Influence of Socio-Economic, Demographic and Climate Factors on the Regional Distribution of Dengue in the United States and Mexico

Matthew J Watts\textsuperscript{1+*}, Panagiota Kotsila\textsuperscript{1}, P Graham Mortyn\textsuperscript{1}, Victor Sarto i Monteys\textsuperscript{1} and Cesira Urzi Brancati\textsuperscript{2}

\textsuperscript{*}Correspondence: matthew.watts@uab.cat
\textsuperscript{1}Institute of Environmental Science and Technology (ICTA), Autonomous University of Barcelona (UAB), Bellaterra, Spain, Barcelona, Spain
\textsuperscript{2}Full list of author information is available at the end of the article
\textsuperscript{+}Equal contributor

Abstract

\textbf{Background:} Dengue is one of the important vector-borne diseases in the world today; it infects tens of millions of people each year and has been on the rise since the 1950s. In this study, we develop a set of indicators that help us examine the impact of socio-economic and demographic factors on the occurrence of dengue in regions of the United States and Mexico.

\textbf{Methods:} We assess the relationship between dengue occurrence in humans, climate factors (temperature and minimum quarterly rainfall), socio-economic factors (such as household income, regional rates of education, housing overcrowding, life expectancy, and medical resources), and demographic factors (such as migration flows, age structure of the population, and population density). Areas at risk of dengue are first selected based on the predicted presence of at least one of the two mosquito vectors responsible for dengue’s transmission: \textit{Aedes aegypti} and \textit{Aedes albopictus}. In those regions where the vectors had a high probability of presence, we assess the impact of the composite socio-economic indicators (derived through factor analysis to account for collinearity), and three composite demographic indicators (also derived from factor analysis) on the regional distribution of dengue cases, controlling for climate and spatial correlation.

\textbf{Results:} We found that an increase of one unit in one of our socio-economic indicators representing labour force with at least secondary education, better broadband access, and rooms per inhabitant, a higher proportions of active physicians is related to a drop in the occurrence of dengue, whereas the demographic indicators such as population density, age structure of the population and population growth showed no significant impact after taking climate into account. More importantly, our socio-economic indicator can also explain differences in the occurrence of dengue across Mexico, whereas simpler measures, such as regional GDP could not.

\textbf{Conclusions:} These results suggest that the set of indicators developed is a better indicator than GDP at predicting the distribution of dengue, by capturing information that is much more tailored to poverty related conditions which aid dengue transmission. Given that data for these indicators are available at a sub-national scale for OECD countries and selected OECD non-member economies, these indices may help us better understand factors responsible for the global distribution of dengue and also, given a warming climate, may help us to better predict vulnerable populations.

\textbf{Keywords:} dengue; climate-change; global-warming; socio-economic; mosquito-borne; vector-borne-diseases; GDP
1 Introduction

The dengue virus (DENV) is one of the most important mosquito-borne viral diseases in the world today; it is a disease caused by any one of four closely related viruses (DENV 1, DENV 2, DENV 3, or DENV 4). Two main arthropod vectors are responsible for transmission of dengue viruses: *Aedes aegypti* (commonly known as yellow fever mosquito) and *Aedes albopictus* (commonly known as tiger mosquito). *A. aegypti* mainly feeds on humans and is highly adapted to human habitations and urban areas; *A. albopictus* feeds on animals and humans and is more prevalent in rural and peri-urban environments. While *A. albopictus* is also responsible for dengue transmission among humans, it is a less likely vector than *A. aegypti* since it is adapted to a wider range of environments and has less restrictive feeding habits [1].

According to the World Health Organisation [2], DENV causes an acute flu-like illness that affects people of all age groups. This can lead to severe dengue, which causes potentially lethal complications, and sometimes death. There is currently no specific antiviral therapy for dengue fever; therefore, once the disease is contracted, there is no way to combat it other than relying on the host’s immune response. Several vaccines are currently in development; however, given the current cost-effectiveness, efficacy, safety and estimated impact of vaccination, the WHO’s current recommendation is to introduce it only in geographic settings (national or sub national) where the disease is particularly problematic [3].

1.1 Motivation for study

Climate change is likely playing an increasing role in dengue transmission and aiding its expanse across the globe. Several studies have predicted that rising temperatures will facilitate the spread of dengue [5–9] since temperature is one of the main drivers of dengue transmission. However, most work in this domain does not account for socio-economic factors other than GDP. Socio-economic conditions in a given location can be vital for a disease to persist once local transmission has occurred [10–15]. While studies have looked at the interaction between climate, socio-economic factors and demographics at a local level [16–20], they only focus on factors specific to local areas and issues arise when trying to extrapolate results from local level studies to macro level studies, since variables or predictors used at a local level may not be comparable across different locations and/or because of a lack of data availability at such scales. To get better estimates of where dengue may spread, there is a need to understand how climate factors, socio-economic factors and demographic factors interact over a greater geographic scale to reveal common global patterns.

The original contribution of this article is that it selects factors shown to predict dengue at a local level and tests whether the association can be generalized to the regional or state level. In addition, we also test if we can find a better socio-economic predictor of dengue than GDP. Although a useful and parsimonious indicator, GDP is a very broad measure and it is not necessarily reflective of well-being and population health, distribution of wealth, discrimination and spending on public welfare [21]. More importantly, GDP alone may not be able to capture cross-regional differences. The predominance of using GDP as an indicator has been largely questioned [22–25]; for some time now researchers in human health geography, critical public health, and social epidemiology have requested more careful
consideration of the contextual social and economic conditions that shape disease at
the local level [26,27]. We explain state and regional differences in the incidence of
dengue by controlling for socio-economic and demographic factors such as household
income, regional rates of education, housing overcrowding, life expectancy, medical
resources, migration flows, age structure of the population (the proportion of people
under 14 and over 65), and population density. The underlying idea is that a more
sophisticated indicator should be able to explain differences that may otherwise be
overlooked when controlling for GDP only. This is important as it can make visible
and explain health inequalities, both in “rich” and “poor” countries.

The study focuses on the occurrence and distribution of dengue in Mexico and
southern regions of the United States (US), because some near-border regions share
very similar environmental conditions but have distinct socio-economic factors [12].
Furthermore, this study takes advantage of time series data between 2012 and 2019
and it is, therefore, able to exploit two sources of variation: cross sectional, between
states and between regions, and over time for each region/state.

