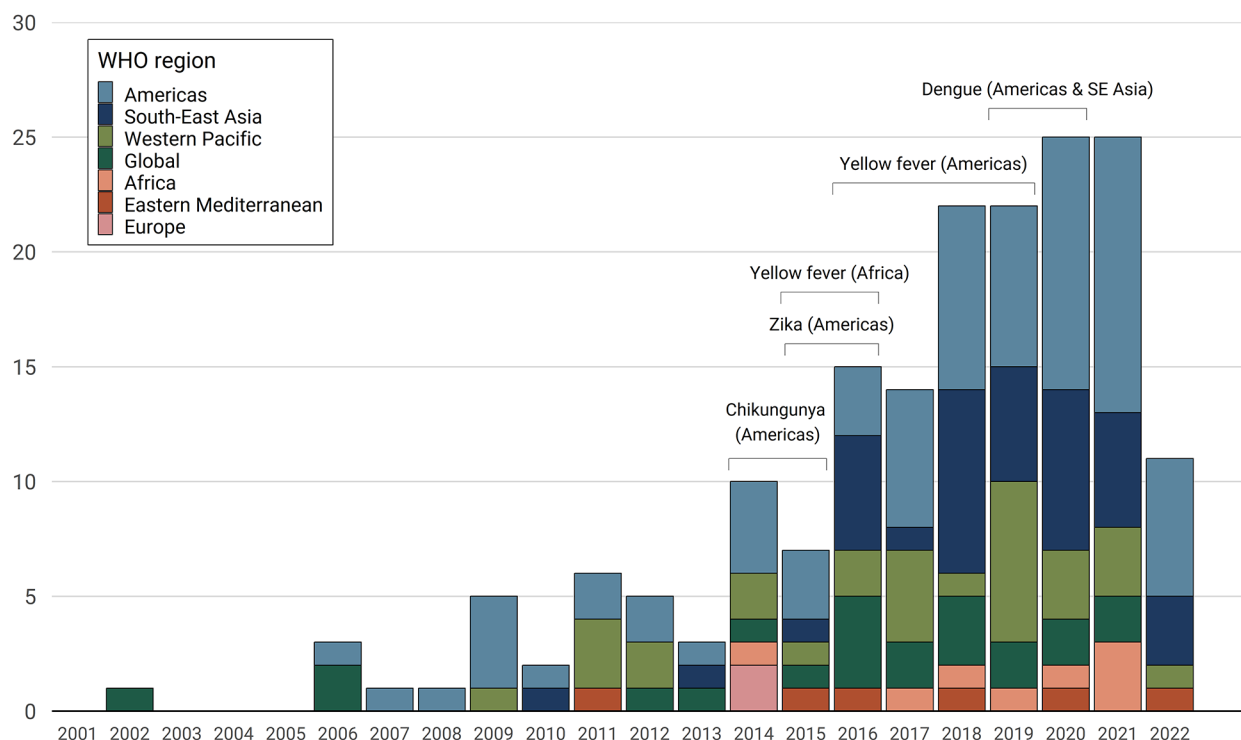

Fig. 2 Number of included studies per year by study region. The brackets represent the key years for *Aedes*-borne arbovirus outbreaks, including chikungunya in the Americas (2014–2015) [25], Zika in the Americas (2015–2016) [7], yellow fever in Brazil (2016–2019) and Angola and Democratic Republic of Congo (2015–2016) [26], and dengue in the Americas & SE Asia (2019–2020) [27]

mapping sub-national risk using surveillance data (e.g., [34, 35]).

Data types

Most studies (n=130, 73.9%) used case count data from routine passive surveillance to fit models, most often aggregated to the administrative district (admin2)- or province (admin1)-level (Fig. 3). Use of occurrence data was also widespread (n=29, 16.5%), particularly for specific use cases, such as the generation of global suitability maps. There were only seven studies (4.0%) that included data from community-representative seroprevalence surveys, and seven studies that included data from at least two different data types. The use of seroprevalence data was limited to dengue (n=8) and yellow fever (n=4), both resulting from widespread seroprevalence surveys in preparation for, or to evaluate, vaccination programmes. Generally the paucity of any one data type for yellow fever meant a more equally distributed use of different data types in models and greater use of multiple types of data [8, 33, 36, 37] (Fig. 3).

Risk maps have been generated across a wide range of spatial scales from global to sub-national (Fig. 4). We identified 22 studies that produced global risk maps of various *Aedes*-borne arboviruses. Despite large gaps in data availability at the global scale, the majority (n=18/22, 81.8%) of these global maps make high resolution predictions at the pixel level, enabled by growing availability of high resolution remotely-sensed climate datasets (Fig. 5). For Zika, yellow fever, and chikungunya, maps were primarily focussed at a continent or national scale with a resolution between city-level and national-level (Figs. 4 and 5), reflecting the more regional scope of their distribution (yellow fever in Africa) or high profile epidemics (the 2015–2016 Zika epidemic in the Americas). While maps are available at all spatial scales for dengue, the majority of models (n=82, 65.6%) are now at sub-national scale, usually at the resolution of city/district (admin2) (Figs. 4 and 5). This reflects the increasing application of these techniques to routinely collected case incidence data to provide country-specific recommendations about targeting of control resources within countries based on the latest local data. There remain strong regional disparities in the scale and resolution of mapping efforts with many high-resolution and country-specific maps in the Americas, while risk estimates for Africa are fewer, of comparatively lower resolution, and are typically derived from global or continent-level modelling efforts (Figure S1, Additional file 3).
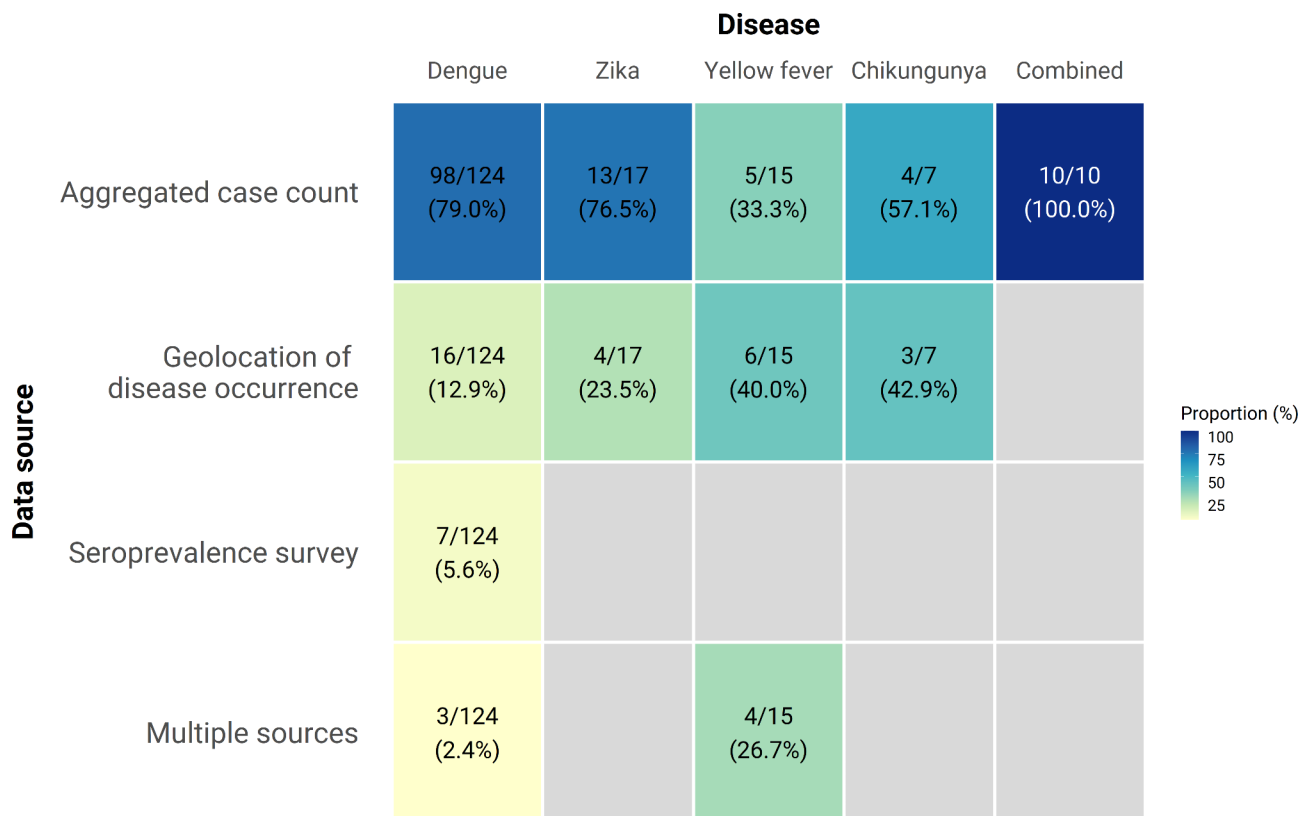

