

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

4
5 **Ah-Young Lim^{1,2*}, Yalda Jafari^{3,4}, Jamie M. Caldwell⁵, Hannah E. Clapham⁶, Katy A. M.**
6 **Gaythorpe⁷, Laith Hussain-Alkhateeb⁸, Michael A. Johansson⁹, Moritz U. G. Kraemer¹⁰, Richard**
7 **J. Maude^{3,4}, Clare P. McCormack⁷, Jane P. Messina^{11,12}, Erin A. Mordecai¹³, Ingrid B. Rabe¹⁴,**
8 **Robert C. Reiner Jr^{15,16}, Sadie J. Ryan¹⁷, Henrik Salje¹⁸, Jan C. Semenza¹⁹, Diana P. Rojas¹⁴,**
9 **Oliver J. Brady^{1,2}**

10
11 **Author affiliations**

12 ¹ Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health,
13 London School of Hygiene & Tropical Medicine, London, United Kingdom

14 ² Centre for Mathematical Modelling of Infectious Diseases, Faculty of Epidemiology and Population
15 Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

16 ³ Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University,
17 Bangkok, Thailand

18 ⁴ Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of
19 Oxford, Oxford, United Kingdom

20 ⁵ High Meadows Environmental Institute, Princeton University, Princeton, New Jersey, United States

21 ⁶ Saw Swee Hock School of Public Health, National University of Singapore and National University
22 Health System, Singapore, Singapore

23 ⁷ MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College
24 London, London, United Kingdom

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

25 ⁸ Global Health, School of Public Health and Community Medicine, Sahlgrenska Academy, Institute of
26 Medicine, University of Gothenburg, Gothenburg, Sweden

27 ⁹ Dengue Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San
28 Juan, Puerto Rico, United States

29 ¹⁰ Department of Biology, University of Oxford, Oxford, United Kingdom

30 ¹¹ School of Geography and the Environment, University of Oxford, Oxford, United Kingdom

31 ¹² Oxford School of Global and Area Studies, University of Oxford, Oxford, United Kingdom

32 ¹³ Department of Biology, Stanford University, Stanford, California, United States

33 ¹⁴ Department of Epidemic and Pandemic Preparedness and Prevention, World Health Organization,
34 Geneva, Switzerland

35 ¹⁵ Institute for Health Metrics and Evaluation, University of Washington, Seattle, Washington, United
36 States

37 ¹⁶ Department of Health Metrics Sciences, School of Medicine, University of Washington, Seattle,
38 Washington, United States

39 ¹⁷ Department of Geography and Emerging Pathogens Institute, University of Florida, Gainesville,
40 Florida, United States

41 ¹⁸ Department of Genetics, University of Cambridge, Cambridge, United Kingdom

42 ¹⁹ Department of Public Health and Clinical Medicine, Section of Sustainable Health, Umeå University,
43 Umeå, Sweden

44

45

46 * Corresponding author

47 E-mail: Ahyoung.Lim@lshtm.ac.uk

48

49

50 **Abstract**

51 **Background**

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

58

59 **Methods**

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

65

66 **Results**

67 We found a growing number of arbovirus risk mapping studies across all endemic regions and
68 arboviral diseases, with a total of 183 papers published 2002-2022 with the largest increases shortly
69 following major epidemics. Three dominant use cases emerged: i) global maps to identify limits of
70 transmission, estimate burden and assess impacts of future global change, ii) regional models used to
71 predict the spread of major epidemics between countries and iii) national and sub-national models that
72 use local datasets to better understand transmission dynamics to improve outbreak detection and
73 response. Temperature and rainfall were the most popular choice of covariates (included in 50% and
74 40% of studies respectively) but variables such as human mobility are increasingly being included.
75 Surprisingly, few studies (22%, 33/148) robustly tested combinations of covariates from different

76 domains (e.g. climatic, sociodemographic, ecological, etc) and only 48% of studies assessed
77 predictive performance via out-of-sample validation procedures.

78

79 **Conclusions**

80 Here we show that approaches to map risk for different arboviruses have diversified in response to
81 changing use cases, epidemiology and data availability. We outline specific recommendations for
82 future studies regarding aims and data choice, covariate selection, model formulation and evaluation.

83

84 **Author Summary**

85 *Aedes*-borne arboviruses such as dengue, Zika, chikungunya, and yellow fever pose a growing global
86 threat. It is crucial to map their risk to target interventions and control their spread. A review of 183
87 studies found that risk mapping methods have evolved over time to respond to changing epidemiology
88 and data availability. Initially, mapping risk involved using data from multiple areas and satellite
89 imagery to develop models predicting transmission risk on a global or continental scale. Following
90 Zika and chikungunya epidemics, mechanistic models based on national-level incidence data have
91 been utilised to track the spread of epidemics across countries. The use of case-based surveillance
92 systems has enabled more precise and detailed predictions at sub-national levels. Of the studies
93 reviewed, half included temperature and rainfall as covariates, and human mobility was increasingly
94 accounted for in arbovirus risk mapping. However, only 33 of the 148 studies robustly selected the
95 variables included in their predictions, and only half of the studies assessed their accuracy against new
96 data. The review suggests that future risk mapping studies should consider the purpose of the map,
97 data quality, and methodological innovations to improve accuracy of risk maps to ensure they are
98 useful for informing control of *Aedes*-borne arboviruses.

99 **Background**

100 Arboviruses, commonly referred to as arthropod-borne viruses, are a wide range of viral
101 pathogens transmitted through the bite of arthropods such as mosquitoes and ticks. The term arbovirus
102 does not refer to a distinct taxonomic group, but the viruses have similar transmission mechanisms,
103 which makes information gained from one virus potentially useful in understanding and preventing
104 the spread of other viruses [1]. In this paper, we focus on *Aedes (Stegomyia)*-borne arboviruses,
105 including dengue, Zika, chikungunya, and yellow fever, which are of particular concern due to their
106 high disease burden and life-threatening health consequences [2]. The geographical spread and burden
107 of this group of arboviruses have been rapidly increasing in recent years. It has been estimated that
108 100-400 million dengue infections occur each year worldwide, mainly in South America and South-
109 East Asia (SE Asia), with the disease threatening to spread to new regions including Europe [3–5].
110 Zika and chikungunya viruses were first identified in Africa and Asia, but emerged and rapidly spread
111 throughout the Americas between 2013 and 2015, likely due to a combination of suitable climatic
112 factors, increasing international air travel and possible immunological drivers [6,7]. The Zika
113 outbreak received global attention due to its link to congenital and neurological complications,
114 resulting in the declaration of a Public Health Emergency of International Concern by the World
115 Health Organization (WHO) in 2016 [7]. Chikungunya is frequently accompanied by joint pain and
116 rheumatic manifestations that can persist for a long time and have a significant impact on the quality
117 of life of affected individuals [4]. Yellow fever is endemic in tropical and subtropical countries of
118 South America and Africa, with an estimated number of 109,000 severe infections and 51,000 deaths
119 in 2018 [8]. Among the *Aedes*-borne arboviruses, yellow fever is the only one that has a safe and
120 effective vaccine available for humans. A sylvatic cycle between non-human primate reservoirs and
121 mosquitoes is the most common source of yellow fever virus infection; however, humans can also
122 become infected through the urban cycle, which can potentially lead to large outbreaks, as recently
123 seen in Angola, Nigeria and the Democratic Republic of the Congo [8,9]. As these *Aedes*-borne
124 arboviruses share a common mechanism of transmission, the WHO launched the Global Arbovirus
125 Initiative in 2022, which includes the aim of developing a comprehensive risk monitoring and early

126 detection tool that will allow countries to assess global risk of different *Aedes*-borne viruses,
127 strengthen vector control, and develop global systems and strategies to monitor and reduce the risk in
128 the local, regional, and national levels. This initiative identified reviewing the drivers of spatial
129 arbovirus risk at global and regional levels as a key priority.

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

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

154 populations. The inherent spatial and temporal patterns of arbovirus transmission are therefore the
155 result of the complex interactions of multiple factors, likely differing between arbovirus, location and
156 spatial scale.

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

166 Previous systematic reviews have been conducted to identify and characterise dengue
167 transmission models focused on predicting trends over time (hindcasting with the goal of developing
168 forecasting systems) as opposed to spatially explicit prediction (risk mapping) [15–17]. Some of these
169 systematic reviews included risk mapping studies but they have been limited to just a single arbovirus,
170 usually dengue [7,18–20]. Although arbovirus risk mapping studies have become more diverse and
171 advanced, to our knowledge, there are no systematic reviews that consider the important similarities
172 and differences among arboviruses. Therefore, this study aims to identify and review studies that map
173 *Aedes* mosquito-transmitted arbovirus risk in humans, and to characterise epidemiological data,
174 covariates, modelling frameworks and methods of evaluation used.

175 **Methods**

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

178

179 **Search strategy**

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

188 **Selection process**

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

194 **Inclusion/exclusion criteria**

195 Articles must be peer-reviewed, published in English and contain a spatial model that
196 investigates the transmission of the arboviruses to humans. Spatial models were defined as models
197 that included geographically realistic and explicit representations of more than one spatial location.
198 While our primary focus was to review spatial models, spatiotemporal models were also included.
199 There were no geographical or publishing date restrictions applied. We only included models that
200 made predictions of some measure of the population-level virus infection risk, including but not
201 limited to occurrence, incidence, prevalence, and proxies of transmission risk (e.g. reproduction

202 number). Studies where the model was developed and/or validated in a previous paper were also
203 included.

204 Articles were excluded if they only modelled transmission to vectors or non-human hosts or
205 were exclusively dealing with occurrence of or suitability for the mosquito (e.g. vector suitability).
206 Studies were excluded if they had only descriptive mapping of incidence using geographic
207 information systems or if the model was not fitted or validated using observation data. Simulation-
208 based and theoretical modelling studies were excluded unless their predictions of *Aedes*-borne disease
209 transmission risk (as opposed to model parameters) were validated using data from real-world
210 settings. Conference and workshop proceedings were excluded, as were review articles. This
211 systematic review is registered on PROSPERO (reference: CRD42022358144).

212 **Data extraction**

213 The following variables were extracted from eligible articles:

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

218 Analysis of the data and visualisations were carried out using R (version 4.2.2) [22]. The
219 complete list of all included studies and data extracted from each study are available in S1 File.

220 **Quality assessment**

221 A quality assessment tool was developed using the EPIFORGE checklist (S2 File), a
222 guideline for standardised reporting of epidemic forecasting and prediction research, to assess the
223 reporting quality of included studies [23]. This guideline assesses whether studies report on the
224 following domains: study goals, data sources, model characteristics and assumptions, model
225 evaluation, and study generalisability. The nine criteria were equally weighted, each with a score of 0

226 (poor) to 2 (good), for a maximum of 18 points. On the basis of the overall score, each paper was
227 rated ‘low’ (<10), ‘medium’ (10–12), ‘high’ (13–15) or ‘very high’ (>15).

228 **Results**

229 A total of 16,625 records were retrieved from the databases and 7,742 titles and abstracts
230 screened after removing duplicates (Fig 1). A total of 83 records were additionally identified through
231 bibliographic searches and contacts with experts. Of 301 records, a total of 118 studies were excluded
232 because the full-text was not available, they were published in other languages, or the topics were
233 irrelevant. One paper included two different models using different datasets so we counted it as two
234 separate studies [24]. As a result, we identified 183 studies published between 2002 and 2022 that
235 were ultimately included in the review (Fig 1).

236

237 **Fig 1. PRISMA flow chart.**

238

239 There has been a rapid increase in the number of arboviral spatial modelling studies over the
240 past 20 years, reflecting the growing public health priority of these diseases and increasing
241 accessibility of data and modelling methods. There was an average of 1.7 studies published per year
242 before 2008, 4.7 studies per year between 2008-2014 and 19.3 per year between 2015-2021 (Fig 2).
243 The distribution of risk mapping studies over geography and by disease closely follow the abundance
244 and availability of data. Using WHO Regions, a total of 40.8% (n = 78) of the studies were conducted
245 in the Americas, followed by 19.4% (n = 37) in SE Asia and 17.3% (n = 33) in the Western Pacific
246 region with a wide geographic diversity of studies over the past five years. Brazil (n = 35) was the
247 most frequently studied country, followed by Colombia (n = 15) and Indonesia (n = 13). The diversity
248 of regions studied has also increased: until 2014 studies tended to focus primarily on the Americas
249 and Western Pacific whereas since 2015 studies focusing on SE Asia and the global scale have been
250 increasingly prevalent (Fig 2). More than 70% (n = 131) of the studies modelled dengue transmission,
251 20 (10.9%) modelled Zika, 15 (8.2%) modelled yellow fever and seven (3.8%) chikungunya. There

252 were six (3.3%) studies that modelled the risk of dengue, Zika, and chikungunya together, while also
253 modelling the diseases individually; two modelled dengue and Zika together and two modelled Zika
254 and chikungunya together.