A conceptual framework has been formulated to illustrate the potential relation-
ships between variables that will be the object of the empirical analysis. It was
developed by carrying out a literature review and identifying a series of environ-
mental and geographical factors, demographic factors, biological factors, and socio-
economic factors potentially affecting the spread of dengue (Figure 1). The concep-
tual framework is formulated in a general way so that different data sources can
be used to test these relationships. The following section gives an overview of the
findings and provides a rationale to explain variable selection.

1.1.1 Environmental and Geographical factors
Since dengue is a vector borne disease, i.e. it relies on a secondary organism as its
primary mode of transport and delivery, understanding the key ecological require-
ments of its vectors is crucial. Both Aedes mosquitoes are ectothermic organisms
and are highly sensitive to colder temperatures and extreme high temperatures. A
study that reviewed the temperature requirements of both vectors under laboratory
conditions found that \textit{A. albopictus} adults can survive in temperatures from 15 to
35°C and \textit{A. aegypti} from 10 to 35°C [28]. It also reported that for both vectors,
growth and development is severely inhibited in ambient temperatures below 13°C
or above 35°C and that, although \textit{A. aegypti} can endure a wider range of tem-
peratures, its survival at temperatures below 14-15°C is limited to short periods,
since its mobility is severely restricted and its ability to imbibe blood impeded. \textit{A.
albopictus} eggs though, can go through diapause (suspended development) when
exposed to extreme cold (down to -10°C). This adaptation allows them to inhabit
environments with a wider annual temperature range, with more distinct seasonal
changes than in tropical climates, where climate is more homogeneous. The latter
study also reported that \textit{A. aegypti} is highly sensitive to fluctuations in tempera-
tures. For most mosquito species, availability of fresh water habitat, humidity and
precipitation are highly indicative of their distribution in the environment. This is
also the case for both \textit{Aedes} mosquitoes in their native habitats, where they are
highly adapted to breeding in aquatic habitats like ponds and lakes, but also mi-
cro habitats, such as tree-holes, rock crevices and even leaf axils [29]. The latter
behavior in recent times has benefited both species by allowing them to exploit a range of man-made aquatic breeding habitats, where water can accumulate, like urban gardens, vases in cemeteries, discarded bottles and plant pots; therefore, both species can survive in drier climates than expected, since they are exploiting artificial water sources. We selected a range of humidity and temperature variables for analysis, which would capture the mosquitoes living requirements. Although both species can exploit similar habitats, *A. aegypti* feeds predominantly on humans and is restricted to urban environments and other human settlements, whereas *A. albopictus* feeds on humans as well as a range of mammals and birds, and therefore tends to occupy rural and peri-urban environments [31].

1.1.2 Socio-economic factors
Density of both the vector and host are fundamental factors in disease transmission, as contact between infected vectors and susceptible hosts (and vice versa) is the source of new infections [32]. It is known that socio-economic factors tend to influence the distribution and intensity of dengue both pre-infection and post-infection [12, 33–37]. One of the main reported risk factors of dengue was home water storage (rather than receiving piped water), poor sanitation, and poor public services (e.g. litter not collected). Such factors were responsible for creating breeding habitat for mosquitoes and bringing them into closer contact with humans, therefore increasing the risk of dengue. Furthermore, use of mosquito nets, insect screens, and air-conditioning, all limited the chance of being bitten. Knowledge and education of mosquito ecology also helped residents make personal interventions and reduce risk of being bitten. However, it is important to note that factors associated with higher economic status can also bring humans into closer contact with mosquitoes, for example homeowners with gardens and potted plants and ponds or having good access to recreational space where mosquitoes can breed [38]. Although we could not use data that would directly measure these factors, we selected co-variates that would work as proxies, capturing a latent variable that would represent vector risk. The rationale is that people living in locations with better socio-economic conditions would do better to avoid contact with mosquitoes and restrict virus transmission, either from the bottom-up (e.g. personal interventions) or the top-down (e.g. regional government pest control). We selected the following variables in the OECD regional data repository to describe the actual socio-economic position of residents in a region: “Household income”, “Life expectancy at birth”, and a measure of housing overcrowding “Number of rooms per person”. Furthermore, “Secondary education” would also help to capture areas where there is a higher proportion of manual labourer, e.g. agricultural workers or people working outdoors who may be more exposed to mosquitoes. We also selected “Self-evaluation of life satisfaction”, and “Perception of corruption” to try to capture additional features of a region. Since these two variables yield some indication of how people perceive their surroundings, we assume that poorer scores will capture poor infrastructure, poor public services, lack of basic provisions and lack of beneficial government intervention. In terms of post-infection factors that influence dengue transmission, access to health care, risk perception and access to information on dengue infection symptoms had positive effects on people’s decision to seek medical help when presented with dengue
infection symptoms [14, 39, 40]. This is likely to be one of the main reasons why
dengue tends to stop at the US border. To reflect this in the conceptual framework
we selected variables that would proxy access to health care i.e., “Physician rate”,
and variables which would represent access to information and personal knowledge
i.e., “Secondary education”, “Broadband access” and “Perceived social network
support”. However, it is important to note that those areas with better access to
health care are also more likely to report cases of dengue, so the relationship may
be harder to infer.

1.1.3 Demographic factors
Considering the points laid out by [32], host population density can have a direct
impact on the intensity of disease. We would expect human population density to
benefit both the vector and the virus. To represent this in the data set, we selected
‘Population density” and “Population density growth”. A study by [41] investigated
some of the causes for increased dengue incidence in Acre, Brazil. The authors
found that unplanned urbanization and population growth contributed to creating
beneficial conditions for *A. aegypti*. Furthermore, well connected areas, in terms of
trade and transport, with considerable human movement, can benefit both mosquito
species and dengue, by facilitating their movement and spread [42,43]. To capture
these factors, we selected “Inter-regional migration rate” and “Population density
growth”. We assume that these factors would be significant in poorer areas where
unplanned urbanization, poor infrastructure and poor public services are more likely
to exist. Finally, younger people are more likely to be infected by dengue [44], so we
controlled for the age structure of the population by selecting “Percentage of Old
Population Group (65+)” and “Percentage of Youth population group (0-14)”.