Fig. 3 Sources of epidemiological data used by diseases. Each cell represents the number and percentage of studies with the denominators summed vertically. Three studies did not specify the data source
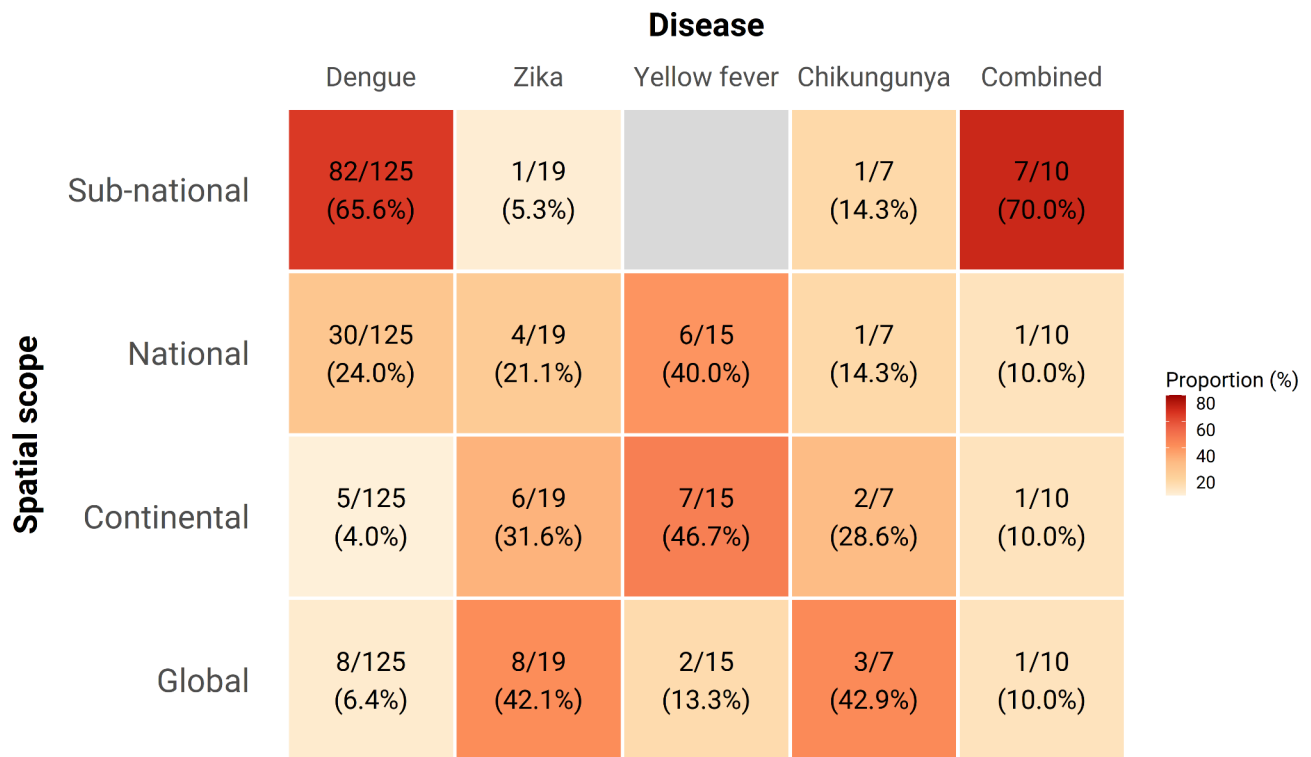
Disease

Spatial scope	Dengue	Zika	Yellow fever	Chikungunya	Combined
Sub-national	82/125 (65.6%)	1/19 (5.3%)		1/7 (14.3%)	7/10 (70.0%)
National	30/125 (24.0%)	4/19 (21.1%)	6/15 (40.0%)	1/7 (14.3%)	1/10 (10.0%)
Continental	5/125 (4.0%)	6/19 (31.6%)	7/15 (46.7%)	2/7 (28.6%)	1/10 (10.0%)
Global	8/125 (6.4%)	8/19 (42.1%)	2/15 (13.3%)	3/7 (42.9%)	1/10 (10.0%)

Proportion (%)
80
60
40
20

Fig. 4 Geographical scope by diseases. Each cell represents the number and percentage of studies with the denominators summed vertically

Spatial resolution

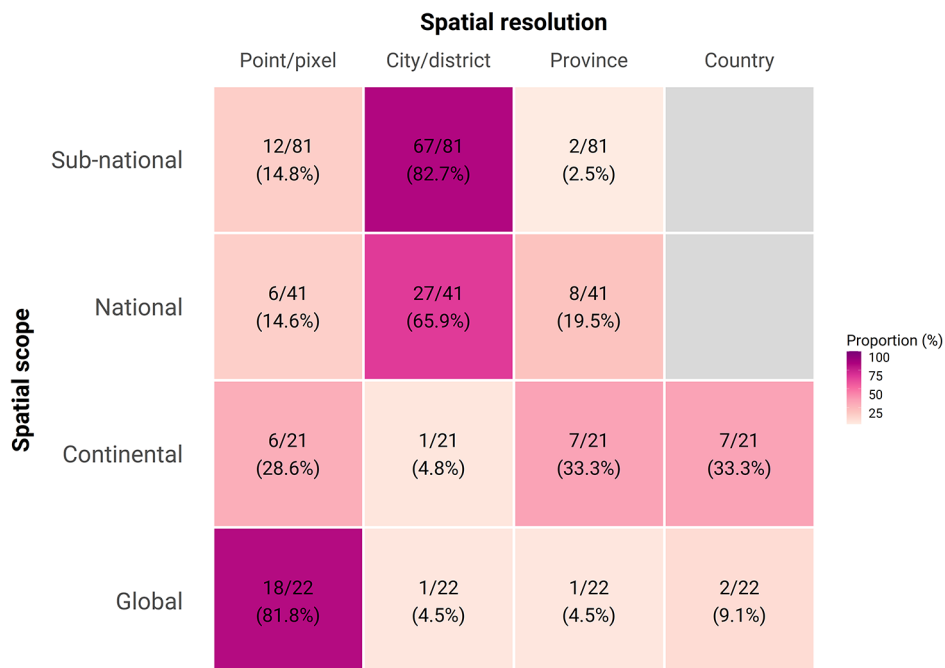| Spatial scope	Point/pixel	City/district	Province	Country
Sub-national	12/81 (14.8%)	67/81 (82.7%)	2/81 (2.5%)	
National	6/41 (14.6%)	27/41 (65.9%)	8/41 (19.5%)	
Continental	6/21 (28.6%)	1/21 (4.8%)	7/21 (33.3%)	7/21 (33.3%)
Global	18/22 (81.8%)	1/22 (4.5%)	1/22 (4.5%)	2/22 (9.1%)

Proportion (%)
100
75
50
25

Fig. 5 Spatial resolution by geographical scope. Each cell represents the number and percentage of studies with the denominators summed horizontally. Eleven studies did not specify the spatial resolution

Spatiotemporal prediction maps were often generated based on monthly or weekly intervals (Table S1, Additional file 3). The longest period of study was for 804 months (67 years), while the shortest period of study was

Table 1 List of covariates included in the studies. Some articles are listed more than once

Covariates	Number of studies	Percentage (%)
Climatic factors		
Temperature	97	55.1
Mean temperature	78	44.3
Min temperature	21	11.9
Max temperature	16	9.1
Rainfall	78	44.3
Total rainfall	42	23.9
Mean rainfall	23	13.1
Max rainfall	6	3.4
Humidity	25	14.2
Temperature suitability	22	12.5
Bioclimatic variables	6	3.4
El Niño Southern Oscillation Index	4	2.3
Soil moisture (water stress/wetness)	4	2.3
Demographic factors		
Population density	43	24.4
Age	25	14.2
Air travel	15	8.5
Human daily mobility	13	7.4
Vaccination coverage	7	4.0
Sex	7	4.0
Socio-economic factors		
Gross domestic product	15	8.5
Household income	11	6.3
Education/literacy rate	6	3.4
Occupation and employment status	5	2.8
Socio-economic strata	6	3.4
Ecology		
Non-human primates species	6	3.4
Location of breeding sites	4	2.3
Breteau index	3	1.7
Adult mosquito abundance	2	1.1
Environmental factors		
Vegetation	27	15.3
Elevation/altitude	25	14.2
Urbanisation	22	12.5
Distance to roads, road density	14	8.0
Land use/land cover	13	7.4
Distance to water bodies/river	9	5.1
Spatiotemporal incidence		
Case count across time periods and neighbouring regions	18	10.2

for 3 months, with an average of 128 months (10 years) and a median of 60 months (5 years). Studies tended to use data from periods with high numbers of reported cases, with dengue data concentrated in between 2010 and 2015, Zika data between 2015 and 2016. For chikungunya many studies use data from 2014 and for yellow fever the data used have been spread over time, with few studies using recent data from 2015 to 2020 (Figure S2, Additional file 3).

Covariates

Studies reviewed included a wide range of covariates in their models (Table 1). We grouped these into six main groups: climatic, demographic, socioeconomic, ecological, environmental and spatiotemporal incidence.

Climatic variables were the most common group of covariates in models with temperature and rainfall dominating. More than half of the studies (n=97, 55.1%) included temperature as a covariate while around 40% of studies had rainfall (n=78, 44.3%). For temperature, the mean temperature (n=78, 44.3%) was most frequently used, while for rainfall, the cumulative amount of rain fallen during the time period (n=42, 23.9%) was more commonly included. Station-based data was the most popular data source for both temperature (n=46) and rainfall (n=36), with satellite images (e.g., NASA MODIS [38, 39], the Climate Prediction Center Morphing Technique [40, 41] and Global Precipitation Climatology Project [42, 43]) and the reanalysis and model assimilation data (e.g., ERA5 [27, 36]) becoming popular. Temperature and rainfall were better fit when lagged one or two months rather than unlagged [44–47]. Temperature and rainfall were considered as significant factors in most studies, but some studies showed that meteorological factors alone are not sufficient to explain spatial heterogeneity in disease transmission, which may be associated more with non-climatic factors [48–50]. Rather than rely on raw measures of temperature, 22 studies (12.5%) instead used "temperature suitability" of *Aedes* mosquito vectors, which incorporates a variety of different methods of modelling the temperature constraints on the vector and virus dynamics that are most critical for virus transmission [51]. Six studies used bioclimatic variables that encompassed annual temperature and precipitation ranges, seasonal fluctuations, as well as extreme or constraining factors that capture broader biological patterns [29, 52–56]. Four studies additionally used indicators associated with El Niño Southern Oscillation as covariates [35, 57–59]. Examples of other climatic variables that were included in the reviewed models were diurnal temperature range [8, 60, 61], atmospheric pressure [62, 63], wind speed [64, 65], and duration of sunshine [44, 66, 67].