255

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

260 Purpose of maps

261 The main groups of purposes or goals of risk maps vary depending on the specific disease and
262 context, but can generally be grouped into four categories: 1) providing a broad overview of the spatial
263 distribution of risk over long-term averages and suggesting how it might change under different
264 scenarios of global changes in climate, economics, and demographics (e.g., [28,29]); 2) predicting the
265 spread of outbreaks and gaining a better understanding of major drivers of geographical spread (e.g,
266 [30,31]); 3) evaluating and planning vaccination programs by estimating disease burden and identifying
267 high-risk areas at the continental or country-level scale (e.g., [32,33]); and 4) informing planning and
268 outbreak response by increasing the precision of risk estimates and mapping sub-national risk using
269 surveillance data (e.g., [34,35]).

270 Data types

271 Most studies (n = 137, 74.9%) used case count data from routine passive surveillance to fit
272 models, most often aggregated to the administrative district (admin2)- or province (admin1)-level (Fig
273 3). Use of occurrence data was also widespread (n = 29, 15.8%), particularly for specific use cases,
274 such as the generation of global suitability maps. There were only seven studies (3.8%) that included
275 data from community-representative seroprevalence surveys, and seven studies that included data
276 from at least two different data types. The use of seroprevalence data was limited to dengue (n = 9)

277 and yellow fever (n = 4), both resulting from widespread seroprevalence surveys in preparation for, or
278 to evaluate, vaccination programmes. Generally the paucity of any one data type for yellow fever
279 meant a more equally distributed use of different data types in models and greater use of multiple
280 types of data [8,33,36,37] (Fig 3).

281

282 **Fig 3. Sources of epidemiological data used by diseases.** Each cell represents the number and
283 percentage of studies with the denominators summed vertically.

284

285 Risk maps have been generated across a wide range of spatial scales from global to sub-
286 national (Fig 4). We identified 22 studies that produced global risk maps of various *Aedes*-borne
287 arboviruses. Despite large gaps in data availability at the global scale, the majority (n = 18/23, 78.3%)
288 of these global maps make high resolution predictions at the pixel level, enabled by growing
289 availability of high resolution remotely-sensed climate datasets (Fig 5). For Zika, yellow fever, and
290 chikungunya, maps were primarily focussed at a continent or national scale with a resolution between
291 city-level and national-level (Figs 4 and 5), reflecting the more regional scope of their distribution
292 (yellow fever in Africa) or high profile epidemics (the 2015-2016 Zika epidemic in the Americas).
293 While maps are available at all spatial scales for dengue, the majority of models (n = 83, 63.4%) are
294 now at sub-national scale, usually at the resolution of city/district (admin-2) (Figs 4 and 5). This
295 reflects the increasing application of these techniques to routinely collected case incidence data to
296 provide country-specific recommendations about targeting of control resources within countries based
297 on the latest local data. There remain strong regional disparities in the scale and resolution of mapping
298 efforts with many high-resolution and country-specific maps in the Americas, while risk estimates for
299 Africa are fewer, of comparatively lower resolution, and are typically derived from global or
300 continent-level modelling efforts (S1 Fig).

301

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

304 **Fig 5. Spatial resolution by geographical scope.** Each cell represents the number and percentage of
 305 studies with the denominators summed horizontally.

306

307 Spatiotemporal prediction maps were often generated based on monthly or weekly intervals
 308 (S1 Table). The longest period of study was for 804 months (67 years), while the shortest period of
 309 study was for 3 months, with an average of 125 months (10 years) and a median of 60 months (5
 310 years). Studies tended to use data from periods with high numbers of reported cases, with dengue data
 311 concentrated in between 2010-2015, Zika data between 2015-2016. For chikungunya many studies
 312 use data from 2014 and for yellow fever the data used have been spread over time, with few studies
 313 using recent data from 2015-2020 (S2 Fig).

314 **Covariates**

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

318

319 **Table 1. List of covariates included in the studies.**

Covariates	Count	Percentage (%)
Climatic		
Temperature	98	53.6
Rainfall	79	43.2
Humidity	23	12.6
Bioclimatic variables	6	3.3
El Nino Southern Oscillation Index	4	2.2
Soil moisture (water stress/wetness)	4	2.2
Demographic factors		
Population density	44	24.0
Age	26	14.2
Air travel	19	10.4
Human daily mobility	13	7.1
Vaccination coverage	7	3.8
Sex	7	3.8
Socio-economic factors		
Gross domestic product	15	8.2
Household income	12	6.6
Education/literacy rate	6	3.3

Covariates	Count	Percentage (%)
Occupation and employment status	5	2.7
Socio-economic strata	6	3.3
Ecology		
Non-human primates species	6	3.3
Location of breeding sites	6	3.3
Breteau index	3	2.2
Adult mosquito abundance	2	1.1
Environmental factors		
Vegetation	27	14.8
Elevation/altitude	25	13.7
Urbanisation	22	12.0
Distance to roads, road density	14	7.7
Land use/land cover	13	7.1
Distance to water bodies/river	9	4.9
Spatiotemporal incidence		
Case count across time periods and neighbouring regions	23	12.6

320

321 Climatic variables were the most common group of covariates in models with temperature
322 and rainfall dominating. More than half of the studies (n = 98, 53.6%) included temperature as a
323 covariate while around 40% of studies had rainfall (n = 79, 43.2%). Temperature and rainfall were
324 better fit when lagged one or two months rather than unlagged [38–41]. Temperature and rainfall were
325 considered as significant factors in most studies, but some studies showed that meteorological factors
326 alone are not sufficient to explain spatial heterogeneity in disease transmission, which may be
327 associated more with non-climatic factors [42–44]. Rather than rely on raw measures of temperature,
328 24 studies (13.1%) instead used “temperature suitability” of *Aedes* mosquito vectors, which
329 incorporates a variety of different methods of modelling the temperature constraints on the vector and
330 virus dynamics that are most critical for virus transmission [45]. Six studies used bioclimatic variables
331 that encompassed annual temperature and precipitation ranges, seasonal fluctuations, as well as
332 extreme or constraining factors that capture broader biological patterns [29,46–50]. Four studies
333 additionally used indicators associated with El Niño Southern Oscillation as covariates [35,51–53].
334 Examples of other climatic variables that were included in the reviewed models were diurnal
335 temperature range [54–56], atmospheric pressure [57,58], wind speed [59,60], and duration of
336 sunshine [38,61,62].

337 Population density (n = 44, 24.0%) and age distributions (n = 26, 14.2%) were often
338 considered in modelling arboviruses. Many studies found population density to be a significant
339 covariate in their models, demonstrating a positive association with disease transmission, but some
340 studies reported a negative [63,64] or null association [39,44,65]. Human mobility between cities or
341 countries (n = 19, 10.4%) was also considered by including travel distance between regions [66,67] or
342 air travel passenger volume [68–73]. Some studies included daily human mobility data (n = 13, 7.1%),
343 mostly mapped at sub-national scale, with the aim of better representing short-distance high frequency
344 movements such as daily commuting [74,75]. Seven studies, for yellow fever and dengue, considered
345 vaccination coverage and measures of population immunity from infection in their models
346 [27,33,37,47,63,76,77].

347 The most common socio-economic variable was gross domestic product (GDP) (n = 15),
348 followed by household poverty/income level (n = 12, 6.6%) and education level (n = 6, 3.3%). A
349 socio-economic strata or a composite index such as human development index, social advantage and
350 disadvantage score (n = 6, 3.3%) were also included as socio-economic predictors in some of the
351 reviewed models. Lower neighbourhood socio-economic status was generally associated with
352 increased risk of *Aedes*-borne arbovirus diseases; in regions with established arboviral circulation,
353 community-level factors such as inadequate garbage collection, low income, and lack of access to
354 health care were associated with elevated risk of dengue infections [78–80].

355 For models fit at the sub-national scale to case incidence data, accompanying direct
356 measurements of the *Aedes* mosquito population improved model predictive performance. Breteau
357 index (BI), which is defined as the number of positive containers per 100 houses, was used as a
358 predictor in three studies [53,81,82]. Six studies included location of *Aedes* breeding sites in their
359 models [74,83–87]. The number of catches of female adult mosquitoes was included in two studies
360 [58,88]. In the absence of direct measurements of the vector abundance, modelled predictions of
361 “suitability for *Aedes* mosquitoes [89]” have been used, particularly at broad global scales and to
362 make early predictions for emerging Zika epidemics. Six studies included the occurrence or species
363 richness of non-human primates in modelling yellow fever.

364 The most common environmental variable was vegetation index (n = 27, 14.8%), followed by
365 altitude/elevation (n = 25, 13.7%) and urbanisation (n = 22, 12.0%). Some studies found that
366 vegetation was not a key predictor variable and had no association with dengue incidence [90,91],
367 whereas those considering vegetation in modelling yellow fever generally found that there was a
368 strong and significant vegetation-disease association possibly because of the greater role of the forest-
369 fringe environment in driving spillover from non-human primate reservoirs [64,92–94]. Road density
370 and proximity to the road were also included as a predictor in 14 studies (7.7%). More generic
371 categories of land use and land cover type have also been considered in another 13 studies.

372 Disease incidence across time periods and neighbouring regions were included as covariates
373 in 23 studies (12.6%) to explain contemporaneous disease transmission. Several studies included past
374 case counts lagged by one week to four months to improve temporal prediction accuracy [51,66,95–
375 98]. Source country's disease incidence rate was included in studies quantifying the risk of
376 importation from endemic to non-endemic settings such as Europe [70,99] and Asia-Pacific regions
377 [69].

378 For each paper, we also examined whether the collinearity among covariates was checked and
379 whether models retained covariates after conducting variable selection procedures. Among the 148
380 studies excluding those that used mechanistic models or only included random effect terms, only 33
381 studies (22.3%) tested different combinations of covariates and checked the multicollinearity among
382 them by calculating the correlation coefficient or variance inflation factor, or using principal
383 component analysis. There were 63 studies (42.6%) that did not include any process for selecting
384 variables or checking collinearity (S3 Table). However, it is worth noting that some of these studies
385 may have had a small number of covariates that were selected based on their known or cited
386 ecological or theoretical relevance to disease transmission, which may explain the lack of variable
387 selection process.

388 For the 33 studies that both checked the multicollinearity of covariates and performed
389 variable selection, we summarised the retention rate of different groups of covariates in the final
390 models (Fig 6A) [27,32,38–44,46–48,54,59,60,72,78,79,86,93,99–112]. Of 33 studies, 25 studies
391 (96.2%) retained climatic variables when tested. Only one study on dengue [111] tested all six

392 categories and rejected demographic, ecological data and spatiotemporal incidence; seven studies
393 tested all categories except for ecological variables. Apart from climatic variables, environmental
394 variables were the most commonly used, with 21 studies tested and only three of them rejected,
395 followed by demographic (23 tested and 6 rejected), socio-economic variables (16 tested but 5
396 rejected). Ecological data (7 tested and 2 rejected) and spatiotemporal incidence (5 tested and 1
397 rejected) were the least tested and included (Fig 6A). The most common combinations of retained
398 categories were climatic, environmental, demography, and socio-economic ($n = 4$) [41,44,103,108].
399 For climatic variables, different measures of temperature and rainfall were tested in reviewed studies.
400 Inclusion of temperature in models differed between studies, with minimum temperature often
401 selected over average and maximum temperature in six out of 14 studies (Fig 6B). We identified that
402 average rainfall was preferred over other measures of rainfall and humidity but only five studies
403 examined the performance of models in which both variables were considered (Fig 6B). We found
404 that 29 studies have included lagged covariates in their models. The length of the lag periods tested
405 for temperature, humidity and precipitation ranged from 0 to 16 weeks, with most being concentrated
406 between 4 to 12 weeks (S3 Fig). The average lag periods for mean temperature and precipitation tend
407 to be longer in the Americas compared to Western Pacific and SE Asia (S3 Fig).

408

409 **Fig 6. Covariates included and rejected. (a) Selected covariate categories; (b) climate variable**
410 **selections.** Mean temp: mean temperature; Min temp: minimum temperature; Max temp: maximum
411 temperature; DTR: Diurnal temperature range; Avg: average. The values in the bottom represent the
412 number and percentage of studies tested and included the corresponding category of covariates.

413 **Modelling framework**

414 Four classes of modelling methods were identified: statistical mixed effect models, statistical
415 fixed effect models, machine learning and mechanistic models (Table 2). Overall, the most common
416 modelling approaches were types of statistical mixed effect models ($n = 69$, 39.5%), with generalised
417 linear mixed models (GLMM) dominating, followed by generalised additive mixed models (GAMM)

418 (n = 4) and distributed lag non-linear models (DLNM) (n = 4). Mixed effect models were often
419 preferred when using areal-type case count data aggregated over distinct geographical areas (e.g.
420 administrative boundaries) (Fig 7).

421

422 **Fig 7. Modelling framework by input data type.**

423

424 Statistical fixed effect models were used in 21.2% of studies, with generalised linear models
425 (GLM) and geographically weighted regression (GWR) as the most used approaches. Since fixed
426 effect models assume that all observations are independent, models used spatial variables to account
427 for spatial relationships. For example, several studies included the coordinates (long, lat) of cases,
428 households, or the centroid of a region [32,36,77,113–116].