2 Materials and Methods
In this study, we compiled a data-set that would reflect the conceptual framework
by including variables that either directly measure factors contributing to the dis-
tribution of the vectors and transmission of dengue; or indirect measures (proxies)
that are reflective of conditions known to influence dengue transmission.

One of the difficulties in compiling such a data-set using disparate sources is
finding data that is comparable in terms of scale, since most health and economic
data are provided at relatively crude and differing resolutions, and presented in the
form of areal data (regional).

2.1 Data extraction and methods to estimate population at risk
Dengue case data is only provided at the state or equivalent level, it is difficult to
obtain a good estimate of the population at risk since the vector can be present
in only certain locations within a region. Therefore, to extract more accurate data
on the population at risk, and in areas where climatic and environmental factors
would contribute to disease transmission, we estimated the distribution of the *Aedes*
mosquitoes using a generalized logistic regression that could predict the likelihood a
vector would occur in a region given annual temperature range, mean temperature of
the coldest quarter, precipitation during the driest quarter and precipitation during
the warmest quarter. We then identified the areas at risk based on the prediction
of the logistic model. We also only selected the most southern regions at risk in the United States as these areas are those most vulnerable because of their close proximity to endemic regions in Mexico. More specific information on statistical methods and results from this analysis can be found in Additional file 1. Figure 2 shows the point *Aedes* data sample locations and results of the modelling.

To parameterize the distribution models, point location occurrence data for *A. aegypti* and *A. albopictus* were obtained from [45,46]. Point occurrence data represent spatial geo-coordinates of a location in which a given individual organism was sampled or sighted. Many of the samples in these data-sets consist of museum records or unpublished studies including national entomological surveys. These data are used to extract information about the underlying factors that influence an organism’s distribution, which can then used to predict distributions across landscapes.

Climate data were extracted using R’s DISMO package in all point locations where mosquitoes occurred. Climate data for the species distribution prediction modelling were sourced from MERRAclim [47]. This data-set was built using 2 m air temperature (Kelvin degrees) and 2 m specific humidity (kg of water/kg of air) hourly data derived from satellite observations from the Modern Era Retrospective Analysis for Research and Applications Reanalysis. Tables providing summary statistics for the climate values at *Aedes* point locations can be found in Additional file 1.

2.2 Regional data: socio-economic, demographic and climate data extraction methods

2.2.1 Climate data

Climate data for the main analysis i.e. measuring the impact of the climate variables on dengue transmission were sourced from the Climate Prediction Center (CPC) of the National Centers for Environmental Prediction (NCEP), see [48]. This data represents a global summary of daily weather data. The CPC extracts surface synoptic weather observations from the Global Telecommunications System (GTS), which collects global data from a combination of weather station and satellite observations. Files were processed in R with the NetCDF, Raster and Dismo packages in order to create annual bio-climatic variables. The bio-climatic variables in this study were derived from daily maximum temperatures, daily minimum temperatures and total daily rainfall.

The Global Administrative Unit Layers [49] data-set along with our *Aedes* distribution maps (results figure 2, bottom right) were used to spatially capture, process and convert all the necessary data (for the main analysis) to a regional level. The GAUL data set contains geographic information in the form of shape files that lay out within country boundaries linked to a unique nomenclature. Countries are broken down into statistical subdivisions e.g., ADM0 representing data at country level (e.g. US), ADM1 at regional level (e.g. California), and ADM2 at subregional level (e.g. Orange County).

2.2.2 Dengue Case Data

Dengue case data for Mexico 2012-2016 were obtained at request from Healthmap [50] and case data for 2017-2019 were extracted from the PLISA Health Information Platform for the Americas (only available years). To extract the most valid and comparable observations, we only selected ProMED reports from the Healthmap
data-set since they are both sourced from the same mandatory notification system covering all of the national territory. All data from these reports are aggregated at the regional level (ADM1 level).

Cases data for the United States were extracted from, since data is provided at the county level (ADM2 level) we needed to aggregate them to the state level (ADM1 level) in order to match them with the main data-set.

### 2.2.3 Socio-Economic and Demographic Data

Socio-economic and demographic data were extracted from the OECD’s Regional Statistics and Indicators Database [51]. This database provides comparable statistics and indicators and is presented in yearly time series. Missing values were filled based on values for previous years or subsequent years, depending on their position in the data set.

Population count data to predict the number of persons at risk in a region were sourced from the Socioeconomic Data and Applications Center’s [52] Gridded Population of the World data set. This data set estimates population count for the years 2000, 2005, 2010, 2015, and 2020, consistent with national censuses and population registers. Data were extracted from areas where vector presence was predicted. R’s Zoo package was used to replace values for missing years, by implementing a linear interpolation method that would predict trends between years. This way increases or decreases in human population were controlled for in the final model.

Table 1 provides summary statistics of all the collected data for the final models.

### 2.3 Statistical Methods

#### 2.3.1 Factor Analysis - Data Processing for Regional Analysis

The socio-economic variables are strongly correlated with each other, as well as some of the demographic variables, and if included in a regression would give rise to multi-collinearity issues. By over-inflating the standard errors, multi-collinearity makes some variables statistically insignificant when they should be significant. To address this problem, following similar methods to [53], a factor analysis by maximum likelihood (VARIMAX rotation) was performed on each of the main subgroups of variables i.e. socio-economic and demographic. Factor analysis is a method for investigating whether a number of variables of interest $Y_1, Y_2, ..., Y_n$, are linearly related to a smaller number of latent (i.e. not directly measured) factors $F_1, F_2, ..., F_k$. The basic concept of factor analysis is that multiple observed variables have similar patterns because they are all associated with a latent variable. The factors are constructed in such a way that they capture the maximum amount of common variance (correlation) of the original items; the eigenvalue is a measure of how much of the variance of the observed variables a factor explains. The factor analysis can be formalized as follows:

\[
Y_1 = \beta_{10} + \beta_{11}F_1 + \beta_{12}F_2 + ... + \beta_{1k}F_k + \epsilon
\]

\[
Y_2 = \beta_{20} + \beta_{21}F_1 + \beta_{22}F_2 + ... + \beta_{2k}F_k + \epsilon
\]

\[
Y_N = \beta_{n0} + \beta_{n1}F_1 + \beta_{n2}F_2 + ... + \beta_{nk}F_k + \epsilon
\]
Before performing factor analysis, all variables had to be standardized to z-scores \((x - \mu)/\sigma\) to ensure that they were on the same scale. After performing the factor analysis, the predicted values for the factors for any individual region can be estimated. These predictions, known as factor scores, are weighted sums of the values of the observed items. Roughly, indicators that are more reliable measures of a factor (i.e. those with larger loadings) will receive higher weights in the calculation of a factor score for that factor.