Population density (n=43, 24.4%) and age distributions (n=25, 14.2%) were often considered in modelling

arboviruses. Many studies found population density to be a significant covariate in their models, demonstrating a positive association with disease transmission, but some studies reported a negative [68, 69] or null association [45, 50, 70]. Human mobility between cities or countries (n=15, 8.5%) was also considered by including travel distance between regions [71, 72] or air travel passenger volume [73–76]. Some studies included daily human mobility data (n=13, 7.4%%), mostly mapped at sub-national scale, with the aim of better representing short-distance high frequency movements such as daily commuting [77, 78]. Seven studies, for yellow fever and dengue, considered vaccination coverage and measures of population immunity from infection in their models [27, 33, 37, 53, 68, 79, 80].

The most common socio-economic variable was gross domestic product (GDP) (n=15, 8.5%), followed by household poverty/income level (n=11, 6.3%) and education level (n=6, 3.4%). A socio-economic strata or a composite index such as human development index, social advantage and disadvantage score (n=6, 3.4%) were also included as socio-economic predictors in some of the reviewed models. Lower neighbourhood socio-economic status was generally associated with increased risk of *Aedes*-borne arbovirus diseases; in regions with established arboviral circulation, community-level factors such as inadequate garbage collection, low income, and lack of access to health care were associated with elevated risk of dengue infections [81–83].

For models fit at the sub-national scale to case incidence data, accompanying direct measurements of the *Aedes* mosquito population improved model predictive performance. Breteau index (BI), which is defined as the number of positive containers per 100 houses, was used as a predictor in three studies [59, 84, 85]. Six studies included location of *Aedes* breeding sites in their models [77, 86–90]. The number of catches of female adult mosquitoes was included in two studies [63, 91]. In the absence of direct measurements of the vector abundance, modelled predictions of "suitability for *Aedes* mosquitoes [92]" have been used, particularly at broad global scales and to make early predictions for emerging Zika epidemics. Six studies included the occurrence or species richness of non-human primates in modelling yellow fever [8, 33, 38, 41, 68, 79].

The most common environmental variable was vegetation index (n=27, 15.3%), followed by altitude/elevation (n=25, 14.2%) and urbanisation (n=22, 12.5%). Some studies found that vegetation was not a key predictor variable and had no association with dengue incidence [39, 93], whereas those considering vegetation in modelling yellow fever generally found that there was a strong and significant vegetation-disease association possibly because of the greater role of the forest-fringe
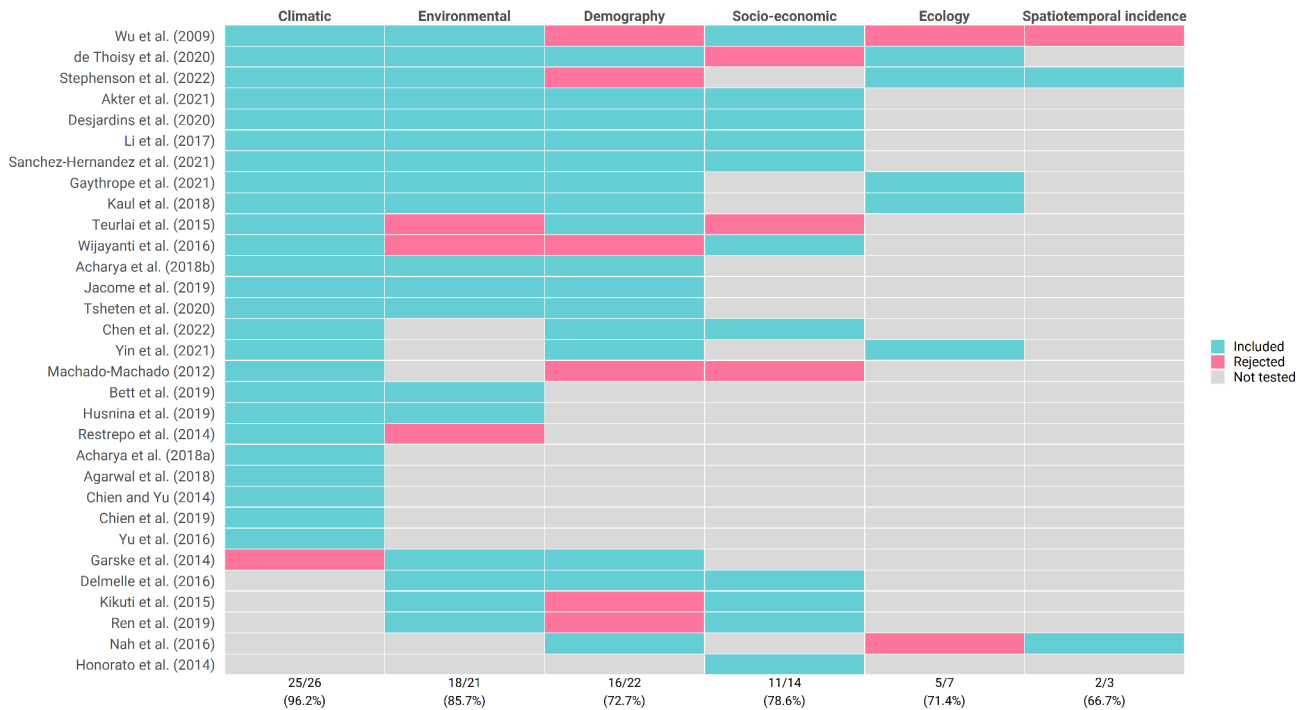
environment in driving spillover from non-human primate reservoirs [38, 40, 41, 69]. Road density and proximity to the road were also included as a predictor in 14 studies (8.0%). More generic categories of land use and land cover type have also been considered in another 13 studies.

Disease incidence across time periods and neighbouring regions were included as covariates in 18 studies (10.2%) to explain contemporaneous disease transmission. Several studies included past case counts lagged by one week to four months to improve temporal prediction accuracy [42, 57, 71, 94–96]. Source country's disease incidence rate was included in studies estimating the risk of Zika virus spread and local transmission in the Americas [73] and Asia-Pacific regions [74].

For each paper, we also examined whether the collinearity among covariates was checked and whether models retained covariates after conducting variable selection procedures. Among the 144 studies excluding those that used mechanistic models or only included random effect terms, only 31 studies (21.5%) tested different combinations of covariates and checked the multicollinearity among them by calculating the correlation coefficient or variance inflation factor, or using principal component analysis. There were 61 studies (42.4%) that did not include any process for selecting variables or checking collinearity (Table S2, Additional file 3). However, it is worth noting that some of these studies may have had a small number of covariates that were selected based on their known or cited ecological or theoretical relevance to disease transmission, which may explain the lack of variable selection process.

For the 31 studies that both checked the multicollinearity of covariates and performed variable selection, we summarised the retention rate of different groups of covariates in the final models (Fig. 6A) [8, 27, 32, 38, 44–50, 52–54, 64, 65, 75, 81, 82, 89, 97–108]. Of 31 studies, 25 studies (96.2%) retained climatic variables when tested. Only one study on dengue [107] tested all six categories and rejected demographic, ecological data and spatiotemporal incidence. Apart from climatic variables, environmental variables were the most commonly used, with 21 studies tested and only three of them rejected, followed by demographic (22 tested and 6 rejected), socio-economic variables (14 tested but 5 rejected). Ecological data (7 tested and 2 rejected) and spatiotemporal incidence (3 tested and 1 rejected) were the least tested and included (Fig. 6A). The most common combinations of retained categories were climatic, environmental, demography, and socio-economic (n=4) [47, 50, 100, 104]. For climatic variables, different measures of temperature and rainfall were tested in reviewed studies. Inclusion of temperature in models differed between studies, with mean temperature often selected over minimum and maximum
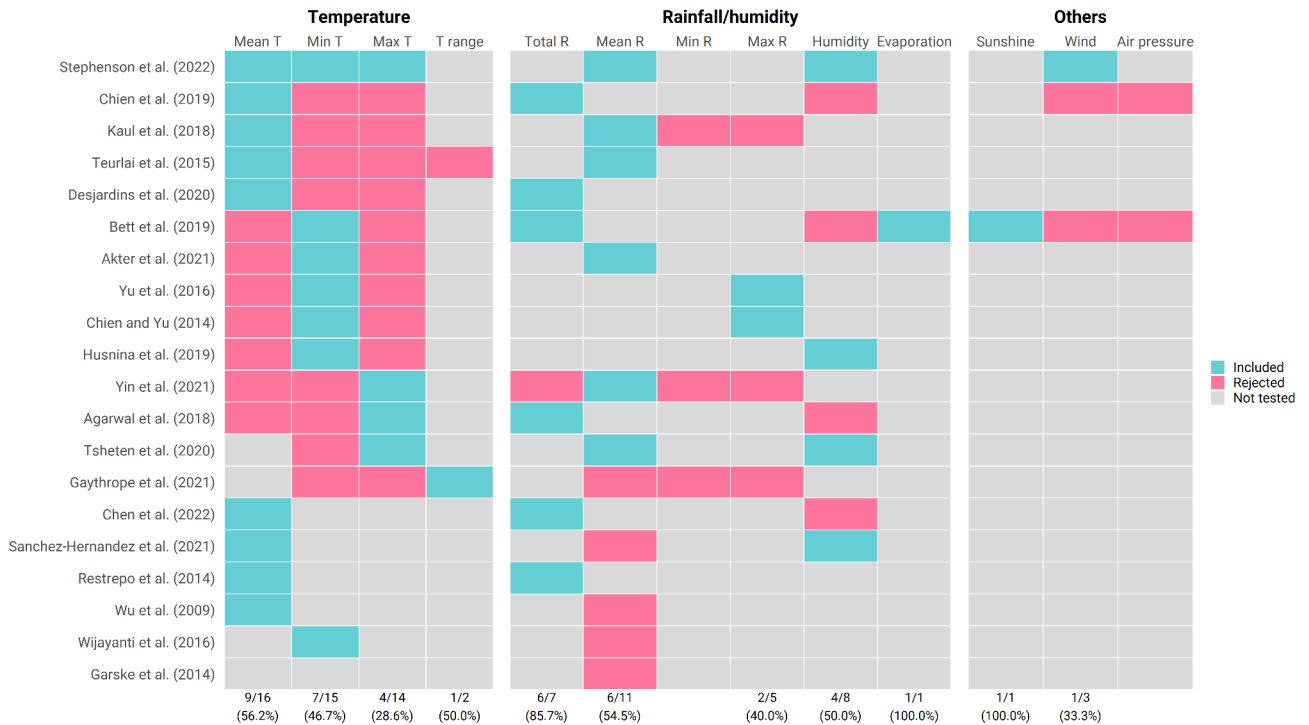
Fig. 6 Covariates included and rejected. (**a**) Selected covariate categories; (**b**) climate variable selections. Mean T: mean temperature; Min T: minimum temperature; Max T: maximum temperature; T range: temperature range; Total R: total rainfall; Mean R: mean rainfall; Min R: minimum rainfall; Max R: maximum rainfall. The values in the bottom represent the number and percentage of studies tested and included the corresponding category of covariates