429 A variety of machine learning methods were employed in 26.1% of studies. The most
430 frequently used machine learning methods were MaxEnt and boosted regression tree (BRT). They
431 were often used when developing ecological niche or species distribution models using point-
432 referenced occurrence data to describe the environmental suitability of arbovirus transmission, and
433 especially for larger geographical scales (e.g., international scale). Of 23 studies that developed a
434 global risk map of different arbovirus transmission, ten studies adopted machine learning methods, six
435 of which used either MaxEnt or BRT [3,28,29,117–119]. Seven studies developed and compared the
436 performance of different machine learning methods. For example, Jiang et al. (2018) adapted three
437 different machine learning models, namely backward propagation neural network, gradient boosting
438 machine and random forest, and reported that backward propagation neural network showed the best
439 performance in predicting the global transmission risk of Zika [120]. Two studies generated ensemble
440 model projections of the spatiotemporal dynamics of Zika in Brazil and burden of yellow fever in
441 Africa [121].

442 Mechanistic models were used in 15.2% of studies, especially compartmental and
443 metapopulation models. Compartmental models e.g. human SEIR - mosquito SIR models were used
444 in six studies to explain the impact of different factors on the transmission dynamics, especially for
445 smaller scales e.g. country or sub-national scale [75,88,122–125]. Eight studies used metapopulation

446 or network models, all of which considered the connectivity between areas or regions by including the
447 patterns of daily human mobility or air travel data [31,67,70,71,73,82,83,126]. Five studies used
448 mechanistic mosquito models to produce estimates of temperature suitability, vectorial capacity or
449 basic reproductive number (R_0) at the continent or global scale [127–131].

450 Surprisingly, only 48.1% of studies ($n = 88$) included in this review assessed the predictive
451 performance using cross-validation procedures, such as K-fold cross-validation or random partitioning
452 of data, commonly referred to as "out-of-sample validation". It was more common to perform this
453 type of validation in studies using machine learning methods than in studies using other modelling
454 methods; only 25% of studies using fixed effect models performed out-of-sample validation (Fig 8).
455 Of these studies, only three studies included model validation on independent test data ("hold-out
456 validation") [55,132,133].

457

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

459

460 The most common model performance evaluation metrics were information criteria ($n = 82$,
461 29.8%), with Akaike information criteria (AIC) and the Bayesian information criterion (BIC) as the
462 most used metric (S2 Table). Confusion matrix-based metrics were used in 21.1% of studies, with the
463 Receiver Operating Characteristics (ROC) curve most frequently used. Correlation-based metrics
464 were used in 14.9% of studies, especially R-squared. 23 studies (8.4%) did not use any of the metrics
465 described above (S2 Table).

466 **Quality assessment**

467 Using the adapted tool for assessment of modelling study reporting, scores for the reviewed
468 paper ranged from 6 to 18 out of 18. Twelve studies were classified as low quality, 50 as medium
469 quality, 76 as high quality and 45 as very high quality. The median score was 13/18, which is
470 categorised as high quality. Discussions on the generalisability of the developed models were lacking

471 in many papers. Study objectives, settings, and data sources were often unclear in poorly scored
472 studies.

473 **Discussion**

474 This review provides a comprehensive overview of risk mapping studies, including their
475 covariates and modelling frameworks to investigate the transmission of arboviruses. We found that
476 the choices of data, covariates and modelling frameworks were largely determined by the purpose of
477 the map. We identified 23 studies that generated global risk maps, using machine learning-based
478 ecological niche modelling. These approaches are designed to give a broad overview of the spatial
479 distribution of risk over long-term averages and suggest how it might change under different scenarios
480 of global changes in climate, economics and demographics. Geolocation of disease occurrence data,
481 often combined with high- resolution environmental datasets, were more common for global risk
482 mapping because they encompass large areas and various environmental conditions and provide
483 information about the extent of transmission. However, caution is needed when utilising the outputs
484 of high resolution global risk maps, particularly for informing local decisions due to large data gaps
485 and biases [133] that are not reflected in their highly geographically precise predictions and
486 sometimes don't align with (typically later published) estimates from country-specific models that use
487 more local data.

488 We found that major epidemics, such as the 2015-2016 Zika epidemic, have acted as catalysts
489 for the development of new risk mapping methods applied in new contexts, possibly due to expanding
490 generation and sharing of data that has accompanied these more recent epidemics. The paucity of data
491 in the early stages of epidemics and similarities between arboviral diseases gives mechanistic
492 modelling approaches an advantage over more data-dependent statistical approaches despite the
493 latter's traditional dominance of the field of risk mapping [14]. As with any model, the predictions are
494 inherently a function of the data available and primary use cases at the time of analysis, and
495 contemporary approaches to mapping risk of diseases like Zika and chikungunya would likely differ
496 substantially from those conducted in the early stages of epidemics. We also show how epidemics

497 have accelerated the use of human movement data in arbovirus risk mapping, and that human
498 movement data is especially valuable to understand long-distance spread since *Aedes* mosquitoes have
499 a limited dispersal capability [134]. Daily commuting and air travel has improved predictions in both
500 statistical and mechanistic modelling approaches, particularly when mapping how the spatial
501 distribution of risk changes over the course of an epidemic.

502 Studies on modelling yellow fever employed multiple datasets and various approaches,
503 mostly motivated by the need to account for sparse, non-standardised data. They tend to be conducted
504 at continental or country-level scale in African and South American countries with high endemicity
505 for yellow fever transmission or recent outbreaks, for the purpose of evaluation and planning
506 vaccination programs. Inclusion of seroprevalence data and vaccination coverage therefore played a
507 significant role in robust estimation of disease burden and approaches used for yellow fever could be
508 increasingly important for mapping dengue risk as vaccines begin to be rolled out in various countries
509 [135].

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

525 Overall, we found that the quantity and variety of covariates included in arbovirus risk
526 mapping studies has increased in line with growing availability of these variables. While
527 developments over the past decade have focussed on global climate datasets, data on human
528 movement [136] and urban infrastructure [137] are becoming increasingly available and may play
529 important roles in future arbovirus risk mapping studies. Historically, limited data availability has
530 made it difficult to quantify human mobility patterns, requiring models that incorporate gravity or
531 radiation as an approximation [31,83,138]. However, the recent emergence of mobile phone data
532 enables real-time tracing of fine-scale movement across large numbers of individuals, although
533 privacy and bias issues remain [139]. The move towards large, open, accessible datasets for vector
534 borne diseases necessitates not just a more robust data science workforce, but a better motivation and
535 capacity planning for data fluency among primary data producers. While issues of human subjects and
536 data privacy must remain foremost in contemplating large-scale studies of vector borne disease risk,
537 nonetheless, leveraging entomological surveillance data, meteorological data, geospatial
538 representation of infrastructure and landscape (e.g., derived from remote sensing, well-resolved built
539 environment enumerations, high resolution travel network data), and climatological modelling output,
540 is less constrained by international regulations, so identifying the necessary investments and key
541 routes of engagement is a high-level first step to addressing the data gaps.

542 We found surprisingly few studies conducted robust variable selection procedures. In
543 addition, out-of-sample validation techniques were explicitly stated in only half of the studies
544 reviewed. Statistical and machine learning models, predominantly used in arbovirus risk mapping
545 studies, require a large amount of data and therefore both variable selection and cross-validation are
546 important steps to reduce overfitting and improve model interpretability and predictive accuracy.
547 Although the majority of studies used traditional cross-validation techniques, the use of spatial cross-
548 validation i.e., spatial block bootstrapping is increasingly popular due to its ability to account for
549 spatial dependence in the data [92,94]. This may help to better test the spatial predictive performance
550 of the model, particularly if there are large heterogeneities in data availability across the study sites,
551 which is common in many arbovirus mapping applications.

552 **Limitations**

553 One limitation of our systematic review is that it focussed on spatial modelling approaches.
554 The conclusions we reach, particularly with reference to drivers of transmission, may differ between
555 risk mapping and temporal prediction models which may be particularly important as the two fields
556 continue to overlap. We also only considered studies published in English, which may affect our
557 conclusions about regional patterns. Additionally, it is possible that some relevant literature,
558 particularly in the form of grey literature, may have been missed as the databases do not contain all
559 journals and university press articles. This is particularly true for locally-relevant geospatial modelling
560 work, which may not have been published in mainstream academic outlets. Finally, we excluded
561 studies that did not assess risk of human infection, excluding a number of studies dealing exclusively
562 with entomological risk or non-human host risk.

563 **Recommendations for future studies**

- 564 • Consider the strengths and weaknesses of different data types for different purposes as the
565 choice of data type imposes specific restrictions on the modelling framework and resolution
566 of the prediction. Historically the most common applications have been: occurrence data to
567 map the changing global limits of transmission, short-term aggregated level incidence data to
568 track the geographic spread of epidemics and high spatiotemporal resolution incidence data to
569 understand the roles of different drivers and forecast epidemics.
- 570 • Include covariates from multiple domains (climatic, environmental, demographic,
571 socioeconomic, ecological) and test whether their inclusion improves prediction.
- 572 • National or subnational studies should consider additional local covariates not available
573 across broader regions, such as data from the arbovirus control programmes, finer scale
574 meteorological resolution data, or infrastructural data from census databases.
- 575 • Even with extensive use of covariates, unobserved confounding will still be an issue,
576 particularly for broad scope (national and above) models, meaning that the use of structured

577 spatio-temporal random effects, ideally in a Bayesian mixed effects statistical modelling
578 framework, is preferable to more simplistic fixed effect statistical models.

- 579 • Use predictive validation metrics on held out datasets. Ideally using procedures that take into
580 account the unique challenges posed by highly spatially and temporally heterogeneous
581 datasets, such as multiple-fold blocked spatial and temporal cross validation.
- 582 • Arbovirus risk mapping is a rapidly developing field with continual improvements in data
583 quantity and representativeness, growing availability of potentially informative covariates and
584 new innovations to model fitting and evaluation. Future arbovirus risk mapping studies should
585 incorporate these new developments and not just rely on the status quo of existing studies.

586 **Conclusion**

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

594 **Acknowledgements**

595 This work was discussed with the Technical Advisory Group on arboviruses (TAG-
596 Arbovirus), the Secretariat of the Global Arbovirus Initiative (Raman Velayudhan, Laurence Cibrelus,
597 Jennifer Horton, Marie-Eve Raguenaud, Maria Van Kerkhove, Qingxia Zhong), and the participants
598 of the arbovirus risk mapping meeting held in Seattle in October 2022 as part of the ASTMH (Isabel
599 Rodriguez-Barraquer, Leo Bastos, Simon Cauchemez, Ilaria Dorigatti, Neil Ferguson, Simon Hay,

600 Wenbiao Hu, Axel Kroeger, Velma Lopez, A. Townsend Peterson, Maile Philips, David Pigott,
601 Krystina Rysava, Sophie von Dobschütz, and Anna Winters).

602 **References**

- 603 1. Jones R, Kulkarni MA, Davidson TMV, Team R-LR, Talbot B. Arbovirus vectors of
604 epidemiological concern in the Americas: A scoping review of entomological studies on Zika,
605 dengue and chikungunya virus vectors. PLOS ONE. 2020;15: e0220753.
606 doi:10.1371/journal.pone.0220753
- 607 2. Leta S, Beyene TJ, Clercq EMD, Amenu K, Kraemer MUG, Revie CW. Global risk mapping
608 for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*. Int J Infect Dis.
609 2018;67: 25–35. doi:10.1016/j.ijid.2017.11.026
- 610 3. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global
611 distribution and burden of dengue. Nature. 2013;496: 504–507. doi:10.1038/nature12060
- 612 4. Paixão ES, Teixeira MG, Rodrigues LC. Zika, chikungunya and dengue: the causes and threats
613 of new and re-emerging arboviral diseases. BMJ Glob Health. 2018;3: e000530.
614 doi:10.1136/bmjgh-2017-000530
- 615 5. Cattarino L, Rodriguez-Barraquer I, Imai N, Cummings DAT, Ferguson NM. Mapping global
616 variation in dengue transmission intensity. Sci Transl Med. 2020;12: eaax4144.
617 doi:10.1126/scitranslmed.aax4144
- 618 6. Puntasecca CJ, King CH, LaBeaud AD. Measuring the global burden of chikungunya and Zika
619 viruses: A systematic review. PLoS Negl Trop Dis. 2021;15: e0009055.
620 doi:10.1371/journal.pntd.0009055
- 621 7. Li SL, Messina JP, Pybus OG, Kraemer MUG, Gardner L. A review of models applied to the
622 geographic spread of Zika virus. Trans R Soc Trop Med Hyg. 2021;115: 956–964.
623 doi:10.1093/trstmh/trab009
- 624 8. Gaythorpe KA, Hamlet A, Jean K, Garkauskas Ramos D, Cibrelus L, Garske T, et al. The
625 global burden of yellow fever. Davenport MP, Flegg J, Flegg J, Perkins AT, editors. eLife.