2.3.2 Socio-economic and demographic indices Mexico/US

The Mexico/US factor analysis yielded two socio-demographic indices (Table 2) and three demographic indices (Table 3). Regions with high scores on the first socio-economic factor tend to have a higher share of labour force with at least secondary education, better broadband access, more rooms per inhabitant and a better perception of social network support and a higher rate of active physicians. The second socio-economic factor mainly captured the variation from the life expectancy at birth, primary income of households, and a higher rate of active physicians and better broadband access. A priori, the socio-economic indicators are expected to have a negative association with dengue. Regions with high scores on the first demographic factor (see Table 3) tend to have faster growing human populations, and higher rates of inter-regional migration. A priori, the indicator is expected to have a positive association with dengue. The second demographic factor captured the age structure of the population and the third demographic factor captured population density in the areas where Aedes mosquito was present.

2.3.3 Socio-economic and demographic indices Mexico

Regions with high scores on the first socio-economic factor (see Table 4) tend to have better broadband coverage, higher share of labour force with at least secondary education, a better perception of social network support, and higher levels of life satisfaction and life expectancy at birth. The second factor mainly captures rooms per inhabitant, number of doctors, and a higher share of labour force with at least secondary education. A priori, the socio-economic indicator is expected to have a negative association with dengue.

Regions with high scores on the first demographic factor (see Table 5) tend to have faster growing human populations, higher rates of inter-regional migration and higher population densities. A priori, the indicator is expected to have a positive association with dengue. The second demographic factor captured the age structure of the population and the third demographic factor captured population density in the areas where Aedes mosquito was present.

2.3.4 Calculating Incidence Rate, Risk and the Standardized Mortality Ratio

After selecting the areas at risk and combining the main predictors into composite indicators, we performed an exploratory analysis of the incidence of dengue. In particular, we used the Standardized Mortality Ratio (SMR) to test if there were “excess cases” in a region. The SMR is a ratio between the observed number of cases in a study population and the number of cases that would be expected, based on the population size of the study population. If the ratio of observed dengue
cases: expected dengue cases is greater than 1.0, we would consider there to be “excess cases” in the study population.

To calculate the SMR we first calculated the Incidence Rate: \( R_i = \frac{N_i}{P_i} \)

we then calculated the Overall Rate: \( R = \frac{\sum N_i}{\sum P_i} \) Overall rate is \( \frac{\text{Total Cases}}{\text{Total Population}} \)

Rate is \( \frac{\text{Number Of Cases}}{\text{Population At Risk}} \)

Followed by the number of Expected Cases:

\[ E_i = R \times P_i \]

Expected Cases is Overall Rate x Population

The final equation for the Standardized Morbidity Ratio is: \( SMR_i = \frac{N_i}{E_i} \) Std Morbidity Rate is \( \frac{\text{No.of Cases}}{\text{Expected}} \)

We postulate that the excess cases can be explained by the socio-economic, demographic and climate variables, which means that regions with an SMR higher than 1.0 should differ from the others along those indicators in line with the predictions of the final model (illustrated in the next section).

2.3.5 Negative binomial regression model to assess impact of independent variables on dengue case data at regional level.

Generally, count data can be modeled using a Poisson distribution. However, the mean of our dependent variable (dengue cases by region and year) was lower than its variance - \( E(Y) < \text{Var}(Y) \), suggesting that the data are over-dispersed. Over-dispersed Poisson processes can be modeled in many alternative ways. The most common approach is to assume a negative binomial distribution, which is particularly suited when the variance is much larger than the mean. We therefore assume that

\[ \text{Dengue}_{it} \sim NB(\mu, k) \]

that is, the number of Dengue cases in region \( i \) at time \( t \) are distributed as a negative binomial with mean \( \mu_i \) and dispersion parameter \( k \). The variance is given by:

\[ \text{VAR}(\text{Dengue}_{it}) = \mu + \left( \frac{\mu^2}{k} \right) \]

As well as accounting for over-dispersion in the data, we want to account for possible non-linearities, so we proceed with a Generalized Linear Model (GLM) with a negative binomial distribution and logarithmic link function. The GLM model is given by:

\[ E(\text{Dengue}_{it}) = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n) \]

The vector of explanatory variables \( X \) includes: Two socio-economic indices, three demographic indices, and two climate variables, namely mean temperature of warmest quarter and precipitation of warmest quarter, plus a vector of time fixed effects.

We also accounted for spatial auto-correlation by introducing spatial random effects based on geographical regions. This involves creating an adjacency matrix of the regions, where if there are \( g \) regions, the adjacency matrix is a
indicator matrix where there is some non-zero value when region i is connected to region j, and 0 otherwise. In addition, an approach similar to that for a random effect is used to incorporate observations belonging to specific regions.

3 Results
We conducted two analyses, one focused on regions within Mexico and the United States and one focused on Mexico only.

3.1 US/Mex analysis
Figure 2 shows the results of the SMR and the factor analysis on the socio-economic variables, and regional GDP for the US/Mex analyses. Our indices are able to demonstrate that there is much more variation between regions than we would capture than using GDP alone. However, it is difficult to infer a relationship by visual inspection alone.

Table 6 shows the results for the regression model comparing the US and Mexico (GLM with negative binomial link) for 2012-2019. Table 7 restricts the analysis to Mexico only. It should be noted for the Mexican analysis we only selected data 2017-2019 using data from the PAHO, so we could combine confirmed, suspected and probable cases that would give us a better idea of the actual incidence rate in Mexico during those years, rather than using confirmed cases of a longer period. Regressions were run on a pooled cross section; time trends were accounted for by controlling for year fixed effects; regional correlation was accounted for by introducing regional weights (neighborhood boundary approach). All models use population in the area at risk as the offset (exposure) variable.

The first column of table 6 (GDP Model) shows the association between regional GDP and dengue cases across the regions. The model presents a negative association between GDP and dengue as predicted however it is not significant.