temperature in nine out of 16 studies (Fig. 6B). We identified that any measures of rainfall (e.g., total or mean rainfall) was preferred over humidity but only seven studies examined the performance of models in which both variables were considered (Fig. 6B). We found that 29 studies have included lagged covariates in their models. The length of the lag periods tested for temperature, humidity and precipitation ranged from 0 to 16 weeks, with most being concentrated between 4 and 12 weeks (Figure S3, Additional file 3). The average lag periods for mean temperature and precipitation tend to be longer in the Americas compared to Western Pacific and SE Asia (Figure S3, Additional file 3).

Modelling framework

Four classes of modelling methods were identified: statistical mixed effect models, statistical fixed effect models, machine learning and mechanistic models (Table S3, Additional file 3). Overall, the most common modelling approaches were types of statistical mixed effect models (n=69, 39.2%), with generalised linear mixed models (GLMM) dominating (n=59), followed by generalised additive mixed models (GAMM) (n=5) and distributed lag non-linear models (DLNM) (n=4). Mixed effect models were often preferred when using areal-type case count data aggregated over distinct geographical areas (e.g. administrative boundaries) (Fig. 7).

Statistical fixed effect models were used in 22.2% of studies, with generalised linear models (GLM) and geographically weighted regression (GWR) as the most used approaches. Since fixed effect models assume that all observations are independent, models used spatial variables to account for spatial relationships. For example, several studies included the coordinates (long, lat) of cases, households, or the centroid of a region [32, 36, 49, 80, 109–111].

A variety of machine learning methods were employed in 27.3% of studies. The most frequently used machine learning methods were maximum entropy (MaxEnt) and boosted regression tree (BRT). They were often used when developing ecological niche or species distribution models using point-referenced occurrence data to describe the environmental suitability of arbovirus transmission, and especially for larger geographical scales (e.g., international scale). Of 22 studies that developed a global risk map of different arbovirus transmission, eleven studies adopted machine learning methods, seven of which used either MaxEnt or BRT [3, 28, 29, 68, 112–114]. Six studies developed and compared the performance of different machine learning methods. For example, Jiang et al. (2018) adapted three different machine learning models, namely backward propagation neural network, gradient boosting machine and random forest, and reported that backward propagation neural network showed the
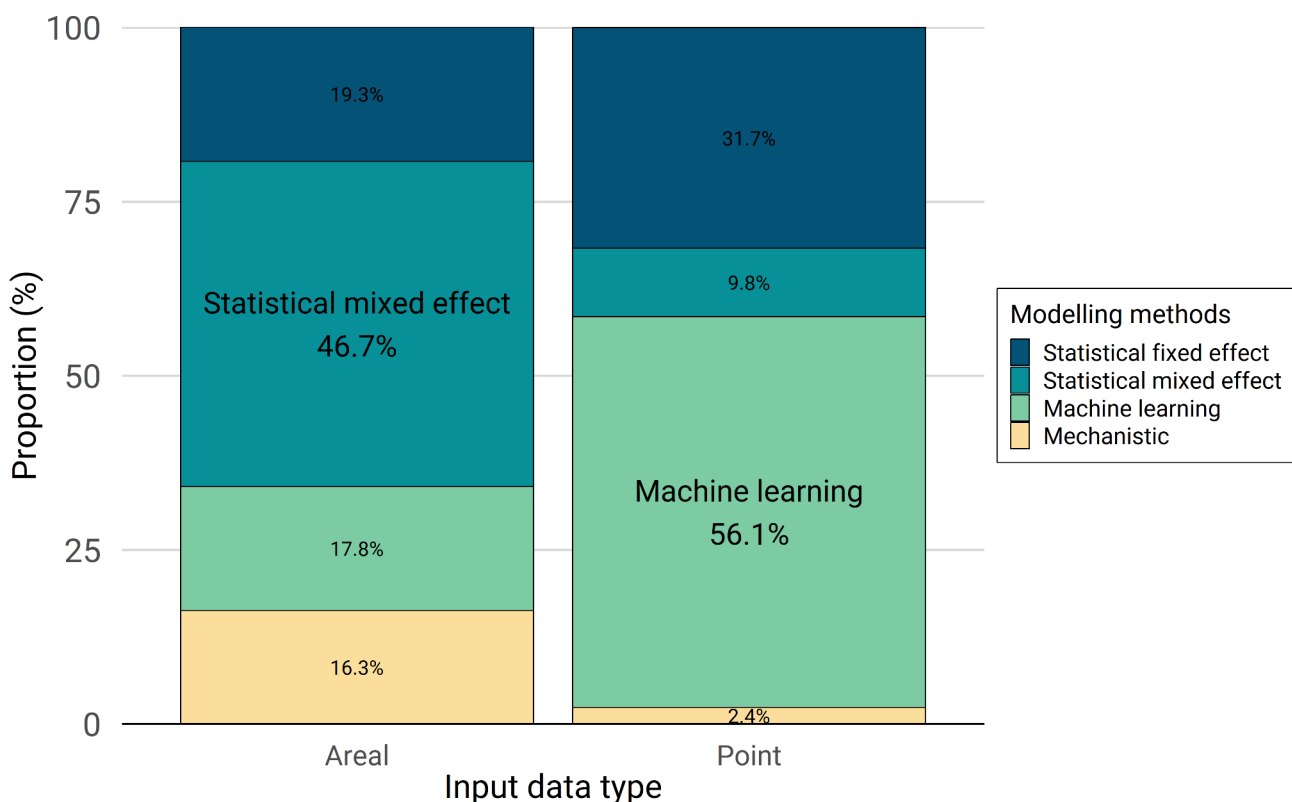

Fig. 7 Modelling framework by input data type

best performance in predicting the global transmission risk of Zika [115]. Two studies generated ensemble model projections of the spatiotemporal dynamics of Zika in Brazil and burden of yellow fever in Africa [33, 116].

Mechanistic models were used in 14.2% of studies, especially compartmental and metapopulation models. Compartmental models e.g. human SEIR - mosquito SIR models were used in six studies to explain the impact of different factors on the transmission dynamics, especially for smaller scales e.g. country or sub-national scale [78, 91, 117–120]. Six studies used metapopulation or network models, all of which considered the connectivity between areas or regions by including the patterns of daily human mobility or air travel data [31, 72, 76, 85, 86, 121]. Five studies used mechanistic mosquito models to produce estimates of temperature suitability, vectorial capacity or basic reproductive number (R0) at the continent or global scale [43, 122–125].

Surprisingly, only 48.9% of studies (n=86) included in this review assessed the predictive performance using cross-validation procedures, such as K-fold cross-validation or random partitioning of data, commonly referred to as "out-of-sample validation". It was more common to perform this type of validation in studies using machine learning methods than in studies using other modelling methods; only 23.1% of studies using fixed effect models performed out-of-sample validation (Fig. 8). Of these

studies, only three studies included model validation on independent test data ("hold-out validation") [60, 126, 127].

The most common model performance evaluation metrics were information criteria (n=82, 46.6%), with Akaike information criteria (AIC) and the Bayesian information criterion (BIC) as the most used metric (Table S4, Additional file 3). Confusion matrix-based metrics were used in 33.0% of studies, with the Receiver Operating Characteristics (ROC) curve most frequently used. Correlation-based metrics were used in 21.6% of studies, especially R-squared. 20 studies (11.4%) did not use any of the metrics described above (Table S4, Additional file 3).