- 626 2021;10: e64670. doi:10.7554/eLife.64670
- 627 9. Bassey BE, Braka F, Onyibe R, Kolude OO, Oluwadare M, Oluwabukola A, et al. Changing
628 epidemiology of yellow fever virus in Oyo State, Nigeria. BMC Public Health. 2022;22: 467.
629 doi:10.1186/s12889-022-12871-0
- 630 10. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of
631 human dengue virus occurrence. Sci Data. 2014;1: 140004. doi:10.1038/sdata.2014.4
- 632 11. Anderson RM, Ferguson NM, Donnelly CA, Anderson RM. Transmission dynamics and
633 epidemiology of dengue: insights from age-stratified sero-prevalence surveys. Philos Trans R
634 Soc Lond B Biol Sci. 1999;354: 757–768. doi:10.1098/rstb.1999.0428
- 635 12. Liu-Helmersson J, Brännström Å, Sewe MO, Semenza JC, Rocklöv J. Estimating Past,
636 Present, and Future Trends in the Global Distribution and Abundance of the Arbovirus Vector
637 *Aedes aegypti* Under Climate Change Scenarios. Front Public Health. 2019;7: 148.
638 doi:10.3389/fpubh.2019.00148
- 639 13. Kolimenakis A, Heinz S, Wilson ML, Winkler V, Yakob L, Michaelakis A, et al. The role of
640 urbanisation in the spread of *Aedes* mosquitoes and the diseases they transmit—A systematic
641 review. PLoS Negl Trop Dis. 2021;15: e0009631. doi:10.1371/journal.pntd.0009631
- 642 14. Lee SA, Jarvis CI, Edmunds WJ, Economou T, Lowe R. Spatial connectivity in mosquito-
643 borne disease models: a systematic review of methods and assumptions. J R Soc Interface.
644 2021;18: 20210096. doi:10.1098/rsif.2021.0096
- 645 15. Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S. Climate change and
646 dengue: a critical and systematic review of quantitative modelling approaches. BMC Infect
647 Dis. 2014;14: 167. doi:10.1186/1471-2334-14-167
- 648 16. Sylvestre E, Joachim C, Cécilia-Joseph E, Bouzillé G, Campillo-Gimenez B, Cuggia M, et al.
649 Data-driven methods for dengue prediction and surveillance using real-world and Big Data: A
650 systematic review. Santos VS, editor. PLoS Negl Trop Dis. 2022;16: e0010056.
651 doi:10.1371/journal.pntd.0010056
- 652 17. Lima CL de, da Silva ACG, Moreno GMM, Cordeiro da Silva C, Musah A, Aldosery A, et al.
653 Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic

- 654 Review. *Front Public Health*. 2022;10: 900077. doi:10.3389/fpubh.2022.900077
- 655 18. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, et al. Modeling
656 tools for dengue risk mapping - a systematic review. *Int J Health Geogr*. 2014;13: 50.
657 doi:10.1186/1476-072X-13-50
- 658 19. Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal
659 approaches to modelling dengue fever: a systematic review. *Epidemiol Infect*. 2019;147: e33.
660 doi:10.1017/S0950268818002807
- 661 20. Yin S, Ren C, Shi Y, Hua J, Yuan H-Y, Tian L-W. A Systematic Review on Modeling
662 Methods and Influential Factors for Mapping Dengue-Related Risk in Urban Settings. *Int J*
663 *Environ Res Public Health*. 2022;19: 15265. doi:10.3390/ijerph192215265
- 664 21. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for
665 Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Med*. 2009;6:
666 e1000097. doi:10.1371/journal.pmed.1000097
- 667 22. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R
668 Foundation for Statistical Computing; 2022. Available: <https://www.R-project.org/>
- 669 23. Pollett S, Johansson MA, Reich NG, Brett-Major D, Del Valle SY, Venkatramanan S, et al.
670 Recommended reporting items for epidemic forecasting and prediction research: The
671 EPIFORGE 2020 guidelines. *PLOS Med*. 2021;18: e1003793.
672 doi:10.1371/journal.pmed.1003793
- 673 24. Rogers DJ, Wilson AJ, Hay SI, Graham AJ. The Global Distribution of Yellow Fever and
674 Dengue. *Adv Parasitol*. 2006;62: 181–220. doi:10.1016/S0065-308X(05)62006-4
- 675 25. Bettis AA, Jackson ML, Yoon I-K, Breugelmans JG, Goios A, Gubler DJ, et al. The global
676 epidemiology of chikungunya from 1999 to 2020: A systematic literature review to inform the
677 development and introduction of vaccines. *PLoS Negl Trop Dis*. 2022;16: e0010069.
678 doi:10.1371/journal.pntd.0010069
- 679 26. Giancchetti E, Cianchi V, Torelli A, Montomoli E. Yellow Fever: Origin, Epidemiology,
680 Preventive Strategies and Future Prospects. *Vaccines*. 2022;10: 372.
681 doi:10.3390/vaccines10030372

- 682 27. Chen Y, Li N, Lourenço J, Wang L, Cazelles B, Dong L, et al. Measuring the effects of
683 COVID-19-related disruption on dengue transmission in southeast Asia and Latin America: a
684 statistical modelling study. *Lancet Infect Dis.* 2022;22: 657–667. doi:10.1016/S1473-
685 3099(22)00025-1
- 686 28. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global
687 environmental suitability for Zika virus. *Jit M*, editor. *eLife.* 2016;5: e15272.
688 doi:10.7554/eLife.15272
- 689 29. Tjaden NB, Suk JE, Fischer D, Thomas SM, Beierkuhnlein C, Semenza JC. Modelling the
690 effects of global climate change on Chikungunya transmission in the 21st century. *Sci Rep.*
691 2017;7: 3813. doi:10.1038/s41598-017-03566-3
- 692 30. Johansson MA, Powers AM, Pesik N, Cohen NJ, Staples JE. Nowcasting the Spread of
693 Chikungunya Virus in the Americas. *PLOS ONE.* 2014;9: e104915.
694 doi:10.1371/journal.pone.0104915
- 695 31. Chadsuthi S, Althouse BM, Iamsirithaworn S, Triampo W, Grantz KH, Cummings DAT.
696 Travel distance and human movement predict paths of emergence and spatial spread of
697 chikungunya in Thailand. *Epidemiol Infect.* 2018;146: 1654–1662.
698 doi:10.1017/S0950268818001917
- 699 32. Garske T, Kerkhove MDV, Yactayo S, Ronveaux O, Lewis RF, Staples JE, et al. Yellow Fever
700 in Africa: Estimating the Burden of Disease and Impact of Mass Vaccination from Outbreak
701 and Serological Data. *PLOS Med.* 2014;11: e1001638. doi:10.1371/journal.pmed.1001638
- 702 33. Perkins TA, Huber JH, Tran QM, Oidtman RJ, Walters MK, Siraj AS, et al. Burden is in the
703 eye of the beholder: Sensitivity of yellow fever disease burden estimates to modeling
704 assumptions. *Sci Adv.* 2021;7: eabg5033. doi:10.1126/sciadv.abg5033
- 705 34. Lauer SA, Sakrejda K, Ray EL, Keegan LT, Bi Q, Suangtho P, et al. Prospective forecasts of
706 annual dengue hemorrhagic fever incidence in Thailand, 2010–2014. *Proc Natl Acad Sci.*
707 2018;115: E2175–E2182. doi:10.1073/pnas.1714457115
- 708 35. Yip S, Che Him N, Jamil NI, He D, Sahu SK. Spatio-temporal detection for dengue outbreaks
709 in the Central Region of Malaysia using climatic drivers at mesoscale and synoptic scale. *Clim*

- 710 Risk Manag. 2022;36: 100429. doi:10.1016/j.crm.2022.100429
- 711 36. Gaythorpe KAM, Jean K, Cibrelus L, Garske T. Quantifying model evidence for yellow fever
712 transmission routes in Africa. PLOS Comput Biol. 2019;15: e1007355.
713 doi:10.1371/journal.pcbi.1007355
- 714 37. Jean K, Hamlet A, Benzler J, Cibrelus L, Gaythorpe KAM, Sall A, et al. Eliminating yellow
715 fever epidemics in Africa: Vaccine demand forecast and impact modelling. PLoS Negl Trop
716 Dis. 2020;14: e0008304. doi:10.1371/journal.pntd.0008304
- 717 38. Bett B, Grace D, Lee HS, Lindahl J, Nguyen-Viet H, Phuc P-D, et al. Spatiotemporal analysis
718 of historical records (2001–2012) on dengue fever in Vietnam and development of a statistical
719 model for forecasting risk. PLOS ONE. 2019;14: e0224353.
720 doi:10.1371/journal.pone.0224353
- 721 39. Tsheten T, Clements ACA, Gray DJ, Wangchuk S, Wangdi K. Spatial and temporal patterns of
722 dengue incidence in Bhutan: a Bayesian analysis. Emerg Microbes Infect. 2020;9: 1360–1371.
723 doi:10.1080/22221751.2020.1775497
- 724 40. Husnina Z, Clements ACA, Wangdi K. Forest cover and climate as potential drivers for
725 dengue fever in Sumatra and Kalimantan 2006-2016: a spatiotemporal analysis. Trop Med Int
726 Health TM IH. 2019;24: 888–898. doi:10.1111/tmi.13248
- 727 41. Desjardins MR, Eastin MD, Paul R, Casas I, Delmelle EM. Space–Time Conditional
728 Autoregressive Modeling to Estimate Neighborhood-Level Risks for Dengue Fever in Cali,
729 Colombia. Am J Trop Med Hyg. 2020;103: 2040–2053. doi:10.4269/ajtmh.20-0080
- 730 42. Chien L-C, Yu H-L. Impact of meteorological factors on the spatiotemporal patterns of dengue
731 fever incidence. Environ Int. 2014;73: 46–56. doi:10.1016/j.envint.2014.06.018
- 732 43. Chien L-C, Sy F, Pérez A. Identifying high risk areas of Zika virus infection by meteorological
733 factors in Colombia. BMC Infect Dis. 2019;19: 888. doi:10.1186/s12879-019-4499-9
- 734 44. Akter R, Hu W, Gatton M, Bambrick H, Cheng J, Tong S. Climate variability, socio-ecological
735 factors and dengue transmission in tropical Queensland, Australia: A Bayesian spatial analysis.
736 Environ Res. 2021;195: 110285. doi:10.1016/j.envres.2020.110285
- 737 45. Brady OJ, Golding N, Pigott DM, Kraemer MUG, Messina JP, Reiner Jr RC, et al. Global

- 738 temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for
739 dengue virus transmission. *Parasit Vectors*. 2014;7: 338. doi:10.1186/1756-3305-7-338
- 740 46. Acharya BK, Cao C, Xu M, Khanal L, Naeem S, Pandit S. Present and Future of Dengue Fever
741 in Nepal: Mapping Climatic Suitability by Ecological Niche Model. *Int J Environ Res Public*
742 *Health*. 2018;15: 187. doi:10.3390/ijerph15020187
- 743 47. de Thoisy B, Silva NIO, Sacchetto L, Trindade G de S, Drumond BP. Spatial epidemiology of
744 yellow fever: Identification of determinants of the 2016-2018 epidemics and at-risk areas in
745 Brazil. *PLoS Negl Trop Dis*. 2020;14: e0008691. doi:10.1371/journal.pntd.0008691
- 746 48. Jácome G, Vilela P, Yoo C. Present and future incidence of dengue fever in Ecuador
747 nationwide and coast region scale using species distribution modeling for climate variability's
748 effect. *Ecol Model*. 2019;400: 60–72. doi:10.1016/j.ecolmodel.2019.03.014
- 749 49. Wu W, Ren H, Lu L. Increasingly expanded future risk of dengue fever in the Pearl River
750 Delta, China. *PLoS Negl Trop Dis*. 2021;15: e0009745. doi:10.1371/journal.pntd.0009745
- 751 50. Flórez-Lozano K, Navarro-Lechuga E, Llinás-Solano H, Tuesca-Molina R, Sisa-Camargo A,
752 Mercado-Reyes M, et al. Spatial distribution of the relative risk of Zika virus disease in
753 Colombia during the 2015–2016 epidemic from a Bayesian approach. *Int J Gynecol Obstet*.
754 2020;148: 55–60. doi:10.1002/ijgo.13048
- 755 51. Lowe R, Bailey TC, Stephenson DB, Jupp TE, Graham RJ, Barcellos C, et al. The
756 development of an early warning system for climate-sensitive disease risk with a focus on
757 dengue epidemics in Southeast Brazil. *Stat Med*. 2013;32: 864–883. doi:10.1002/sim.5549
- 758 52. Lowe R, Bailey TC, Stephenson DB, Graham RJ, Coelho CAS, Sá Carvalho M, et al. Spatio-
759 temporal modelling of climate-sensitive disease risk: Towards an early warning system for
760 dengue in Brazil. *Comput Geosci*. 2011;37: 371–381. doi:10.1016/j.cageo.2010.01.008
- 761 53. Yu H-L, Yang S-J, Yen H-J, Christakos G. A spatio-temporal climate-based model of early
762 dengue fever warning in southern Taiwan. *Stoch Environ Res Risk Assess*. 2011;25: 485–494.
763 doi:10.1007/s00477-010-0417-9
- 764 54. Gaythorpe KA, Hamlet A, Jean K, Garkauskas Ramos D, Cibrelus L, Garske T, et al. The
765 global burden of yellow fever. Davenport MP, Flegg J, Flegg J, Perkins AT, editors. *eLife*.