The second column (SE Model) of Table 6 shows the association between regional dengue cases and the socio-economic indicators derived through factor analysis. The coefficient for the first socio-economic indicator is negative, more than twice as large in magnitude as the coefficient of GDP alone, and statistically significant.

Given the weight of each variable in the factor analysis, we can say that an increase in education, household income, broadband access and a decrease in housing overcrowding is associated with fewer dengue cases; the third column (Dem Model) of table 6 shows the regression controlling for the three demographic indicators derived from factor analysis, the first index representing population flows and population density has a positive impact on dengue. Column 4 (Clim Model) includes the climate variables alone; mean temperature of the warmest quarter is significantly associated with dengue (and the relationship is positive as predicted), while precipitation does not show any significant association.

The "full model" in column 5 shows the relationship between dengue and all variables. When controlling for demographic and climate variables, the impact of the socio-economic indicator 1 becomes slightly larger (from -0.16 to -0.16) and highly statistically significant. The impact of temperature increases slightly and remains highly significant.
3.2 Mex analysis

Figure 3 shows the results of the SMR and the factor analysis on the socio-economic variables, and regional GDP for the Mex analyses. Again, our indices are able to demonstrate that there is much variation between regions than we would capture than using GDP alone.

The Mexican only factor analysis yielded slightly different results to the US/Mex factor analysis. Our first index captures better broadband coverage, higher share of labour force with at least secondary education, a better perception of social network support, higher levels of life satisfaction and life expectancy at birth. The first result worthy of notice is that if we look at Mexican regions alone, the socio-economic index 1 is no longer associated with dengue. The demographic factor 1 is no longer associated either. However when we control for temperature it becomes significant, which is a similar result to the US/Mex regression analysis.

4 Discussion

4.1 Impact of socio-economic, demographic index and Aedes predicted distribution on dengue

The results confirmed a strong association between our novel indices of socio-economic factors and dengue cases per region. Such results are consistent with the findings reported by [12,14,34–37,40,54]. Our findings enrich the field by offering a set of more refined indicators that can be generalized to other diseases with similar transmission circles. Furthermore, while GDP did not appear to capture differences between Mexico and the US states included in the study, the composite indicator we created was robust enough to capture more nuanced differences within Mexico, as the analysis looking at Mexico alone demonstrated.

Although the variables used in this study do not represent disease transmission mechanisms, understanding the relative significance of these factors on disease outcomes can help risk assessors predict where diseases are likely to occur in the future, by identifying locations with vulnerabilities in public health systems and/or by identifying impoverished areas that tend to be susceptible to disease. Furthermore, this study can also have important policy implications as it points to the important role of social and economic equality, social cohesion, and transparent government, among others, in public health.

Future studies seeking to test the robustness of the indicators developed in this study should try to source data at a more refined scale over a wider geographical area. Furthermore, statistical analysis should be conducted using more refined temporal dimensions to further increase observations and statistical power, and allow us to better investigate the causal impacts of socio-economic, demographic and environment factors on dengue and other vector-borne diseases.

One of the main obstacles in conducting this study was finding adequate data at refined scales. Because of a lack of data at a desirable scale, we resorted to using data at first statistical country subdivisions e.g. FAO ADM 1 US and Mexican states. The shortcomings of this study are therefore a lack of detail and variation within regions and a relatively small sample size, since the number of observations was limited to the number of US and Mexican states. We may also question the quality of the cases data since many people living in Mexico and the United States
do not have adequate access to health care and so cases may go unreported. This is particularly true in the case of the US where more vulnerable groups like illegal migrants, the poor and homeless may not have any access to health care and the disease will go unreported. This is also the case for socio-economic data since huge disparities exist between Native American communities living on reservations and the rest of the population, and it is not clear if data from these self-governing regions are included in the regions.

5 Conclusions
This study investigated the potential development of poverty-related indicators using regional and internationally available data, and also investigated the role of socio-economic factors on the occurrence of dengue in the US and Mexico. To identify which regions are most at risk, we estimated where dengue vectors are likely to occur given their suitability to climate conditions. By estimating the chance of a vector occurring in a region, we could then assess the impact of socio-economic, demographic and climate factors on dengue. The results of this study demonstrate how the use of new composite socio-economic indices can explain with greater accuracy, the differences in the spread of disease in places with similar physical geography and ecological characteristics, more than GDP alone can. Our findings are not only significant for public health, proving the usefulness of our proposed methodology in explaining and predicting disease, but also contribute to a wider scholarly debate on whether and to what extent can economic growth (measured via GDP) contribute to better outcomes of health and well-being. Overall, these results, therefore, suggest the mitigating role of socio-economic factors, other than simply GDP, in reducing the incidence of dengue.

5.1 Abbreviations
GDP: Gross Domestic Product; Mex: Mexico; US: United States of America; SDM: Species distribution model, ADM: Administrative division.

5.2 Ethics approval and consent to participate
Not applicable

5.3 Consent for publication
All authors and co-authors have approved the publication.

5.4 Availability of data and materials
The main data-set is available to download as a csv in the supporting information documents, all other data is publicly available.

5.5 Competing interests
No competing interests

5.6 Authors’ contributions
MW led the work and was responsible for the conceptualization of the project, data curation, data processing, formulation of the methodology, statistical analysis, modelling and writing the original draft and interpreting the results. PK made substantive contributions to writing and formulating some of the concepts and interpretation of the results. All contributors revised the manuscript and copy-edited the final submission version. All the authors were also involved in revising the manuscript critically for important intellectual content. CUB supervised statistical analysis and made substantive contributions to formulating some of the statistical models. All authors read and approved the final manuscript. PK, PGM VSM supervised the project and are responsible for formulating ideas for the umbrella project Impacts of Climate Change (CC) on Human Health (HH) at ICTA-UAB: Integrating socio-economic and policy studies with natural science studies to enhance consilience of climate policy science.
5.7 Acknowledgements

This research was funded by ICTA’s Maria de Maeztu Unit of Excellence, awarded by the Spanish Ministry of Economy and Competitiveness. The award is the highest institutional recognition of scientific research in Spain. Thanks also to Patrizia Ziveri and Pedro Manuel Gonzalez Hernandez for supporting the project.

Author details

1Institute of Environmental Science and Technology (ICTA), Autonomous University of Barcelona (UAB), Bellaterra, Spain, Barcelona, Spain. 2Joint Research Centre (JRC-Seville), European Commission, Sevilla, Spain.