Quality assessment

Using the adapted tool for assessment of modelling study reporting, scores for the reviewed paper ranged from 6 to 18 out of 18. Eleven studies were classified as low quality, 48 as medium quality, 74 as high quality and 43 as very high quality. The median score was 13/18, which is categorised as high quality. Discussions on the generalisability of the developed models were lacking in many papers. Study objectives, settings, and data sources were often unclear in poorly scored studies.
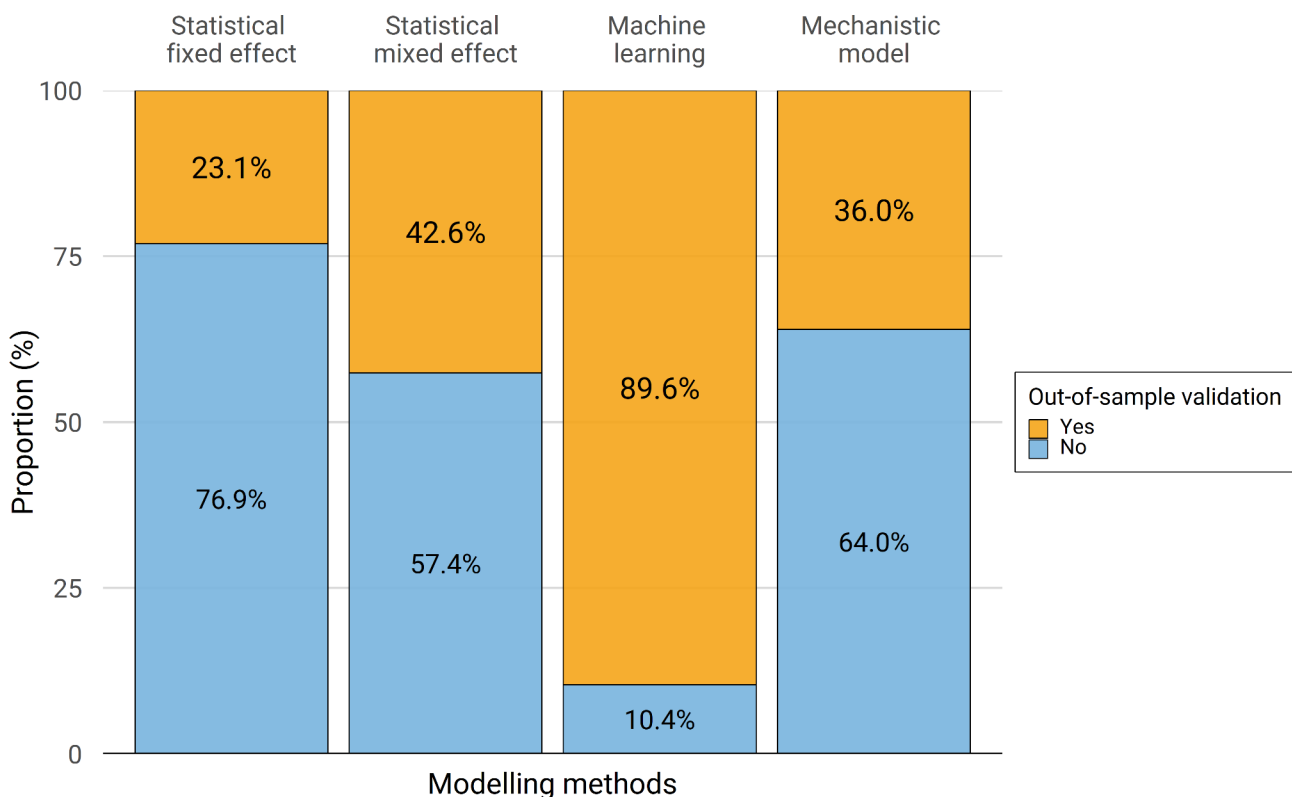

Fig. 8 Out-of-sample validation by modelling framework

Discussion

This review provides a comprehensive overview of arbovirus risk mapping studies in order to identify current best practices and research gaps for each disease. We found that the choices of data, covariates and modelling frameworks were largely determined by the purpose of the map. We identified 22 studies that generated global risk maps, with machine learning-based ecological niche modelling being the most commonly used. These approaches are designed to give a broad overview of the spatial distribution of risk over long-term averages and suggest how it might change under different scenarios of global changes in climate, economics and demographics. Geolocation of disease occurrence data, often combined with high-resolution environmental datasets, were more common for global risk mapping because they encompass large areas and various environmental conditions and provide information about the extent of transmission. However, caution is needed when utilising the outputs of high resolution global risk maps, particularly for informing local decisions due to large data gaps and biases [127] that are not reflected in their highly geographically precise predictions and sometimes do not align with (typically later published) estimates from country-specific models that use more local data.

We found that major epidemics, such as the 2015–2016 Zika epidemic, have acted as catalysts for the development of new risk mapping methods applied in new contexts, possibly due to expanding generation and sharing of data that has accompanied these more recent epidemics. The paucity of data in the early stages of epidemics and similarities between arboviral diseases gives mechanistic modelling approaches an advantage over more data-dependent statistical approaches despite the latter's traditional dominance of the field of risk mapping [14]. As with any model, the predictions are inherently a function of the data available and primary use cases at the time of analysis, and contemporary approaches to mapping risk of diseases like Zika and chikungunya would likely differ substantially from those conducted in the early stages of epidemics. We also show how epidemics have accelerated the use of human movement data in arbovirus risk mapping, and that human movement data is especially valuable to understand long-distance spread since *Aedes* mosquitoes have a limited dispersal capability [128]. Daily commuting and air travel has improved predictions in both statistical and mechanistic modelling approaches, particularly when mapping how the spatial distribution of risk changes over the course of an epidemic.

Studies on modelling yellow fever employed multiple datasets and various approaches, mostly motivated by the need to account for sparse, non-standardised data. They tend to be conducted at continental or country-level scale in African and South American countries with high endemicity for yellow fever transmission or recent outbreaks, for the purpose of evaluation and planning vaccination programs. Inclusion of seroprevalence data and vaccination coverage therefore played a significant role in robust estimation of disease burden and approaches used for yellow fever could be increasingly important for mapping dengue risk as vaccines begin to be rolled out in various countries [129].

In contrast, the majority of publications that use predictive risk mapping for dengue (which accounted for more than 70% of the studies included in this review) now focus on mapping sub-national risk using case incidence data from a country's passive surveillance system. Such models theoretically offer the most potential for direct integration with country surveillance systems and would allow risk maps to directly inform planning, intervention targeting and outbreak response. The proliferation of risk mapping in this domain closely aligns with improvements in routine dengue disease surveillance and sharing of sub-nationally disaggregated data and could be applied to other emerging disease threats if similar approaches to surveillance are adopted. We found that statistical mixed effect models were more commonly implemented than machine-learning approaches for sub-national models, which allow more constraints over the effects of environmental covariates and are easier to implement in Bayesian frameworks, both assets that allow more stability and better representation of uncertainty when making spatio-temporal predictions. Such models blur the boundaries between pure risk mapping (predicting to new spatial locations) and pure hindcasting/forecasting (predicting to new periods of time) and show the added value considering both spatial and temporal information can contribute to each of these applications.

Overall, we found that the quantity and variety of covariates included in arbovirus risk mapping studies has increased in line with growing availability of these variables. While developments over the past decade have focussed on global climate datasets, data on human movement [130] and urban infrastructure [131] are becoming increasingly available and may play important roles in future arbovirus risk mapping studies. Historically, limited data availability has made it difficult to quantify human mobility patterns, requiring models that incorporate gravity or radiation as an approximation [31, 86, 132]. However, the recent emergence of mobile phone data enables real-time tracing of fine-scale movement across large numbers of individuals, although privacy and bias issues remain [133]. The move towards large, open, accessible datasets for vector-borne diseases necessitates not just a more robust data science workforce, but a better motivation and capacity planning for data fluency among primary data producers. While issues of human

subjects and data privacy must remain foremost in contemplating large-scale studies of vector-borne disease risk, nonetheless, leveraging entomological surveillance data, meteorological data, geospatial representation of infrastructure and landscape (e.g., derived from remote sensing, well-resolved built environment enumerations, high-resolution travel network data), and climatological modelling output, is less constrained by international regulations, so identifying the necessary investments and key routes of engagement is a high-level first step to addressing the data gaps.

We found surprisingly few studies conducted robust variable selection procedures. In addition, out-of-sample validation techniques were explicitly stated in only half of the studies reviewed. Statistical and machine learning models, predominantly used in arbovirus risk mapping studies, require a large amount of data and therefore both variable selection and cross-validation are important steps to reduce overfitting and improve model interpretability and predictive accuracy. Although the majority of studies used traditional cross-validation techniques, the use of spatial cross-validation i.e., spatial block bootstrapping is increasingly popular due to its ability to account for spatial dependence in the data [40, 41]. This may help to better test the spatial predictive performance of the model, particularly if there are large heterogeneities in data availability across the study sites, which is common in many arbovirus mapping applications.

Limitations
One limitation of our systematic review is that it focussed on spatial modelling approaches. The conclusions we reach, particularly with reference to drivers of transmission, may differ between risk mapping and temporal prediction models which may be particularly important as the two fields continue to overlap. We also only considered studies published in English, which may affect our conclusions about regional patterns. Additionally, it is possible that some relevant literature, particularly in the form of grey literature, may have been missed as the databases do not contain all journals and university press articles. This is particularly true for locally-relevant geospatial modelling work, which may not have been published in mainstream academic outlets. Finally, we excluded studies that did not assess risk of human infection, excluding a number of studies dealing exclusively with entomological risk or non-human host risk. Finally, to improve study comparability, we chose to exclude studies that were not fit or validated against autochthonous human case data. This excluded several studies that predict risk measures such as vectorial capacity based on laboratory-derived mosquito and virus data. Consequently, we do not include studies that make risk predictions for areas that do not currently experience widespread arbovirus transmission, e.g. Europe. Different factors may influence the risk of importation and establishment in transmission free areas where other vector species may exist (e.g. *Aedes albopictus*) and there is an absence of population-wide arbovirus immunity.

Recommendations for future studies
Based on the best practices and gaps identified in this systematic review we suggest the following recommendations for future arbovirus risk mapping studies:

- Consider the strengths and weaknesses of different data types for different purposes as the choice of data type imposes specific restrictions on the modelling framework and resolution of the prediction. Historically the most common applications have been: occurrence data to map the changing global limits of transmission, short-term aggregated level incidence data to track the geographic spread of epidemics and high spatiotemporal resolution incidence data to understand the roles of different drivers and forecast epidemics.
- Include covariates from multiple domains (climatic, environmental, demographic, socioeconomic, ecological) and test whether their inclusion improves prediction.
- National or subnational studies should consider additional local covariates not available across broader regions, such as data from the arbovirus control programmes, finer scale meteorological resolution data, or infrastructural data from census databases.
- Even with extensive use of covariates, unobserved confounding will still be an issue, particularly for broad scope (national and above) models, meaning that the use of structured spatio-temporal random effects, ideally in a Bayesian mixed effects statistical modelling framework, is preferable to more simplistic fixed effect statistical models.
- Use predictive validation metrics on held out datasets. Ideally using procedures that take into account the unique challenges posed by highly spatially and temporally heterogeneous datasets, such as multiple-fold blocked spatial and temporal cross validation.
- Arbovirus risk mapping is a rapidly developing field with continual improvements in data quantity and representativeness, growing availability of potentially informative covariates and new innovations to model fitting and evaluation. Future arbovirus risk mapping studies should incorporate these new developments and not just rely on the status quo of existing studies.