- 766 2021;10: e64670. doi:10.7554/eLife.64670
- 767 55. Ferdousi T, Cohnstaedt LW, Scoglio CM. A Windowed Correlation-Based Feature Selection
768 Method to Improve Time Series Prediction of Dengue Fever Cases. *IEEE Access*. 2021;9:
769 141210–141222. doi:10.1109/ACCESS.2021.3120309
- 770 56. Sharmin S, Glass K, Viennet E, Harley D. Geostatistical mapping of the seasonal spread of
771 under-reported dengue cases in Bangladesh. *PLoS Negl Trop Dis*. 2018;12: e0006947.
772 doi:10.1371/journal.pntd.0006947
- 773 57. Mussumeci E, Codeço Coelho F. Large-scale multivariate forecasting models for Dengue -
774 LSTM versus random forest regression. *Spat Spatio-Temporal Epidemiol*. 2020;35: 100372.
775 doi:10.1016/j.sste.2020.100372
- 776 58. Sedda L, Taylor BM, Eiras AE, Marques JT, Dillon RJ. Using the intrinsic growth rate of the
777 mosquito population improves spatio-temporal dengue risk estimation. *Acta Trop*. 2020;208:
778 105519. doi:10.1016/j.actatropica.2020.105519
- 779 59. Stephenson C, Coker E, Wisely S, Liang S, Dinglasan RR, Lednicky JA. Imported Dengue
780 Case Numbers and Local Climatic Patterns Are Associated with Dengue Virus Transmission in
781 Florida, USA. *Insects*. 2022;13: 163. doi:10.3390/insects13020163
- 782 60. Siddiq A, Shukla N, Pradhan B. Spatio-temporal modelling of dengue fever cases in Saudi
783 Arabia using socio-economic, climatic and environmental factors. *Geocarto Int*. 2022; 1–25.
784 doi:10.1080/10106049.2022.2072005
- 785 61. Jaya IGNM, Folmer H. Spatiotemporal high-resolution prediction and mapping: methodology
786 and application to dengue disease. *J Geogr Syst*. 2022 [cited 3 Aug 2022]. doi:10.1007/s10109-
787 021-00368-0
- 788 62. Jaya IGNM, Folmer H. Identifying Spatiotemporal Clusters by Means of Agglomerative
789 Hierarchical Clustering and Bayesian Regression Analysis with Spatiotemporally Varying
790 Coefficients: Methodology and Application to Dengue Disease in Bandung, Indonesia. *Geogr*
791 *Anal*. 2021;53: 767–817. doi:10.1111/gean.12264
- 792 63. Shearer FM, Longbottom J, Browne AJ, Pigott DM, Brady OJ, Kraemer MUG, et al. Existing
793 and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet*

- 794 Glob Health. 2018;6: e270–e278. doi:10.1016/S2214-109X(18)30024-X
- 795 64. Servadio JL, Muñoz-Zanzi C, Convertino M. Environmental determinants predicting
796 population vulnerability to high yellow fever incidence. R Soc Open Sci. 9: 220086.
797 doi:10.1098/rsos.220086
- 798 65. McHale TC, Romero-Vivas CM, Fronterre C, Arango-Padilla P, Waterlow NR, Nix CD, et al.
799 Spatiotemporal Heterogeneity in the Distribution of Chikungunya and Zika Virus Case
800 Incidences during their 2014 to 2016 Epidemics in Barranquilla, Colombia. Int J Environ Res
801 Public Health. 2019;16: 1759. doi:10.3390/ijerph16101759
- 802 66. Lim JT, Dickens BSL, Mao Y, Kwak CW, Ng LC, Cook AR. Explicit characterization of
803 human population connectivity reveals long run persistence of interregional dengue shocks. J R
804 Soc Interface. 2020;17. doi:10.1098/rsif.2020.0340
- 805 67. Yu H-L, Angulo JM, Cheng M-H, Wu J, Christakos G. An online spatiotemporal prediction
806 model for dengue fever epidemic in Kaohsiung (Taiwan). Biom J. 2014;56: 428–440.
807 doi:10.1002/bimj.201200270
- 808 68. Gardner LM, Bóta A, Gangavarapu K, Kraemer MUG, Grubaugh ND. Inferring the risk
809 factors behind the geographical spread and transmission of Zika in the Americas. PLoS Negl
810 Trop Dis. 2018;12: e0006194. doi:10.1371/journal.pntd.0006194
- 811 69. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Model-informed risk assessment for Zika
812 virus outbreaks in the Asia-Pacific regions. J Infect. 2017;74: 484–491.
813 doi:10.1016/j.jinf.2017.01.015
- 814 70. Gardner LM, Fajardo D, Waller ST, Wang O, Sarkar S. A Predictive Spatial Model to
815 Quantify the Risk of Air-Travel-Associated Dengue Importation into the United States and
816 Europe. J Trop Med. 2012;2012: 1–11. doi:10.1155/2012/103679
- 817 71. Liebig J, Jansen C, Paini D, Gardner L, Jurdak R. A global model for predicting the arrival of
818 imported dengue infections. PLOS ONE. 2019;14: e0225193.
819 doi:10.1371/journal.pone.0225193
- 820 72. Nah K, Mizumoto K, Miyamatsu Y, Yasuda Y, Kinoshita R, Nishiura H. Estimating risks of
821 importation and local transmission of Zika virus infection. PeerJ. 2016;4: e1904.

- 822 doi:10.7717/peerj.1904
- 823 73. Zhang Q, Sun K, Chinazzi M, Pastore y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika
824 virus in the Americas. *Proc Natl Acad Sci.* 2017;114: E4334–E4343.
825 doi:10.1073/pnas.1620161114
- 826 74. Ong J, Liu X, Rajarethinam J, Kok SY, Liang S, Tang CS, et al. Mapping dengue risk in
827 Singapore using Random Forest. *PLoS Negl Trop Dis.* 2018;12: e0006587.
828 doi:10.1371/journal.pntd.0006587
- 829 75. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact
830 of human mobility on the emergence of dengue epidemics in Pakistan. *Proc Natl Acad Sci.*
831 2015;112: 11887–11892. doi:10.1073/pnas.1504964112
- 832 76. Childs ML, Nova N, Colvin J, Mordecai EA. Mosquito and primate ecology predict human
833 risk of yellow fever virus spillover in Brazil. *Philos Trans R Soc B Biol Sci.* 2019;374:
834 20180335. doi:10.1098/rstb.2018.0335
- 835 77. Honório NA, Nogueira RMR, Codeço CT, Carvalho MS, Cruz OG, Magalhães M de AFM, et
836 al. Spatial Evaluation and Modeling of Dengue Seroprevalence and Vector Density in Rio de
837 Janeiro, Brazil. *PLoS Negl Trop Dis.* 2009;3: e545. doi:10.1371/journal.pntd.0000545
- 838 78. Kikuti M, Cunha GM, Paploski IA, Kasper AM, Silva MM, Tavares AS, et al. Spatial
839 distribution of dengue in a Brazilian urban slum setting: role of socioeconomic gradient in
840 disease risk. *PLoS Negl Trop Dis.* 2015/07/22 ed. 2015;9: e0003937.
841 doi:10.1371/journal.pntd.0003937
- 842 79. Honorato T, Lapa PP de A, Sales CMM, Reis-Santos B, Tristão-Sá R, Bertolde AI, et al.
843 Spatial analysis of distribution of dengue cases in Espírito Santo, Brazil, in 2010: use of
844 Bayesian model. *Rev Bras Epidemiol.* 2014;17: 150–159. doi:10.1590/1809-
845 4503201400060013
- 846 80. Costa JV, Donalisio MR, Silveira LV de A. Spatial distribution of dengue incidence and socio-
847 environmental conditions in Campinas, São Paulo State, Brazil, 2007. *Cad Saúde Pública.*
848 2013;29: 1522–1532. doi:10.1590/0102-311X00110912
- 849 81. Lin C-H, Wen T-H. Using Geographically Weighted Regression (GWR) to Explore Spatial

- 850 Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of
851 Dengue. *Int J Environ Res Public Health*. 2011;8: 2798–2815. doi:10.3390/ijerph8072798
- 852 82. Zhu G, Liu T, Xiao J, Zhang B, Song T, Zhang Y, et al. Effects of human mobility,
853 temperature and mosquito control on the spatiotemporal transmission of dengue. *Sci Total*
854 *Environ*. 2019;651: 969–978. doi:10.1016/j.scitotenv.2018.09.182
- 855 83. Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, et al.
856 Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med*. 2015;13: 102.
857 doi:10.1186/s12916-015-0348-x
- 858 84. Cordeiro R, Donalisio MR, Andrade VR, Mafra AC, Nucci LB, Brown JC, et al. Spatial
859 distribution of the risk of dengue fever in southeast Brazil, 2006-2007. *BMC Public Health*.
860 2011;11: 355. doi:10.1186/1471-2458-11-355
- 861 85. Abdul Rahm S, Rahim A, Mallongi A. Forecasting of Dengue Disease Incident Risks Using
862 Non-stationary Spatial of Geostatistics Model in Bone Regency Indonesia. *J Entomol*.
863 2016;14: 49–57. doi:10.3923/je.2017.49.57
- 864 86. Yin MS, Bicout DJ, Haddawy P, Schöning J, Laosiritaworn Y, Sa-angchai P. Added-value of
865 mosquito vector breeding sites from street view images in the risk mapping of dengue
866 incidence in Thailand. *PLoS Negl Trop Dis*. 2021;15: e0009122.
867 doi:10.1371/journal.pntd.0009122
- 868 87. Withanage GP, Gunawardana M, Viswakula SD, Samaraweera K, Gunawardana NS,
869 Hapugoda MD. Multivariate spatio-temporal approach to identify vulnerable localities in
870 dengue risk areas using Geographic Information System (GIS). *Sci Rep*. 2021;11: 4080.
871 doi:10.1038/s41598-021-83204-1
- 872 88. Li R, Xu L, Bjørnstad ON, Liu K, Song T, Chen A, et al. Climate-driven variation in mosquito
873 density predicts the spatiotemporal dynamics of dengue. *Proc Natl Acad Sci*. 2019;116: 3624–
874 3629. doi:10.1073/pnas.1806094116
- 875 89. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global
876 distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Jit M*, editor. *eLife*.
877 2015;4: e08347. doi:10.7554/eLife.08347

- 878 90. Rogers DJ, Suk JE, Semenza JC. Using global maps to predict the risk of dengue in Europe.
879 Acta Trop. 2014;129: 1–14. doi:10.1016/j.actatropica.2013.08.008
- 880 91. Ye J, Moreno-Madriñán MJ. Comparing different spatio-temporal modeling methods in
881 dengue fever data analysis in Colombia during 2012–2015. Spat Spatio-Temporal Epidemiol.
882 2020;34: 100360. doi:10.1016/j.sste.2020.100360
- 883 92. Hamlet A, Ramos DG, Gaythorpe KAM, Romano APM, Garske T, Ferguson NM. Seasonality
884 of agricultural exposure as an important predictor of seasonal yellow fever spillover in Brazil.
885 Nat Commun. 2021;12: 3647. doi:10.1038/s41467-021-23926-y
- 886 93. Kaul RB, Evans MV, Murdock CC, Drake JM. Spatio-temporal spillover risk of yellow fever
887 in Brazil. Parasit Vectors. 2018;11: 488. doi:10.1186/s13071-018-3063-6
- 888 94. Hamlet A, Gaythorpe KAM, Garske T, Ferguson NM. Seasonal and inter-annual drivers of
889 yellow fever transmission in South America. PLoS Negl Trop Dis. 2021;15: e0008974.
890 doi:10.1371/journal.pntd.0008974
- 891 95. Lowe R, Barcellos C, Coelho CAS, Bailey TC, Coelho GE, Graham R, et al. Dengue outlook
892 for the World Cup in Brazil: an early warning model framework driven by real-time seasonal
893 climate forecasts. Lancet Infect Dis. 2014;14: 619–626. doi:10.1016/S1473-3099(14)70781-9
- 894 96. Liu K, Zhang M, Xi G, Deng A, Song T, Li Q, et al. Enhancing fine-grained intra-urban
895 dengue forecasting by integrating spatial interactions of human movements between urban
896 regions. PLoS Negl Trop Dis. 2020;14: e0008924. doi:10.1371/journal.pntd.0008924
- 897 97. Liu K, Yin L, Zhang M, Kang M, Deng A-P, Li Q-L, et al. Facilitating fine-grained intra-urban
898 dengue forecasting by integrating urban environments measured from street-view images.
899 Infect Dis Poverty. 2021;10: 40. doi:10.1186/s40249-021-00824-5
- 900 98. Lowe R, Coelho CA, Barcellos C, Carvalho MS, Catão RDC, Coelho GE, et al. Evaluating
901 probabilistic dengue risk forecasts from a prototype early warning system for Brazil. eLife. 5:
902 e11285. doi:10.7554/eLife.11285
- 903 99. Salami D, Sousa CA, Martins M do RO, Capinha C. Predicting dengue importation into
904 Europe, using machine learning and model-agnostic methods. Sci Rep. 2020;10: 9689.
905 doi:10.1038/s41598-020-66650-1