References

1. Murray NEA, Quam MB, Wilder-Smith A. Epidemiology of dengue: Past, present and future prospects. Clinical epidemiology.2013;5:299.
2. WHO. Dengue and severe dengue. 2018; Available from: http://www.who.int/immunization/research/development/dengue_vaccines/en/
3. Bernardes Terzian AC, Mondini A, Moraes Bronzoni RV de, Drumond BP, Ferro BP, Sotello Cabrera EM, et al. Detection of Saint Louis encephalitis virus in dengue-suspected cases during a dengue 3 outbreak. Vector-Borne and Zoonotic Disease.2011;11(3):291–300.
4. WHO. Immunization, vaccines and biologicals. 2017:2018. Available from: http://www.who.int/immunization/research/development/dengue_vaccines/en/
5. Butterworth MK, Morin CW, Comrie AC. An analysis of the potential impact of climate change on dengue transmission in the southeastern united states. Environmental health perspectives 2017;125(4):579–85.
6. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of aedes-albopictus virus transmission risk with climate change. PLOS Neglected Tropical Diseases.2019;13(3):e0007213.
7. Messina JP, Brady OJ, Goldberg N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nature Microbiology.2019;4(9):1508–15.
8. Xu Z, Bambrick H, Frentiu FD, Devine G, Yakob L, Williams G, et al. Projecting the future of dengue under climate change scenarios: Progress, uncertainties and research needs. PLOS Neglected Tropical Diseases.2020;14(3):e0008118.
9. Ebi KL, Nealon J. Dengue in a changing climate. Environmental Research.2016;151:115–23.
10. Bouzid M, Colon-Gonzalez FJ, Lung T, Lake IR, Hunter PR. Climate change and the emergence of vector-borne diseases in Europe: Case study of dengue fever. Bmc Public Health.2014;14
11. Messina JP, Brady OJ, Goldberg N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nature Microbiology.2019;4(9):1508–15.
12. Brunkard JM, Lopez JLR. Ramirez J, Cifuentes E, Rothenberg SJ, Hunsperger EA, et al. Dengue fever seroprevalence and risk factors, Texas-Mexico border, 2004. Emerging Infectious Diseases .2007;13(10):1477–83.
13. Abeita A, Smith B, Oh JO, Jiw S, Fokte M, Betz T, Gaul L, Robles-Lopez JL, et al. Dengue hemorrhagic fever - US-Mexico border, 2005. Morbidity and Mortality Weekly Report.2007;56(31):785–9.
14. Ramos EF. Hemoterapia e febre dengue. Revista Brasileira de Hematologia e Hemoterapia.2008;30(1):64–6.
15. Magori K, Drake JM. The population dynamics of vector-borne diseases.Vol.4.2013.
16. Vincenti-Gonzalez MF, Gillette ME, Velasco-Salas ZI, Lizarazo EF, Amariara MA, Sierra GM, et al. Spatial analysis of dengue seroprevalence and modeling of transmission risk factors in a dengue hyperendemic city of Venezuela. Plos Neglected Tropical Diseases.2017;11(1).
17. Toan DTT, Hoat LN, Hu W, Wright P, Martens P. Risk factors associated with an outbreak of dengue fever/dengue hemorrhagic fever in hanoi, Vietnam. Epidemiology and Infection.2015;143(6):1594–8.
18. Tipayamongkholgul M, Lisakulruk S. Socio-geographical factors in vulnerability to dengue in Thai villages: A spatial regression analysis. Geospat Health.2011;5.
19. Teurlai M, Menkes 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 Neglected Tropical Diseases.2015;9(12):e0004211.
20. Akter R, Naish S, Hu W, Tong S. Socio-demographic, ecological factors and dengue infection trends in Australia. PLoS One.2017;12(10):e0185551.
21. Robert C, Kubiszewski I, Giovannini E, Lovins H, McGlade J, Pickett K, et al. Time to leave gdp behind. Nature.2014;505(7483).
22. Stiglitz JE, Sen A, Fitoussi J-P. Mismeasuring our lives: Why gdp doesn’t add up. The New Press; 2010.
23. Bleys B. Beyond gdp: Classifying alternative measures for progress. Social Indicators Research.2012;109(3):355–76.
24. Van den Bergh JC. The gdp paradox. Journal of Economic Psychology.2009;30(2):117–35.
25. Costanza R, Kubiszewski I, Giovannini E, Lovins H, McGlade J, Pickett KE, et al. Development: Time to leave gdp behind. Nature News.2014;505(7483).
26. Berkman LF, Kawachi I, Glymour MM. Social epidemiology. Oxford University Press; 2014.
27. Navarro V. Assessment of the world health report 2000. The Lancet.2000;356(9241):1598–601.
28. Brady OJ, Johansson MA, Guerra CA, Bhatt S, Gething PW, Moyes CL, et al. The current and future global distribution and population at risk of dengue fever in the tropical rainforest of imo state, south-east Nigeria. Ann Agric Environ Med. 2007;14(1):31–8.
29. Anosike JC, Nwoke BE, Okere AN, Oku EE, Asor JE, Emmy-Egiebo IO, et al. Dengue hemorrhagic fever/dengue hemorrhagic fever and its public health implications in dares salaam, tanzania anitha philbert1* and jasper. N. Ijumba2. Vol. 4.2013.
30. Philbert A. Preferred breeding habitats of aedes aegypti (diptera-culicidae) mosquito and its public health implications in dares salaam, tanzania anitha philbert1* and jasper. N. Ijumba2. Vol. 4.2013.
31. Waldock J, Chandra NL, Lelieveld J, Proestos Y, Michael E, Christophides G, et al. The role of environmental variables on aedes albopictus biology and chikungunya epidemiology. Pathogens and Global Health.2013;107(5):224–41.
32. Begon M. Ecological epidemiology. In: The princeton guide to ecology. Princeton University Press; 2009. 33.
33. Thammapalo S, Chongsuwiwatwong V, McNeil D, Geater A. The climatic factors influencing the occurrence of
dengue hemorrhagic fever in thailand. Southeast Asian J Trop Med Public Health. 2005;36.
34. Stewart Ibarra AM, Ryan SJ, Beltrán E, Mejía R, Silva M, Muñoz A. Dengue vector dynamics (aedes aegypti)
influenced by climate and social factors in ecuador: Implications for targeted control. PLOS ONE.2013;8(11):e78263.
35. Thammapalo S, Chongsuwiwatwong V, Geater A, Lim A, Choomalee K. Socio-demographic and environmental factors
associated with aedes breeding places in phuket, thailand. Southeast Asian J Trop Med Public Health. 2005;36(2):426–33.
36. Qi X, Wang Y, Li Y, Meng Y, Chen Q, Ma J, et al. The effects of socioeconomic and environmental factors on
the incidence of dengue fever in the pearl river delta, china. 2013. PLOS Neglected Tropical Diseases.2015;9(10):e0004159.
37. Clark GG. Dengue and dengue hemorrhagic fever in northern mexico and south texas: Do they really respect the
border? Am J Trop Med Hyg. 2008;78(3):361–2.
38. Umu I, Farajollahi A, Strickman D, Fonseca DM. Crouching tiger, hidden trouble: Urban sources of aedes
albopictus (diptera: Culicidae) refractory to source-reduction. PloS one. 2013;8(10):e77999–9.
39. Khun S, Manderson L. Health seeking and access to care for children with suspected dengue in cambodia: An
ethnographic study. BMC Public Health.2007;7:262.
40. Elsainga J, Lizarazo EF, Vincenti MF, Schmidt M, Velasco-Salas ZI, Arias L, et al. Health seeking behaviour and
treatment intentions of dengue and fever: A household survey of children and adults in venezuela. PLOS Neglected
Tropical Diseases. 2015;9(12):e0004237.
41. Lana RM, Costa Gomes MF da, Melo de Lima TF, Honorio NA, Codeco CT. The introduction of dengue follows
transportation infrastructure changes in the state of aoc, brazil: A network-based analysis. Plos Neglected Tropical
Diseases.2017;11(11).
42. Gubler DJ. Dengue, urbanization and globalization: The unholy trinity of the 21(st) century. Tropical Medicine
and Health . 2011;39(4 Suppl):3–11.
43. Gubler D. Prevention and control of aedes aegypti-borne diseases: Lesson learned from past successes and
failures. Vol. 19. 2013.
44. ACAPS. ACAPS briefing note: Mexico - dengue fever (16 september 2019). 2019; Available from:
https://reliefweb.int/report/mexico/acaps-briefing-note-mexico-dengue-fever-16-september-2019
45. Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, et al. The global compendium of aedes
aegypti and ae. Albopictus occurrence. Scientific Data.
46. GBIF.org. GBIF occurrence download. 2018; Available from: https://doi.org/10.15468/d1.ygqvsj
47. C. Vega G, Pertierria LR, Ollalla-Tárraga MÁ. MERRAclim, a high-resolution global dataset of remotely sensed
bioclimatic variables for ecological modelling. Scientific Data. 2017;4:170078.
48. CPC/NCEP. National center for atmospheric research. 1987; Available from:
http://rda.ucar.edu/datasets/de512.0/
49. FAO-UN. Global administrative unit layers (gaul). 2014; Available from:
http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691
50. Healthmap. Dengue case reports 2011-2017. 2017; Available from: http://www.healthmap.org/en/
51. OECD. Regional statistics and indicators database. 2018; Available from:
http://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR
52. SEDAC. Gridded population of the world, version 4 (gpw4): Population count, revision 11. 2018; Available from:
https://doi.org/10.7927/W6JW8X85
53. GFC. Spatial data analysis and modeling with r. 2018:2018. Available from: http://rep.spatial.org/index.html
54. Bruenkard JM, Cifuentes E, Rothenberg SJ. Assessing the roles of temperature, precipitation, and enso in dengue
re-emergence on the texas-mexico border region. Salud Publica De Mexico . 2008;50(3):227–34.
55. Johnson TL, Haque U, Monaghan AJ, Eisen L, Hahn MB, Hayden MH, et al. Modeling the environmental
suitability for aedes (stegomyia) aegypti and aedes (stegomyia) albopictus (diptera: Culicidae) in the contiguous
united states.J Med Entomol. 2017;54(6):1605–14.
Figure 1  *Aedes* sample locations and species distribution modelling results.