Conclusion

Spatial modelling can help identify potential risk factors for arbovirus transmission and provide a better understanding of the current and future distribution of arboviruses. We provide a synthesis of covariates and modelling frameworks used for risk mapping of arbovirus, providing an evidence base for developing up-to-date arbovirus risk maps based on current best practices. Although approaches to map arbovirus risk have diversified, it is important to select the data, covariates, models, and evaluation methods based on the purpose of maps, data availability and epidemiological contexts.

List of abbreviations

SE Asia	South-East Asia
WHO	World Health Organization
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
MaxEnt	Maximum entropy
BRT	Boosted regression tree
AIC	Akaike information criteria
BIC	Bayesian information criterion
ROC	Receiver Operating Characteristics

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-023-08717-8.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Data availability

All data generated or analysed during this study are included in this article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Infectious Disease Epidemiology and Dynamics, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
[2]Centre for Mathematical Modelling of Infectious Diseases, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
[3]Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
[4]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[5]High Meadows Environmental Institute, Princeton University, Princeton, NJ, USA
[6]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
[7]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, UK
[8]School of Public Health and Community Medicine, Sahlgrenska Academy, Institute of Medicine, Global Health, University of Gothenburg, Gothenburg, Sweden
[9]Population Health Research Section, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia
[10]Dengue Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico, USA
[11]Department of Biology, University of Oxford, Oxford, UK
[12]School of Geography and the Environment, University of Oxford, Oxford, UK
[13]Oxford School of Global and Area Studies, University of Oxford, Oxford, UK
[14]Department of Biology, Stanford University, Stanford, CA, USA
[15]Department of Epidemic and Pandemic Preparedness and Prevention, World Health Organization, Geneva, Switzerland

[16]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA
[17]Department of Health Metrics Sciences, School of Medicine, University of Washington, Seattle, WA, USA
[18]Department of Geography and Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA
[19]Department of Genetics, University of Cambridge, Cambridge, UK
[20]Department of Public Health and Clinical Medicine, Section of Sustainable Health, Umeå University, Umeå, Sweden

References

1. Jones R, Kulkarni MA, Davidson TMV, Team R-LR, Talbot B. Arbovirus vectors of epidemiological concern in the Americas: a scoping review of entomological studies on Zika, dengue and chikungunya virus vectors. PLoS ONE. 2020;15:e0220753.
2. Leta S, Beyene TJ, Clercq EMD, Amenu K, Kraemer MUG, Revie CW. Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*. Int J Infect Dis. 2018;67:25–35.
3. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496:504–7.
4. Paixão ES, Teixeira MG, Rodrigues LC. Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases. BMJ Glob Health. 2018;3(Suppl 1):e000530.
5. Cattarino L, Rodriguez-Barraquer I, Imai N, Cummings DAT, Ferguson NM. Mapping global variation in dengue transmission intensity. Sci Transl Med. 2020;12:eaax4144.
6. Puntasecca CJ, King CH, LaBeaud AD. Measuring the global burden of chikungunya and Zika viruses: a systematic review. PLoS Negl Trop Dis. 2021;15:e0009055.
7. Li SL, Messina JP, Pybus OG, Kraemer MUG, Gardner L. A review of models applied to the geographic spread of Zika virus. Trans R Soc Trop Med Hyg. 2021;115:956–64.
8. Gaythorpe KA, Hamlet A, Jean K, Garkauskas Ramos D, Cibrelus L, Garske T, et al. The global burden of yellow fever. eLife. 2021;10:e64670.
9. Bassey BE, Braka F, Onyibe R, Kolude OO, Oluwadare M, Oluwabukola A, et al. Changing epidemiology of yellow fever virus in Oyo State, Nigeria. BMC Public Health. 2022;22:467.
10. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of human dengue virus occurrence. Sci Data. 2014;1:140004.
11. Anderson RM, Ferguson NM, Donnelly CA, Anderson RM. Transmission dynamics and epidemiology of dengue: insights from age-stratified sero-prevalence surveys. Philos Trans R Soc Lond B Biol Sci. 1999;354:757–68.
12. Liu-Helmersson J, Brännström Å, Sewe MO, Semenza JC, Rocklöv J. Estimating past, present, and future trends in the global distribution and abundance of the Arbovirus Vector *Aedes aegypti* under climate change scenarios. Front Public Health. 2019;7:148.
13. Kolimenakis A, Heinz S, Wilson ML, Winkler V, Yakob L, Michaelakis A, et al. The role of urbanisation in the spread of *Aedes* mosquitoes and the diseases they transmit—A systematic review. PLoS Negl Trop Dis. 2021;15:e0009631.
14. Lee SA, Jarvis CI, Edmunds WJ, Economou T, Lowe R. Spatial connectivity in mosquito-borne disease models: a systematic review of methods and assumptions. J R Soc Interface. 2021;18:20210096.
15. Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S. Climate change and dengue: a critical and systematic review of quantitative modelling approaches. BMC Infect Dis. 2014;14:167.
16. Sylvestre E, Joachim C, Cécilia-Joseph E, Bouzillé G, Campillo-Gimenez B, Cuggia M, et al. Data-driven methods for dengue prediction and surveillance using real-world and Big Data: a systematic review. PLoS Negl Trop Dis. 2022;16:e0010056.
17. de Lima CL, da Silva ACG, Moreno GMM, Cordeiro da Silva C, Musah A, Aldosery A, et al. Temporal and spatiotemporal arboviruses forecasting by machine learning: a systematic review. Front Public Health. 2022;10:900077.
18. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, et al. Modeling tools for dengue risk mapping - a systematic review. Int J Health Geogr. 2014;13:50.
19. Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. Epidemiol Infect. 2019;147:e33.
20. Yin S, Ren C, Shi Y, Hua J, Yuan H-Y, Tian L-W. A systematic review on modeling methods and influential factors for mapping dengue-related risk in urban settings. Int J Environ Res Public Health. 2022;19:15265.
21. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. PLOS Med. 2009;6:e1000097.
22. R Core Team. R: A language and environment for statistical computing. 2022.
23. Pollett S, Johansson MA, Reich NG, Brett-Major D, Del Valle SY, Venkatramanan S, et al. Recommended reporting items for epidemic forecasting and prediction research: the EPIFORGE 2020 guidelines. PLOS Med. 2021;18:e1003793.
24. Rogers DJ, Wilson AJ, Hay SI, Graham AJ. The global distribution of yellow fever and dengue. Adv Parasitol. 2006;62:181–220.
25. Bettis AA, Jackson ML, Yoon I-K, Breugelmans JG, Goios A, Gubler DJ, et al. The global epidemiology of chikungunya from 1999 to 2020: a systematic literature review to inform the development and introduction of vaccines. PLoS Negl Trop Dis. 2022;16:e0010069.
26. Gianchecchi E, Cianchi V, Torelli A, Montomoli E. Yellow fever: origin, epidemiology, preventive strategies and future prospects. Vaccines. 2022;10:372.
27. Chen Y, Li N, Lourenço J, Wang L, Cazelles B, Dong L, et al. Measuring the effects of COVID-19-related disruption on dengue transmission in southeast Asia and Latin America: a statistical modelling study. Lancet Infect Dis. 2022;22:657–67.
28. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global environmental suitability for Zika virus. eLife. 2016;5:e15272.
29. Tjaden NB, Suk JE, Fischer D, Thomas SM, Beierkuhnlein C, Semenza JC. Modelling the effects of global climate change on Chikungunya transmission in the 21st century. Sci Rep. 2017;7:3813.
30. Johansson MA, Powers AM, Pesik N, Cohen NJ, Staples JE. Nowcasting the spread of chikungunya virus in the Americas. PLoS ONE. 2014;9:e104915.
31. Chadsuthi S, Althouse BM, Iamsirithaworn S, Triampo W, Grantz KH, Cummings DAT. Travel distance and human movement predict paths of emergence and spatial spread of Chikungunya in Thailand. Epidemiol Infect. 2018;146:1654–62.
32. Garske T, Kerkhove MDV, Yactayo S, Ronveaux O, Lewis RF, Staples JE, et al. Yellow fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. PLOS Med. 2014;11:e1001638.
33. Perkins TA, Huber JH, Tran QM, Oidtman RJ, Walters MK, Siraj AS, et al. Burden is in the eye of the beholder: sensitivity of yellow fever disease burden estimates to modeling assumptions. Sci Adv. 2021;7:eabg5033.
34. Lauer SA, Sakrejda K, Ray EL, Keegan LT, Bi Q, Suangtho P, et al. Prospective forecasts of annual dengue hemorrhagic fever incidence in Thailand, 2010–2014. Proc Natl Acad Sci. 2018;115:E2175–82.
35. Yip S, Che Him N, Jamil NI, He D, Sahu SK. Spatio-temporal detection for dengue outbreaks in the Central Region of Malaysia using climatic drivers at mesoscale and synoptic scale. Clim Risk Manag. 2022;36:100429.
36. Gaythorpe KAM, Jean K, Cibrelus L, Garske T. Quantifying model evidence for yellow fever transmission routes in Africa. PLOS Comput Biol. 2019;15:e1007355.
37. Jean K, Hamlet A, Benzler J, Cibrelus L, Gaythorpe KAM, Sall A, et al. Eliminating yellow fever epidemics in Africa: vaccine demand forecast and impact modelling. PLoS Negl Trop Dis. 2020;14:e0008304.
38. Kaul RB, Evans MV, Murdock CC, Drake JM. Spatio-temporal spillover risk of yellow fever in Brazil. Parasit Vectors. 2018;11:488.
39. Rogers DJ, Suk JE, Semenza JC. Using global maps to predict the risk of dengue in Europe. Acta Trop. 2014;129:1–14.
40. Hamlet A, Ramos DG, Gaythorpe KAM, Romano APM, Garske T, Ferguson NM. Seasonality of agricultural exposure as an important predictor of seasonal yellow fever spillover in Brazil. Nat Commun. 2021;12:3647.
41. Hamlet A, Gaythorpe KAM, Garske T, Ferguson NM. Seasonal and inter-annual drivers of yellow fever transmission in South America. PLoS Negl Trop Dis. 2021;15:e0008974.
42. Lowe R, Coelho CA, Barcellos C, Carvalho MS, Catão RDC, Coelho GE et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. eLife 5:e11285.
43. Caminade C, Turner J, Metelmann S, Hesson JC, Blagrove MSC, Solomon T, et al. Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. Proc Natl Acad Sci U S A. 2017;114:119–24.
44. Bett B, Grace D, Lee HS, Lindahl J, Nguyen-Viet H, Phuc P-D, et al. Spatiotemporal analysis of historical records (2001–2012) on dengue fever in Vietnam

and development of a statistical model for forecasting risk. PLoS ONE. 2019;14:e0224353.