- 906 100. Acharya BK, Cao C, Lakes T, Chen W, Naeem S, Pandit S. Modeling the spatially varying risk
907 factors of dengue fever in Jhapa district, Nepal, using the semi-parametric geographically
908 weighted regression model. *Int J Biometeorol.* 2018;62: 1973–1986. doi:10.1007/s00484-018-
909 1601-8
- 910 101. Agarwal N, Koti SR, Saran S, Kumar AS. Data mining techniques for predicting dengue
911 outbreak in geospatial domain using weather parameters for New Delhi, India. *Curr Sci.*
912 2018;114: 2281–2291.
- 913 102. Delmelle E, Hagenlocher M, Kienberger S, Casas I. A spatial model of socioeconomic and
914 environmental determinants of dengue fever in Cali, Colombia. *Acta Trop.* 2016;164: 169–
915 176. doi:10.1016/j.actatropica.2016.08.028
- 916 103. Li Q, Ren H, Zheng L, Cao W, Zhang A, Zhuang D, et al. Ecological Niche Modeling
917 Identifies Fine-Scale Areas at High Risk of Dengue Fever in the Pearl River Delta, China. *Int J*
918 *Environ Res Public Health.* 2017;14: 619. doi:10.3390/ijerph14060619
- 919 104. Machado-Machado EA. Empirical mapping of suitability to dengue fever in Mexico using
920 species distribution modeling. *Appl Geogr.* 2012;33: 82–93. doi:10.1016/j.apgeog.2011.06.011
- 921 105. Ren H, Wu W, Li T, Yang Z. Urban villages as transfer stations for dengue fever epidemic: A
922 case study in the Guangzhou, China. *PLoS Negl Trop Dis.* 2019;13: e0007350.
923 doi:10.1371/journal.pntd.0007350
- 924 106. Restrepo AC, Baker P, Clements ACA. National spatial and temporal patterns of notified
925 dengue cases, Colombia 2007–2010. *Trop Med Int Health.* 2014;19: 863–871.
926 doi:10.1111/tmi.12325
- 927 107. Salami D, Capinha C, Martins M do RO, Sousa CA. Dengue importation into Europe: A
928 network connectivity-based approach. *PLOS ONE.* 2020;15: e0230274.
929 doi:10.1371/journal.pone.0230274
- 930 108. Sánchez-Hernández D, Aguirre-Salado CA, Sánchez-Díaz G, Aguirre-Salado AI, Soubervielle-
931 Montalvo C, Reyes-Cárdenas O, et al. Modeling spatial pattern of dengue in North Central
932 Mexico using survey data and logistic regression. *Int J Environ Health Res.* 2021;31: 872–888.
933 doi:10.1080/09603123.2019.1700938

- 934 109. Teurlai M, Menkès CE, Cavarero V, Degallier N, Descloux E, Grangeon J-P, et al. Socio-
935 economic and Climate Factors Associated with Dengue Fever Spatial Heterogeneity: A
936 Worked Example in New Caledonia. *PLoS Negl Trop Dis.* 2015;9: e0004211.
937 doi:10.1371/journal.pntd.0004211
- 938 110. Wijayanti SPM, Porphyre T, Chase-Topping M, Rainey SM, McFarlane M, Schnettler E, et al.
939 The Importance of Socio-Economic Versus Environmental Risk Factors for Reported Dengue
940 Cases in Java, Indonesia. *PLoS Negl Trop Dis.* 2016;10: e0004964.
941 doi:10.1371/journal.pntd.0004964
- 942 111. Wu P-C, Lay J-G, Guo H-R, Lin C-Y, Lung S-C, Su H-J. Higher temperature and urbanization
943 affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Sci Total*
944 *Environ.* 2009;407: 2224–2233. doi:10.1016/j.scitotenv.2008.11.034
- 945 112. Yu H-L, Lee C-H, Chien L-C. A spatiotemporal dengue fever early warning model accounting
946 for nonlinear associations with hydrological factors: a Bayesian maximum entropy approach.
947 *Stoch Environ Res Risk Assess.* 2016;30: 2127–2141. doi:10.1007/s00477-016-1328-1
- 948 113. Siqueira-Junior JB, Maciel IJ, Barcellos C, Souza WV, Carvalho MS, Nascimento NE, et al.
949 Spatial point analysis based on dengue surveys at household level in central Brazil. *BMC*
950 *Public Health.* 2008;8: 361. doi:10.1186/1471-2458-8-361
- 951 114. Lee SA, Economou T, de Castro Catão R, Barcellos C, Lowe R. The impact of climate
952 suitability, urbanisation, and connectivity on the expansion of dengue in 21st century Brazil.
953 *PLoS Negl Trop Dis.* 2021;15: e0009773. doi:10.1371/journal.pntd.0009773
- 954 115. Mutucumarana CP, Bodinayake CK, Nagahawatte A, Devasiri V, Kurukulasooriya R,
955 Anuradha T, et al. Geospatial analysis of dengue emergence in rural areas in the Southern
956 Province of Sri Lanka. *Trans R Soc Trop Med Hyg.* 2020;114: 408–414.
957 doi:10.1093/trstmh/trz123
- 958 116. Chien L-C, Sy F, Pérez A. Identifying high risk areas of Zika virus infection by meteorological
959 factors in Colombia. *BMC Infect Dis.* 2019;19: 888. doi:10.1186/s12879-019-4499-9
- 960 117. Samy AM, Thomas SM, Wahed AAE, Cohoon KP, Peterson AT. Mapping the global
961 geographic potential of Zika virus spread. *Mem Inst Oswaldo Cruz.* 2016;111: 559.

- 962 doi:10.1590/0074-02760160149
- 963 118. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and
964 future global distribution and population at risk of dengue. *Nat Microbiol.* 2019;4: 1508–1515.
965 doi:10.1038/s41564-019-0476-8
- 966 119. Nsoesie EO, Kraemer MU, Golding N, Pigott DM, Brady OJ, Moyes CL, et al. Global
967 distribution and environmental suitability for chikungunya virus, 1952 to 2015.
968 *Eurosurveillance.* 2016;21: 30234. doi:10.2807/1560-7917.ES.2016.21.20.30234
- 969 120. Jiang D, Hao M, Ding F, Fu J, Li M. Mapping the transmission risk of Zika virus using
970 machine learning models. *Acta Trop.* 2018;185: 391–399.
971 doi:10.1016/j.actatropica.2018.06.021
- 972 121. Soliman M, Lyubchich V, Gel YR. Ensemble forecasting of the Zika space-time spread with
973 topological data analysis. *Environmetrics.* 2020;31: e2629. doi:10.1002/env.2629
- 974 122. Benkimoun S, Atyame C, Haramboure M, Degenne P, Thébault H, Dehecq J-S, et al. Dynamic
975 mapping of dengue basic reproduction number. *Results Phys.* 2021;29: 104687.
976 doi:10.1016/j.rinp.2021.104687
- 977 123. Mukhsar, Ahmar A, A M, El-Khawaga H, Sayed M. Bayesian Convolution for Stochastic
978 Epidemic Model. *Intell Autom Soft Comput.* 2022;34: 1175–1186.
979 doi:10.32604/iasc.2022.025214
- 980 124. Romeo-Aznar V, Picinini Freitas L, Gonçalves Cruz O, King AA, Pascual M. Fine-scale
981 heterogeneity in population density predicts wave dynamics in dengue epidemics. *Nat*
982 *Commun.* 2022;13: 996. doi:10.1038/s41467-022-28231-w
- 983 125. Butterworth MK, Morin CW, Comrie AC. An analysis of the potential impact of climate
984 change on dengue transmission in the Southeastern United States. *Env Health Perspect.*
985 2016/10/08 ed. 2017;125: 579–585. doi:10.1289/EHP218
- 986 126. O'Reilly KM, Lowe R, Edmunds WJ, Mayaud P, Kucharski A, Eggo RM, et al. Projecting the
987 end of the Zika virus epidemic in Latin America: a modelling analysis. *BMC Med.* 2018;16:
988 180. doi:10.1186/s12916-018-1158-8
- 989 127. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, et al. Detecting the

- 990 impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic
991 models. *PLoS Negl Trop Dis*. 2017;11: e0005568. doi:10.1371/journal.pntd.0005568
- 992 128. Ryan SJ, Carlson CJ, Tesla B, Bonds MH, Ngonghala CN, Mordecai EA, et al. Warming
993 temperatures could expose more than 1.3 billion new people to Zika virus risk by 2050. *Glob
994 Change Biol*. 2021;27: 84–93. doi:10.1111/gcb.15384
- 995 129. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of
996 *Aedes*-borne virus transmission risk with climate change. *PLoS Negl Trop Dis*. 2019;13:
997 e0007213. doi:10.1371/journal.pntd.0007213
- 998 130. Caminade C, Turner J, Metelmann S, Hesson JC, Blagrove MSC, Solomon T, et al. Global risk
999 model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proc Natl
1000 Acad Sci U S A*. 2017;114: 119–124. doi:10.1073/pnas.1614303114
- 1001 131. Tesla B, Demakovskiy LR, Mordecai EA, Ryan SJ, Bonds MH, Ngonghala CN, et al.
1002 Temperature drives Zika virus transmission: evidence from empirical and mathematical
1003 models. *Proc R Soc B Biol Sci*. 2018;285: 20180795. doi:10.1098/rspb.2018.0795
- 1004 132. Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk
1005 of Zika in real time. *BMC Med*. 2019;17: 171. doi:10.1186/s12916-019-1389-3
- 1006 133. O'Reilly KM, Hendrickx E, Kharisma DD, Wilastonegoro NN, Carrington LB, Elyazar IRF, et
1007 al. Estimating the burden of dengue and the impact of release of wMel *Wolbachia*-infected
1008 mosquitoes in Indonesia: a modelling study. *BMC Med*. 2019;17: 172. doi:10.1186/s12916-
1009 019-1396-4
- 1010 134. Verdonschot PFM, Besse-Lototskaya AA. Flight distance of mosquitoes (*Culicidae*): A
1011 metadata analysis to support the management of barrier zones around rewetted and newly
1012 constructed wetlands. *Limnologica*. 2014;45: 69–79. doi:10.1016/j.limno.2013.11.002
- 1013 135. Zorlu G, Fleck F. Dengue vaccine roll-out: getting ahead of the game. *Bull World Health
1014 Organ*. 2011;89: 476–477. doi:10.2471/BLT.11.030711
- 1015 136. Kraemer MUG, Sadilek A, Zhang Q, Marchal NA, Tuli G, Cohn EL, et al. Mapping global
1016 variation in human mobility. *Nat Hum Behav*. 2020;4: 800–810. doi:10.1038/s41562-020-
1017 0875-0

- 1018 137. Tusting LS, Bisanzio D, Alabaster G, Cameron E, Cibulskis R, Davies M, et al. Mapping
1019 changes in housing in sub-Saharan Africa from 2000 to 2015. *Nature*. 2019;568: 391–394.
1020 doi:10.1038/s41586-019-1050-5
- 1021 138. Kraemer MUG, Faria NR, Reiner RC, Golding N, Nikolay B, Stasse S, et al. Spread of yellow
1022 fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a
1023 modelling study. *Lancet Infect Dis*. 2017;17: 330–338. doi:10.1016/S1473-3099(16)30513-8
- 1024 139. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJE. Connecting Mobility to Infectious
1025 Diseases: The Promise and Limits of Mobile Phone Data. *J Infect Dis*. 2016;214: S414–S420.
1026 doi:10.1093/infdis/jiw273
- 1027

1028 **Supplementary information**

1029 **S1 Fig. Spatial scale (a) and resolution (b) by study region.** Each cell represents the number and
1030 percentage of studies with the denominators summed vertically.

1031 **S2 Fig. Time span of data used by disease.**

1032 **S3 Fig. Summary of lagged covariates used. (a) lagged week per covariate; (b) average lag**
1033 **period of climatic covariates by region.** The numbers represent the mean (standard deviation) of the
1034 lag period in weeks.