1. Top left: *Aedes* point locations.
2. Top right: Results of *Aedes aegypti* SDM
3. Bottom left: Results of *Aedes albopictus* SDM
4. Bottom right: Population and climate data extraction locations given predicted aedes distributions (probability cutoff 75 percent).

Figure 2  SMR and distribution of socio-economic factors Mex/US.

Figure 3  SMR and distribution of socio-economic factors Mexico.
| Statistic                                      | N    | Min       | Max       | Mean     | St. Dev.   |
|-----------------------------------------------|------|-----------|-----------|----------|------------|
| Primary Income of Private Households (USD per head) | 351  | 2,541.8   | 1,886,380.0 | 128,928.1 | 358,536.5 |
| Regional Gross Domestic Product per Capita     | 351  | 2,883.4   | 1,632,730.0 | 116,785.2 | 314,862.4 |
| Share of labour force with at least secondary education | 351  | 36.5      | 90.1      | 64.7     | 11.0       |
| Share of households with internet broadband access | 351  | 7.3       | 88.7      | 45.6     | 20.9       |
| Self-evaluation of life satisfaction           | 351  | 6         | 9         | 7.2      | 0.6        |
| Perceived social network support              | 351  | 59        | 96        | 81.1     | 10.1       |
| Perception of corruption                      | 351  | 0.7       | 4.8       | 1.7      | 0.6        |
| Active Physicians Rate (physicians for 1000 population) | 351  | 70.5      | 80.8      | 75.3     | 1.6        |
| Life Expectancy at Birth                      | 351  | 0.7       | 2.7       | 1.2      | 0.6        |
| Number of rooms per person                    | 351  | −1.7      | 3.8       | 0.2      | 1.0        |
| Inter-regional net flows migration rate, (% net flows over population) | 351  | 0.5       | 7.0       | 2.0      | 1.2        |
| Inter-regional migration rate, (% migrants over population) | 351  | 0.7       | 7.9       | 208.4    | 703.7      |
| Population density growth                     | 351  | 31.1      | 20.5      | 8.1      | 3.3        |
| Population Density in Predicted Aedes Areas (pp km2) | 351  | 35.1     | 49.4      | 1.9      | 7.7        |
| Percentage of Old Population Group (65+)      | 351  | 64.3      | 26.5      | 3.7      |            |
| Percentage of Youth Population Group (0-14)   | 351  | 35.2      | 25.9      | 3.5      |            |
| Mean temperature of warmest quarter           | 351  | 0.0       | 9.1       | 1.6      | 2.1        |
| Precipitation of warmest quarter              | 351  | 0.0       | 9.1       | 1.6      | 2.1        |