45. Tsheten T, Clements ACA, Gray DJ, Wangchuk S, Wangdi K. Spatial and temporal patterns of dengue incidence in Bhutan: a Bayesian analysis. Emerg Microbes Infect. 2020;9:1360–71.

46. Husnina Z, Clements ACA, Wangdi K. Forest cover and climate as potential drivers for dengue fever in Sumatra and Kalimantan 2006–2016: a spatiotemporal analysis. Trop Med Int Health TM IH. 2019;24:888–98.

47. Desjardins MR, Eastin MD, Paul R, Casas I, Delmelle EM. Space–time conditional autoregressive modeling to estimate neighborhood-level risks for dengue fever in Cali, Colombia. Am J Trop Med Hyg. 2020;103:2040–53.

48. Chien L-C, Yu H-L. Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence. Environ Int. 2014;73:46–56.

49. Chien L-C, Sy F, Pérez A. Identifying high risk areas of Zika virus infection by meteorological factors in Colombia. BMC Infect Dis. 2019;19:888.

50. Akter R, Hu W, Gatton M, Bambrick H, Cheng J, Tong S. Climate variability, socio-ecological factors and dengue transmission in tropical Queensland, Australia: a Bayesian spatial analysis. Environ Res. 2021;195:110285.

51. Brady OJ, Golding N, Pigott DM, Kraemer MUG, Messina JP, Reiner RC Jr, et al. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. Parasit Vectors. 2014;7:338.

52. Acharya BK, Cao C, Xu M, Khanal L, Naeem S, Pandit S. Present and future of dengue fever in Nepal: mapping climatic suitability by ecological niche model. Int J Environ Res Public Health. 2018;15:187.

53. de Thoisy B, Silva NIO, Sacchetto L, Trindade G, de Drumond S. Spatial epidemiology of yellow fever: identification of determinants of the 2016–2018 epidemics and at-risk areas in Brazil. PLoS Negl Trop Dis. 2020;14:e0008691.

54. Jácome G, Vilela P, Yoo C. Present and future incidence of dengue fever in Ecuador nationwide and coast region scale using species distribution modeling for climate variability's effect. Ecol Model. 2019;400:60–72.

55. Wu W, Ren H, Lu L. Increasingly expanded future risk of dengue fever in the Pearl River Delta, China. PLoS Negl Trop Dis. 2021;15:e0009745.

56. Flórez-Lozano K, Navarro-Lechuga E, Llinás-Solano H, Tuesca-Molina R, Sisa-Camargo A, Mercado-Reyes M, et al. Spatial distribution of the relative risk of Zika virus disease in Colombia during the 2015–2016 epidemic from a Bayesian approach. Int J Gynecol Obstet. 2020;148:55–60.

57. Lowe R, Bailey TC, Stephenson DB, Jupp TE, Graham RJ, Barcellos C, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. Stat Med. 2013;32:864–83.

58. Lowe R, Bailey TC, Stephenson DB, Graham RJ, Coelho CAS, Sá Carvalho M, et al. Spatio-temporal modelling of climate-sensitive disease risk: towards an early warning system for dengue in Brazil. Comput Geosci. 2011;37:371–81.

59. Yu H-L, Yang S-J, Yen H-J, Christakos G. A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. Stoch Environ Res Risk Assess. 2011;25:485–94.

60. Ferdousi T, Cohnstaedt LW, Scoglio CM. A windowed correlation-based feature selection method to improve time series prediction of dengue fever cases. IEEE Access. 2021;9:141210–22.

61. Sharmin S, Glass K, Viennet E, Harley D. Geostatistical mapping of the seasonal spread of under-reported dengue cases in Bangladesh. PLoS Negl Trop Dis. 2018;12:e0006947.

62. Mussumeci E, Codeço Coelho F. Large-scale multivariate forecasting models for dengue - LSTM versus random forest regression. Spat Spatio-Temporal Epidemiol. 2020;35:100372.

63. Sedda L, Taylor BM, Eiras AE, Marques JT, Dillon RJ. Using the intrinsic growth rate of the mosquito population improves spatio-temporal dengue risk estimation. Acta Trop. 2020;208:105519.

64. Stephenson C, Coker E, Wisely S, Liang S, Dinglasan RR, Lednicky JA. Imported dengue case numbers and local climatic patterns are associated with dengue virus transmission in Florida, USA. Insects. 2022;13:163.

65. Siddiq A, Shukla N, Pradhan B. Spatio-temporal modelling of dengue fever cases in Saudi Arabia using socio-economic, climatic and environmental factors. Geocarto Int. 2022;1–25.

66. Jaya IGNM, Folmer H. Spatiotemporal high-resolution prediction and mapping: methodology and application to dengue disease. J Geogr Syst. 2022. https://doi.org/10.1007/s10109-021-00368-0.

67. Jaya IGNM, Folmer H. Identifying spatiotemporal clusters by means of agglomerative hierarchical clustering and Bayesian regression analysis with spatiotemporally varying coefficients: methodology and application to dengue disease in Bandung, Indonesia. Geogr Anal. 2021;53:767–817.

68. Shearer FM, Longbottom J, Browne AJ, Pigott DM, Brady OJ, Kraemer MUG, et al. Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. Lancet Glob Health. 2018;6:e270–8.

69. Servadio JL, Muñoz-Zanzi C, Convertino M. Environmental determinants predicting population vulnerability to high yellow fever incidence. R Soc Open Sci. 9:220086.

70. McHale TC, Romero-Vivas CM, Fronterre C, Arango-Padilla P, Waterlow NR, Nix CD, et al. Spatiotemporal heterogeneity in the distribution of chikungunya and Zika virus case incidences during their 2014 to 2016 epidemics in Barranquilla, Colombia. Int J Environ Res Public Health. 2019;16:1759.

71. Lim JT, Dickens BSL, Mao Y, Kwak CW, Ng LC, Cook AR. Explicit characterization of human population connectivity reveals long run persistence of interregional dengue shocks. J R Soc Interface. 2020;17.

72. Yu H-L, Angulo JM, Cheng M-H, Wu J, Christakos G. An online spatiotemporal prediction model for dengue fever epidemic in Kaohsiung (Taiwan). Biom J. 2014;56:428–40.

73. Gardner LM, Bóta A, Gangavarapu K, Kraemer MUG, Grubaugh ND. Inferring the risk factors behind the geographical spread and transmission of Zika in the Americas. PLoS Negl Trop Dis. 2018;12:e0006194.

74. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Model-informed risk assessment for Zika virus outbreaks in the Asia-Pacific regions. J Infect. 2017;74:484–91.

75. Nah K, Mizumoto K, Miyamatsu Y, Yasuda Y, Kinoshita R, Nishiura H. Estimating risks of importation and local transmission of Zika virus infection. PeerJ. 2016;4:e1904.

76. Zhang Q, Sun K, Chinazzi M, Pastore y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika virus in the Americas. Proc Natl Acad Sci. 2017;114:E4334–43.

77. Ong J, Liu X, Rajarethinam J, Kok SY, Liang S, Tang CS, et al. Mapping dengue risk in Singapore using Random Forest. PLoS Negl Trop Dis. 2018;12:e0006587.

78. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proc Natl Acad Sci. 2015;112:11887–92.

79. Childs ML, Nova N, Colvin J, Mordecai EA. Mosquito and primate ecology predict human risk of yellow fever virus spillover in Brazil. Philos Trans R Soc B Biol Sci. 2019;374:20180335.

80. Honório NA, Nogueira RMR, Codeço CT, Carvalho MS, Cruz OG, de Magalhães M. Spatial evaluation and modeling of dengue seroprevalence and vector density in Rio De Janeiro, Brazil. PLoS Negl Trop Dis. 2009;3:e545.

81. Kikuti M, Cunha GM, Paploski IA, Kasper AM, Silva MM, Tavares AS, et al. Spatial distribution of dengue in a Brazilian urban slum setting: role of socioeconomic gradient in disease risk. PLoS Negl Trop Dis. 2015;9:e0003937.

82. Honorato T, Lapa PP, de Sales A, Reis-Santos CMM, Tristão-Sá B, Bertolde R. Spatial analysis of distribution of dengue cases in Espírito Santo, Brazil, in 2010: use of Bayesian model. Rev Bras Epidemiol. 2014;17:150–9.

83. Costa JV, Donalisio MR, de Silveira LV. Spatial distribution of dengue incidence and socio-environmental conditions in Campinas, São Paulo State, Brazil, 2007. Cad Saúde Pública. 2013;29:1522–32.

84. Lin C-H, Wen T-H. Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. Int J Environ Res Public Health. 2011;8:2798–815.