1035 **S1 Table. Temporal resolution of predictions in reviewed studies.**

1036 **S2 Table. Modelling methods used in arbovirus risk mapping.**

1037 **S3 Table. Number of studies that used robust variable selection procedures.**

1038 **S1 File. Data extracted from the studies reviewed.**

1039 **S2 File. A modified EPIFORGE checklist.**

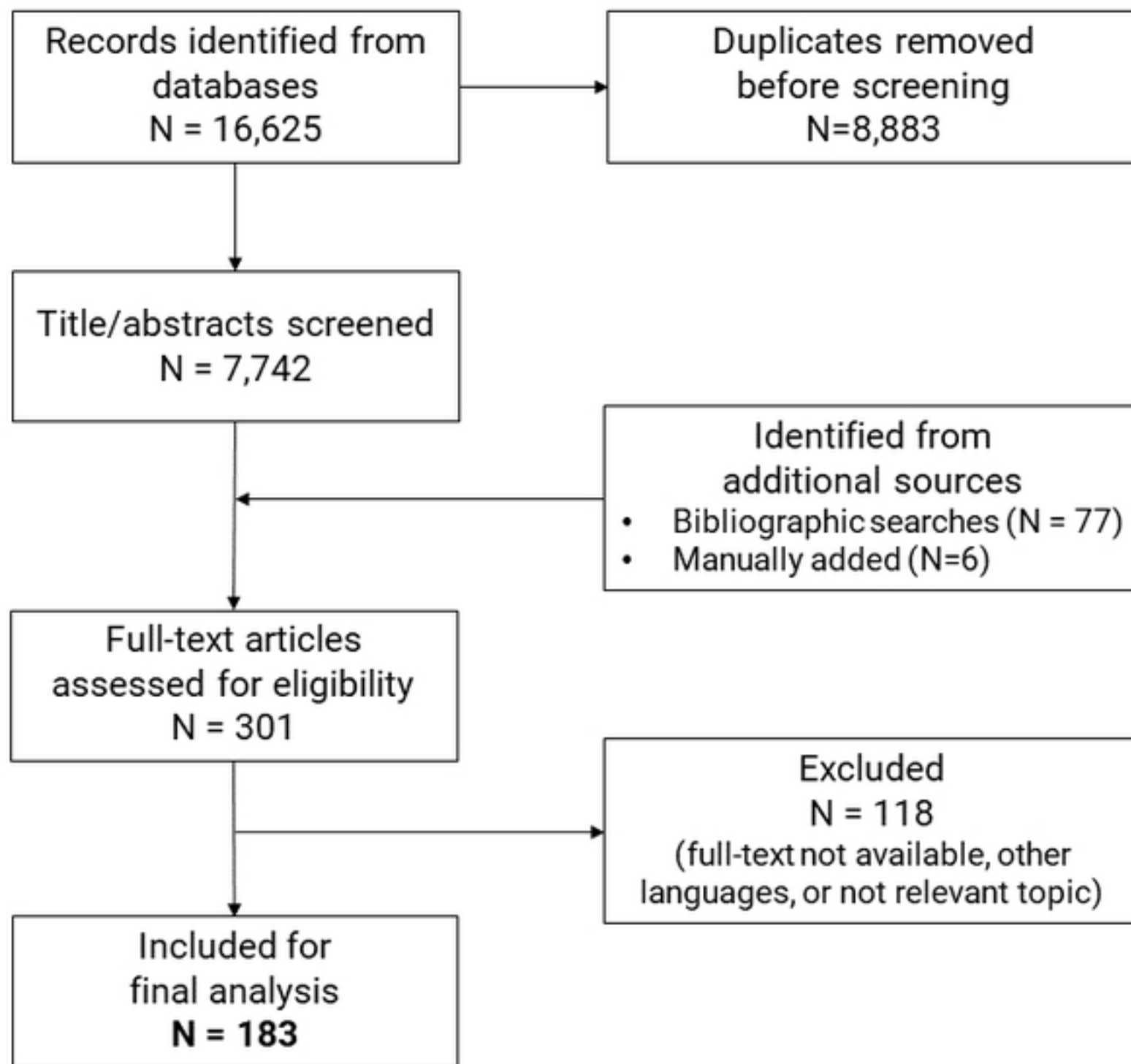


Figure 1

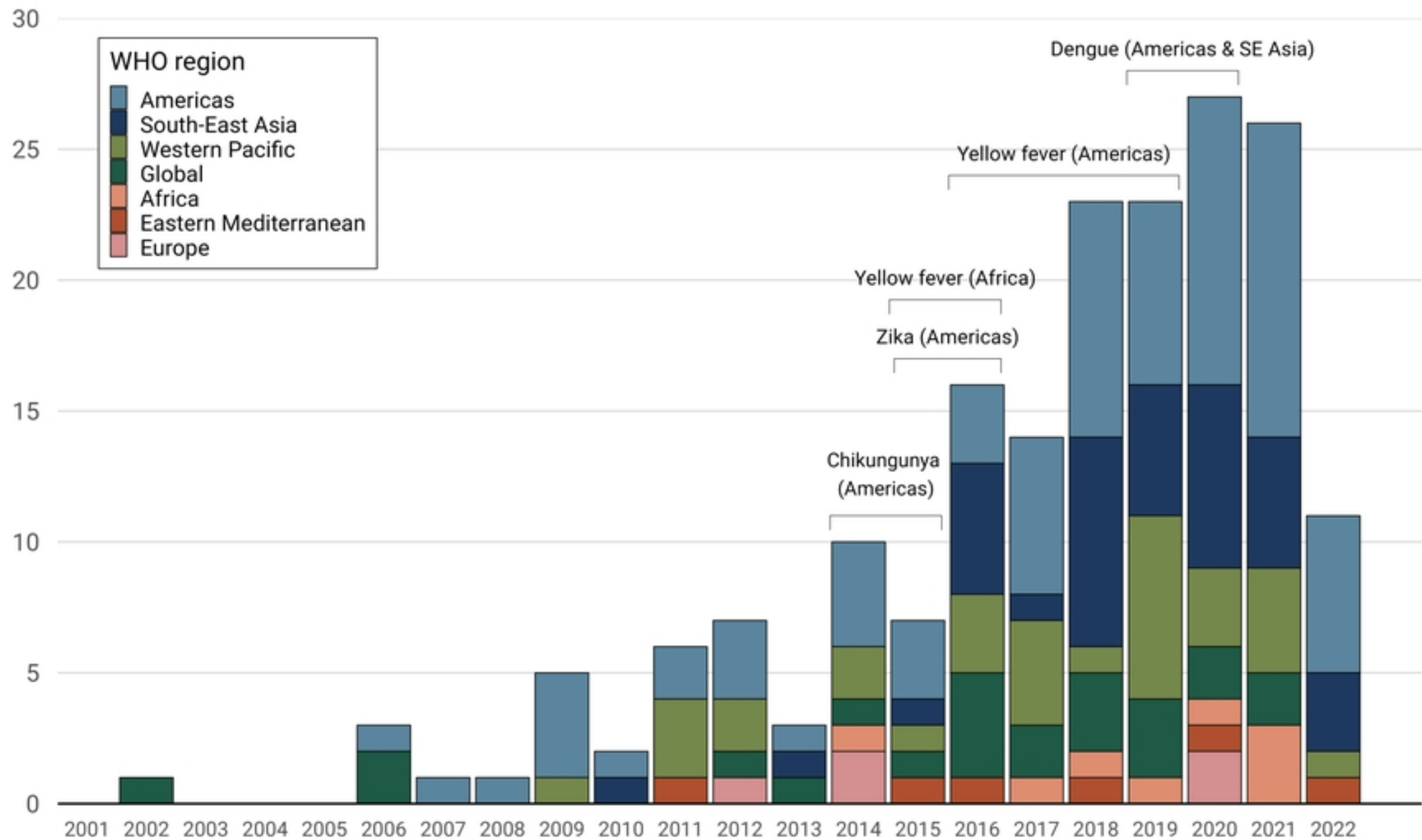


Figure2

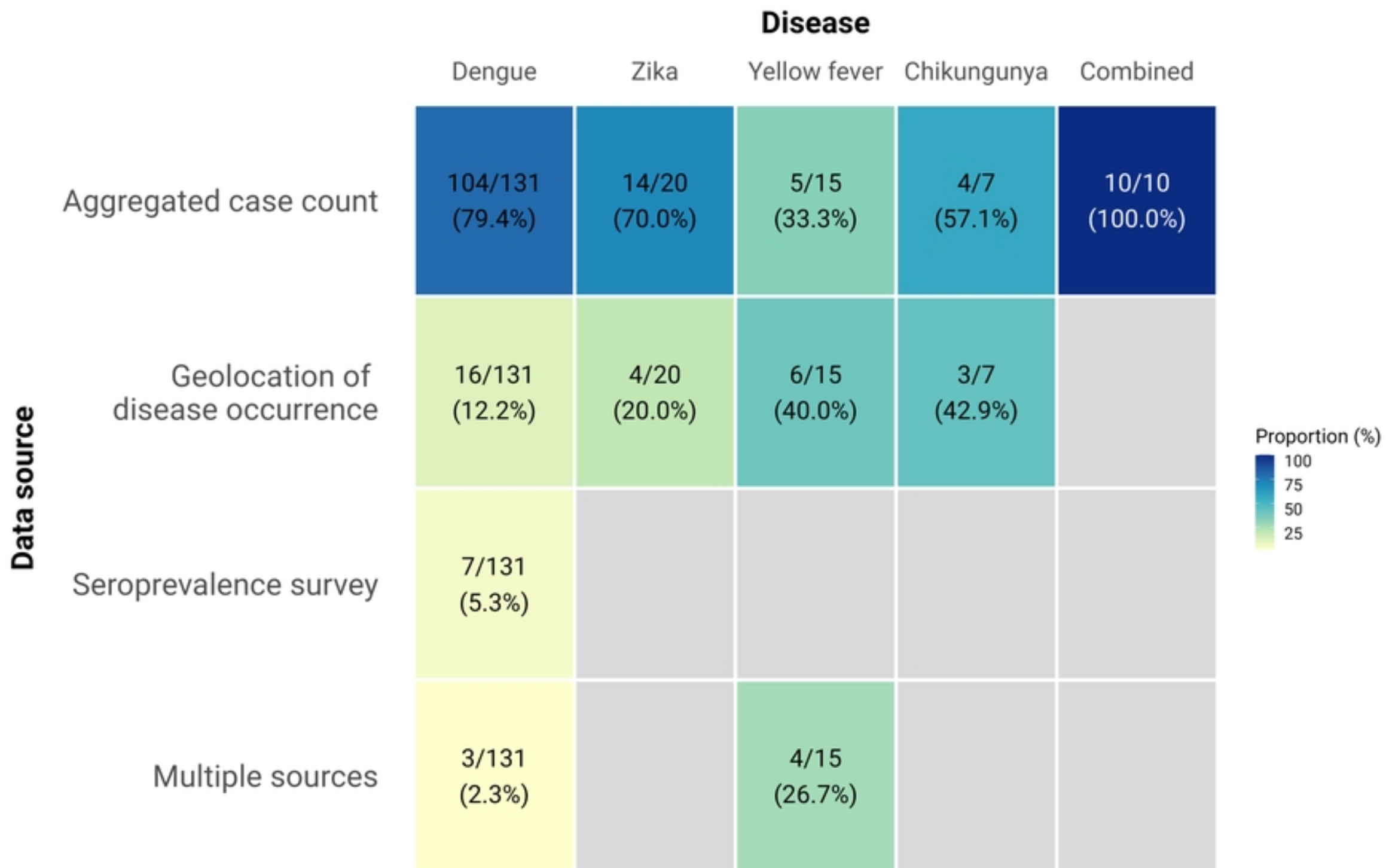


Figure3

Spatial scope

Disease

Dengue Zika Yellow fever Chikungunya Combined

Sub-national

83/131
(63.4%)

2/20
(10.0%)

0/15
(0.0%)

1/7
(14.3%)

7/10
(70.0%)

National

31/131
(23.7%)

4/20
(20.0%)

6/15
(40.0%)

1/7
(14.3%)

1/10
(10.0%)

Continental

8/131
(6.1%)

6/20
(30.0%)

7/15
(46.7%)

2/7
(28.6%)

1/10
(10.0%)

Global

9/131
(6.9%)

8/20
(40.0%)

2/15
(13.3%)

3/7
(42.9%)

1/10
(10.0%)

Proportion (%)