85. Zhu G, Liu T, Xiao J, Zhang B, Song T, Zhang Y, et al. Effects of human mobility, temperature and mosquito control on the spatiotemporal transmission of dengue. Sci Total Environ. 2019;651:969–78.

86. Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, et al. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med. 2015;13:102.

87. Cordeiro R, Donalisio MR, Andrade VR, Mafra AC, Nucci LB, Brown JC, et al. Spatial distribution of the risk of dengue fever in southeast Brazil, 2006–2007. BMC Public Health. 2011;11:355.

88. Abdul Rahm S, Rahim A, Mallongi A. Forecasting of dengue disease incident risks using non-stationary spatial of geostatistics model in Bone Regency Indonesia. J Entomol. 2016;14:49–57.

89. Yin MS, Bicout DJ, Haddawy P, Schöning J, Laosiritaworn Y, Sa-angchai P. Added-value of mosquito vector breeding sites from street view images in the risk mapping of dengue incidence in Thailand. PLoS Negl Trop Dis. 2021;15:e0009122.

90. Withanage GP, Gunawardana M, Viswakula SD, Samaraweera K, Gunawardena NS, Hapugoda MD. Multivariate spatio-temporal approach to identify vulnerable localities in dengue risk areas using Geographic Information System (GIS). Sci Rep. 2021;11:4080.

91. Li R, Xu L, Bjørnstad ON, Liu K, Song T, Chen A, et al. Climate-driven variation in mosquito density predicts the spatiotemporal dynamics of dengue. Proc Natl Acad Sci. 2019;116:3624–9.

92. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. Albopictus*. eLife. 2015;4:e08347.

93. Ye J, Moreno-Madriñán MJ. Comparing different spatio-temporal modeling methods in dengue fever data analysis in Colombia during 2012–2015. Spat Spatio-Temporal Epidemiol. 2020;34:100360.

94. Lowe R, Barcellos C, Coelho CAS, Bailey TC, Coelho GE, Graham R, et al. Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. Lancet Infect Dis. 2014;14:619–26.

95. Liu K, Zhang M, Xi G, Deng A, Song T, Li Q, et al. Enhancing fine-grained intra-urban dengue forecasting by integrating spatial interactions of human movements between urban regions. PLoS Negl Trop Dis. 2020;14:e0008924.

96. Liu K, Yin L, Zhang M, Kang M, Deng A-P, Li Q-L, et al. Facilitating fine-grained intra-urban dengue forecasting by integrating urban environments measured from street-view images. Infect Dis Poverty. 2021;10:40.

97. Acharya BK, Cao C, Lakes T, Chen W, Naeem S, Pandit S. Modeling the spatially varying risk factors of dengue fever in Jhapa district, Nepal, using the semiparametric geographically weighted regression model. Int J Biometeorol. 2018;62:1973–86.

98. Agarwal N, Koti SR, Saran S, Kumar AS. Data mining techniques for predicting dengue outbreak in geospatial domain using weather parameters for New Delhi, India. Curr Sci. 2018;114:2281–91.

99. Delmelle E, Hagenlocher M, Kienberger S, Casas I. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. Acta Trop. 2016;164:169–76.

100. Li Q, Ren H, Zheng L, Cao W, Zhang A, Zhuang D, et al. Ecological niche modeling identifies fine-scale areas at high risk of dengue fever in the Pearl River Delta, China. Int J Environ Res Public Health. 2017;14:619.

101. Machado-Machado EA. Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. Appl Geogr. 2012;33:82–93.

102. Ren H, Wu W, Li T, Yang Z. Urban villages as transfer stations for dengue fever epidemic: a case study in the Guangzhou, China. PLoS Negl Trop Dis. 2019;13:e0007350.

103. Restrepo AC, Baker P, Clements ACA. National spatial and temporal patterns of notified dengue cases, Colombia 2007–2010. Trop Med Int Health. 2014;19:863–71.

104. Sánchez-Hernández D, Aguirre-Salado CA, Sánchez-Díaz G, Aguirre-Salado AI, Soubervielle-Montalvo C, Reyes-Cárdenas O, et al. Modeling spatial pattern of dengue in North Central Mexico using survey data and logistic regression. Int J Environ Health Res. 2021;31:872–88.

105. Teurlai M, Menkès CE, Cavarero V, Degallier N, Descloux E, Grangeon J-P, et al. Socio-economic and climate factors associated with dengue fever spatial heterogeneity: a worked example in New Caledonia. PLoS Negl Trop Dis. 2015;9:e0004211.

106. Wijayanti SPM, Porphyre T, Chase-Topping M, Rainey SM, McFarlane M, Schnettler E, et al. The importance of socio-economic versus environmental risk factors for reported dengue cases in Java, Indonesia. PLoS Negl Trop Dis. 2016;10:e0004964.

107. Wu P-C, Lay J-G, Guo H-R, Lin C-Y, Lung S-C, Su H-J. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. Sci Total Environ. 2009;407:2224–33.

108. Yu H-L, Lee C-H, Chien L-C. A spatiotemporal dengue fever early warning model accounting for nonlinear associations with hydrological factors: a Bayesian maximum entropy approach. Stoch Environ Res Risk Assess. 2016;30:2127–41.

109. Siqueira-Junior JB, Maciel IJ, Barcellos C, Souza WV, Carvalho MS, Nascimento NE, et al. Spatial point analysis based on dengue surveys at household level in central Brazil. BMC Public Health. 2008;8:361.

110. Lee SA, Economou T, de Castro Catão R, Barcellos C, Lowe R. The impact of climate suitability, urbanisation, and connectivity on the expansion of dengue in 21st century Brazil. PLoS Negl Trop Dis. 2021;15:e0009773.

111. Mutucumarana CP, Bodinayake CK, Nagahawatte A, Devasiri V, Kurukulasooriya R, Anuradha T, et al. Geospatial analysis of dengue emergence in rural areas in the Southern Province of Sri Lanka. Trans R Soc Trop Med Hyg. 2020;114:408–14.

112. Samy AM, Thomas SM, Wahed AAE, Cohoon KP, Peterson AT. Mapping the global geographic potential of Zika virus spread. Mem Inst Oswaldo Cruz. 2016;111:559.

113. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nat Microbiol. 2019;4:1508–15.

114. Nsoesie EO, Kraemer MU, Golding N, Pigott DM, Brady OJ, Moyes CL, et al. Global distribution and environmental suitability for Chikungunya virus, 1952 to 2015. Eurosurveillance. 2016;21:30234.

115. Jiang D, Hao M, Ding F, Fu J, Li M. Mapping the transmission risk of Zika virus using machine learning models. Acta Trop. 2018;185:391–9.

116. Soliman M, Lyubchich V, Gel YR. Ensemble forecasting of the Zika space-time spread with topological data analysis. Environmetrics. 2020;31:e2629.

117. Benkimoun S, Atyame C, Haramboure M, Degenne P, Thébault H, Dehecq J-S, et al. Dynamic mapping of dengue basic reproduction number. Results Phys. 2021;29:104687.

118. Mukhsar, Ahmar A, El-Khawaga AM, Sayed H. Bayesian convolution for stochastic epidemic model. Intell Autom Soft Comput. 2022;34:1175–86.

119. Romeo-Aznar V, Picinini Freitas L, Gonçalves Cruz O, King AA, Pascual M. Finescale heterogeneity in population density predicts wave dynamics in dengue epidemics. Nat Commun. 2022;13:996.

120. Butterworth MK, Morin CW, Comrie AC. An analysis of the potential impact of climate change on dengue transmission in the Southeastern United States. Env Health Perspect. 2017;125:579–85.

121. O'Reilly KM, Lowe R, Edmunds WJ, Mayaud P, Kucharski A, Eggo RM, et al. Projecting the end of the Zika virus epidemic in Latin America: a modelling analysis. BMC Med. 2018;16:180.

122. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. PLoS Negl Trop Dis. 2017;11:e0005568.

123. Ryan SJ, Carlson CJ, Tesla B, Bonds MH, Ngonghala CN, Mordecai EA, et al. Warming temperatures could expose more than 1.3 billion new people to Zika virus risk by 2050. Glob Change Biol. 2021;27:84–93.

124. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of *Aedes*-borne virus transmission risk with climate change. PLoS Negl Trop Dis. 2019;13:e0007213.

125. Tesla B, Demakovsky LR, Mordecai EA, Ryan SJ, Bonds MH, Ngonghala CN, et al. Temperature drives Zika virus transmission: evidence from empirical and mathematical models. Proc R Soc B Biol Sci. 2018;285:20180795.

126. Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk of Zika in real time. BMC Med. 2019;17:171.

127. O'Reilly KM, Hendrickx E, Kharisma DD, Wilastonegoro NN, Carrington LB, Elyazar IRF, et al. Estimating the burden of dengue and the impact of release of wMel *Wolbachia*-infected mosquitoes in Indonesia: a modelling study. BMC Med. 2019;17:172.

128. Verdonschot PFM, Besse-Lototskaya AA. Flight distance of mosquitoes (Culicidae): a metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. Limnologica. 2014;45:69–79.

129. Zorlu G, Fleck F. Dengue vaccine roll-out: getting ahead of the game. Bull World Health Organ. 2011;89:476–7.

130. Kraemer MUG, Sadilek A, Zhang Q, Marchal NA, Tuli G, Cohn EL, et al. Mapping global variation in human mobility. Nat Hum Behav. 2020;4:800–10.

131. Tusting LS, Bisanzio D, Alabaster G, Cameron E, Cibulskis R, Davies M, et al. Mapping changes in housing in sub-saharan Africa from 2000 to 2015. Nature. 2019;568:391–4.

132. Kraemer MUG, Faria NR, Reiner RC, Golding N, Nikolay B, Stasse S, et al. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. Lancet Infect Dis. 2017;17:330–8.

133. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJE. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. J Infect Dis. 2016;214 suppl4:414–20.

Publisher's Note