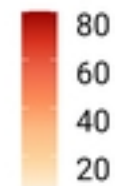


Figure4

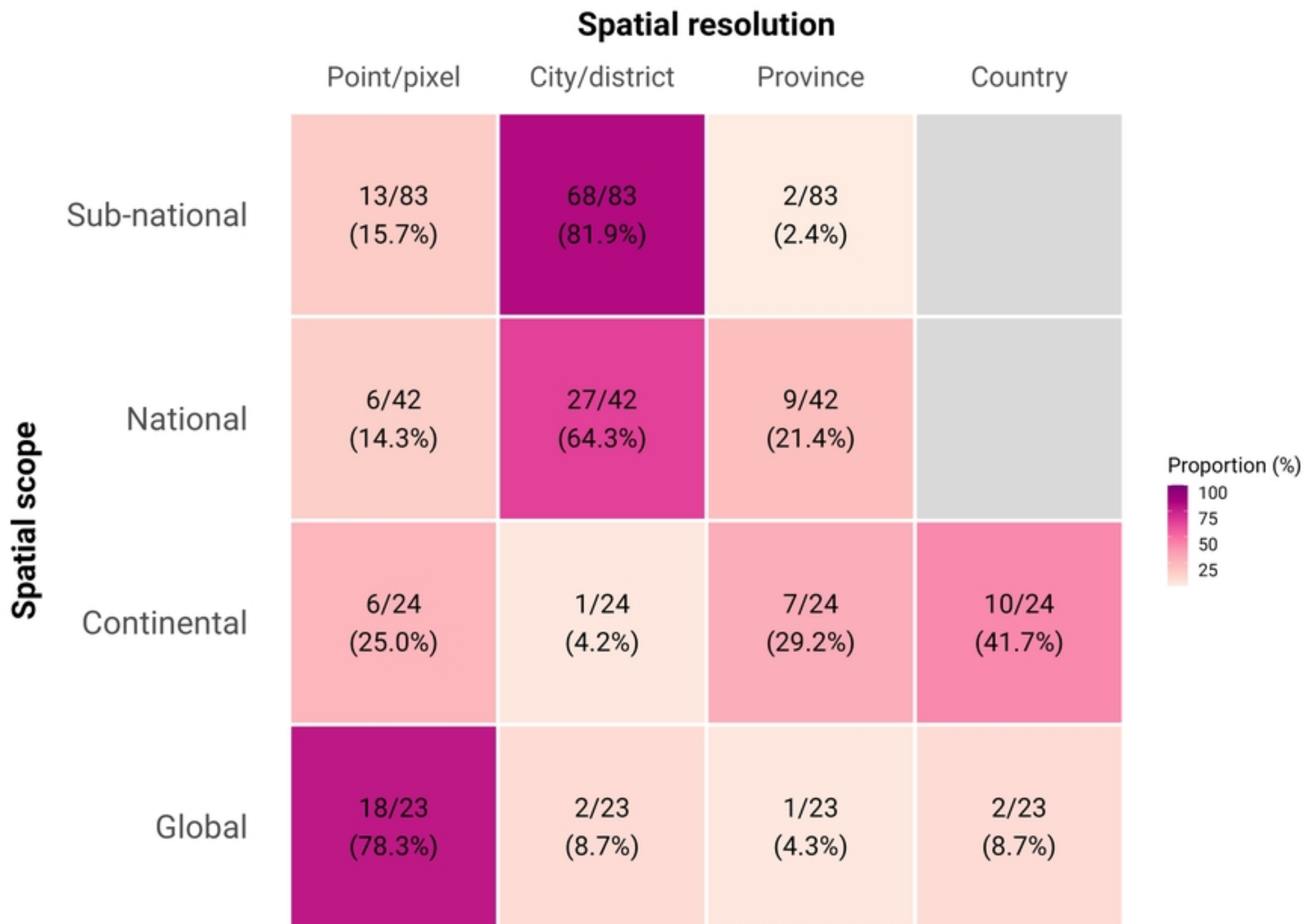
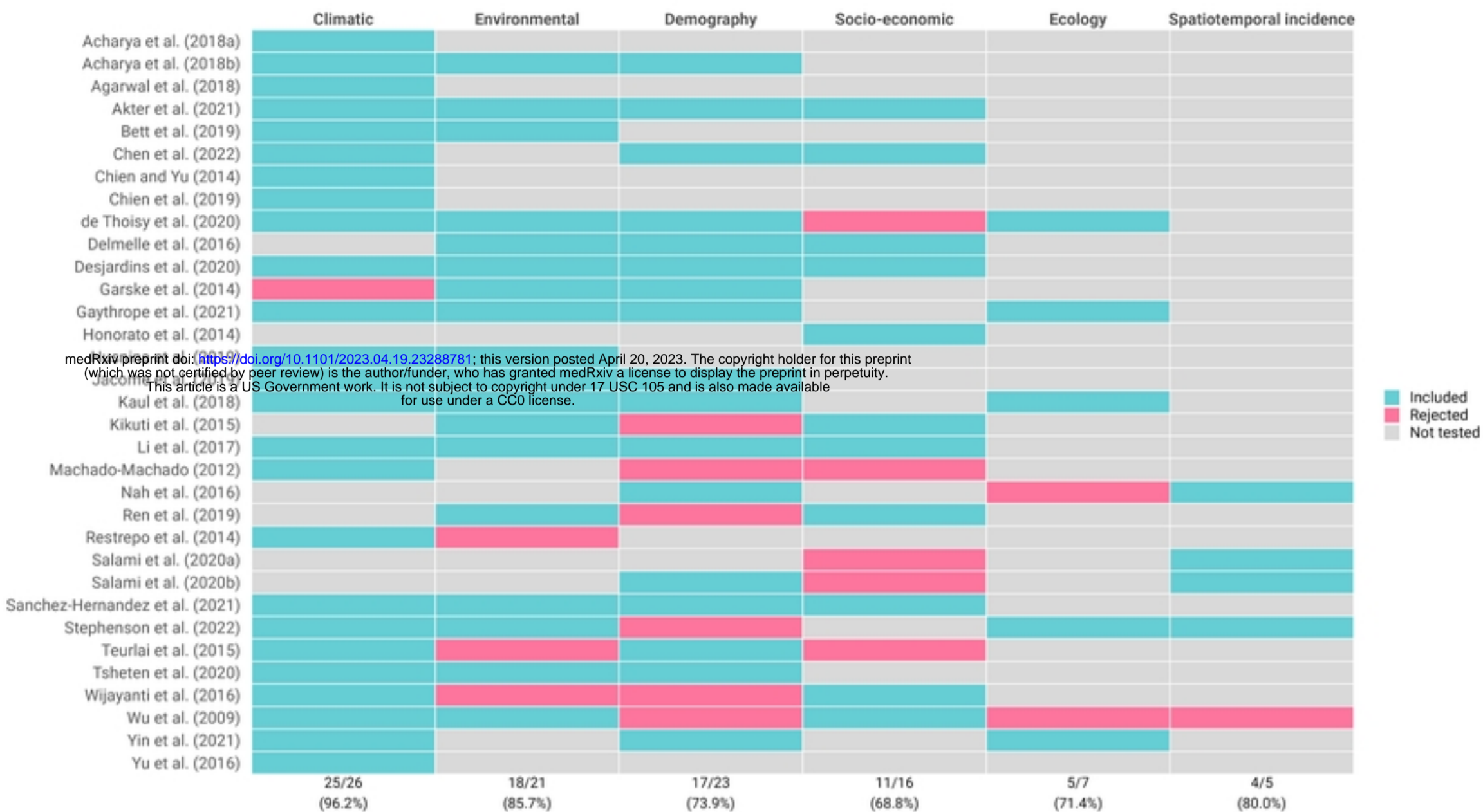


Figure5

A Selected covariate categories



medRxiv preprint doi: <https://doi.org/10.1101/2023.04.19.23288781>; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

B Climate variable selections



Figure 6

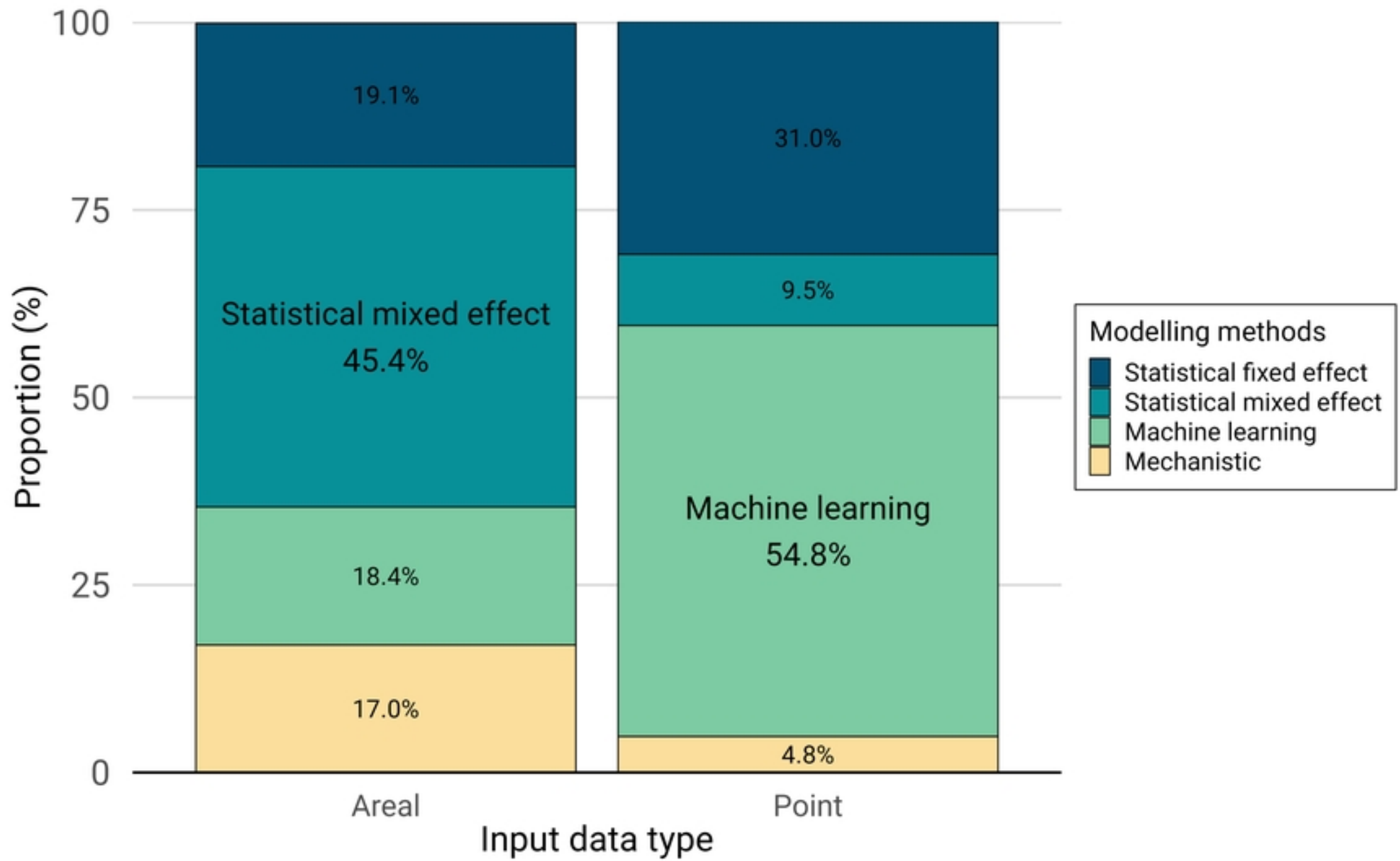


Figure7

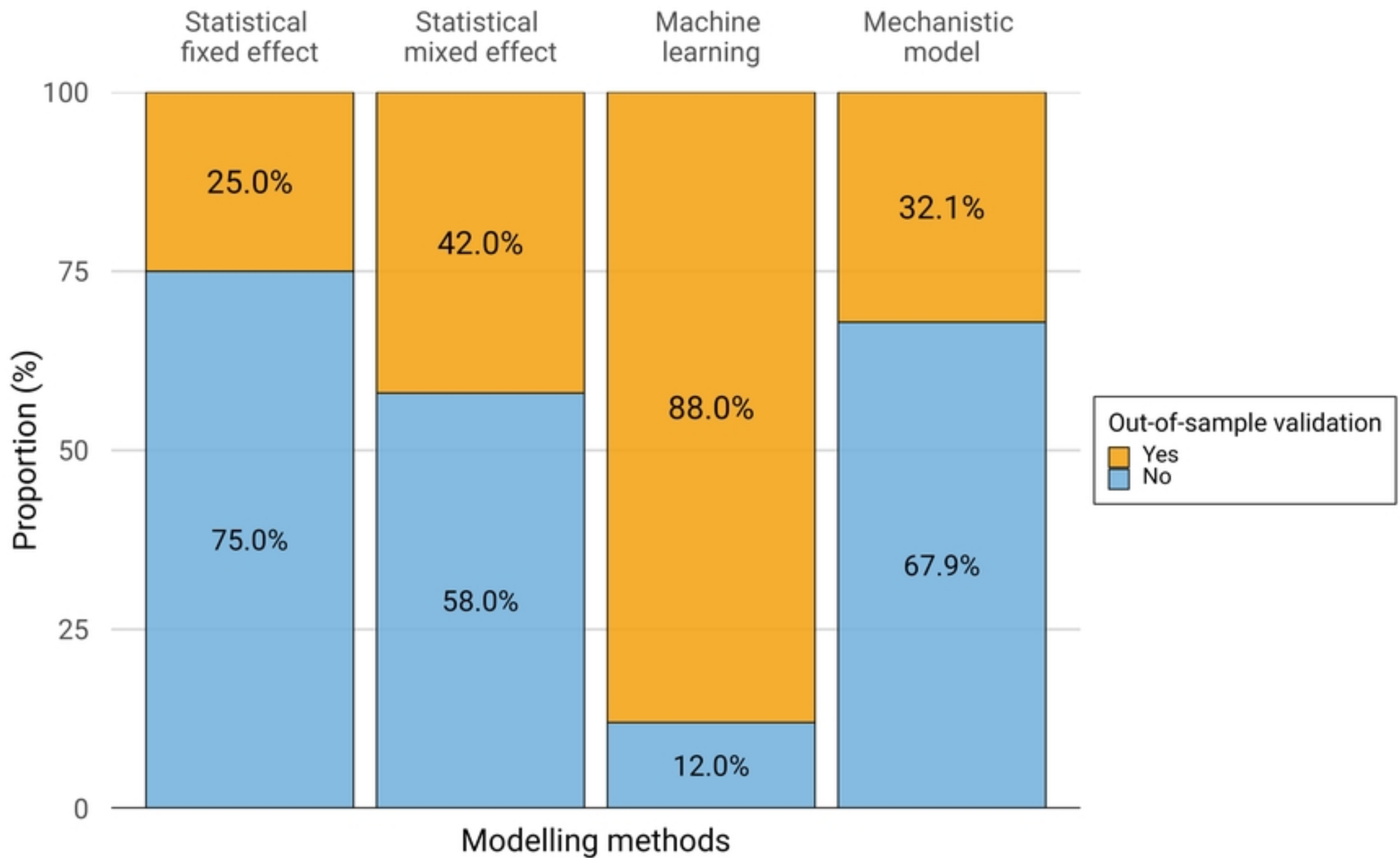


Figure8