| Table 1 Final dataset 2012-2019 |

| Factor 1 | Factor 2 |
|----------|----------|
| 0.37     | 0.71     |
| 0.89     | 0.44     |
| 0.56     | 0.59     |
| 0.07     | 0.23     |
| 0.48     | 0.24     |
| 0.40     | 0.11     |
| 0.44     | 0.42     |
| 0.28     | 0.94     |
| 0.93     | 0.29     |

| Table 2 Socio-economic Factor Analysis Results US/MEX |

| Factor 1 | Factor 2 | Factor 3 |
|----------|----------|----------|
| 0.84     | -0.05    | 0.06     |
| 0.62     | 0.27     | -0.19    |
| -0.09    | -0.06    | 0.99     |
| -0.19    | -0.98    | -0.05    |
| -0.03    | 0.91     | -0.14    |

| Table 3 Demographic Factor Analysis Results US/MEX |
### Table 4 Socio-economic Factor Analysis Results (Mexico)

| Factor          | Factor1 | Factor2 |
|-----------------|---------|---------|
| Primary Income  | 0.13    | 0.47    |
| Private         |         |         |
| Households (USD | per head)|         |
| Share of labour | 0.72    | 0.57    |
| force with      |         |         |
| at least        |         |         |
| secondary       |         |         |
| education       |         |         |
| Share of        | 0.79    | 0.41    |
| households      |         |         |
| with internet   |         |         |
| broadcast       |         |         |
| access          |         |         |
| Self-evaluation | 0.74    | -0.11   |
| of life         |         |         |
| satisfaction    |         |         |
| Perceived social| 0.16    | 0.09    |
| network support |         |         |
| Perception of   | 0.07    | 0.50    |
| corruption      |         |         |
| Active Physicians| 0.36    | 0.55    |
| Rate (physicians|         |         |
| for 1000        |         |         |
| population)     |         |         |
| Life Expectancy | 0.57    | 0.41    |
| at Birth        |         |         |
| Number of rooms | 0.20    | 0.83    |
| per person      |         |         |

### Table 5 Demographic Factor Analysis Results Mexico

| Factor | Factor1 | Factor2 | Factor3 |
|--------|---------|---------|---------|
| Inter-regional | 0.86 | -0.13 | -0.49 |
| migration rate |         |         |         |
| (Inter-regional | -0.17 | -0.68 | 0.19   |
| net flows migration |         |         |         |
| rate, (Population | -0.64 | -0.47 | 0.39   |
| density growth |         |         |         |
| Population density |         |         |         |
| (pop. per km2)  |         |         |         |
| Percentage of Old| 0.74  | 0.03   | 0.39   |
| Population Group |         |         |         |
| (65+)            |         |         |         |
| Percentage of Youth| -0.03 | 0.99   | 0.07   |
| Population Group |         |         |         |
| (0-14)           |         |         |         |

### Table 6 Final Regression Models US/MEX

| GDP Model | SE Model | Dem Model | Clim Model | Full Model |
|-----------|----------|-----------|------------|------------|
| Intercept | -14.65*** | -11.93*** | -13.01**   | -26.01***  |
|           | (0.56)   | (2.47)    | (4.15)     | (3.78)     |
| Regional GDP | -0.00    | 0.04      |            | (6.47)     |
|          | (0.03)   |           |            |            |
| Socio economic index 1 | -0.10* | -0.16** |            |            |
|          | (0.04)   | (0.05)    |            |            |
| Socio economic index 2 | 0.02 | -0.06 |            |            |
|          | (0.04) |           |            |            |
| Demographic index 1 | 0.04* | 0.03 |            |            |
|          | (0.02) | (0.05)    |            |            |
| Demographic index 2 | -0.03 | -0.06 |            |            |
|          | (0.05) | (0.05)    |            |            |
| Demographic index norm 3 | -0.07 | 0.01 |            |            |
|          | (0.04) | (0.04)    |            |            |
| Mean temperature of warmest quarter | 0.45** | 0.54*** |            |            |
|          | (0.14) | (0.15)    |            |            |
| Precipitation of warmest quarter | -0.43 | -0.15 |            |            |
|          | (0.27) | (0.28)    |            |            |
| Year      | *        | *         | *          | *          |
| Regions   |          |          |            |            |
| 31.62**** | (38.00) | 29.99*** | (38.00)    | 30.51***   |
|          | (38.00) | (38.00)  | (38.00)    | (38.00)    |
| AIC       | 2615.68  | 2615.31  | 2612.86    | 2618.69    |
| BIC       | 2791.09  | 2791.80  | 2792.07    | 2792.47    |
| Log Likelihood | -1262.41 | -1261.94 | -1260.01   | -1264.33   |
| Deviance  | 235.76   | 235.73   | 234.54     | 237.23     |
| Deviance explained | 0.62 | 0.62 | 0.62 | 0.61 |
| Dispersion | 1.00 | 1.00 | 1.00 | 1.00 |
| R²        | -3.87    | -4.04    | -2.86      | -1.92      |
| GCV score | 1317.72  | 1316.01  | 1319.25    | 1310.70    |
| Num. obs. | 351      | 351      | 351        | 351        |
| Num. smooth terms | 1 | 1 | 1 | 1 |

* **p < 0.001, *p < 0.01, *p < 0.05
Table 7 Final Regression Models Mexico

|                          | GDP Model | SE Model | Dem Model | Clim Model | Full Model |
|--------------------------|-----------|----------|-----------|------------|------------|
| Intercept                | -7.57***  | -6.00*** | -8.03***  | -19.31***  | -13.78**   |
|                          | (0.26)    | (1.10)   | (1.62)    | (2.06)     | (5.18)     |
| Regional GDP             | -0.00     |          | -6.00     | -8.03      | -13.78     |
|                          | (0.01)    |          | (0.01)    | (0.01)     | (0.01)     |
| Socio economic index 1   | -0.01     |          | -0.04*    |            |            |
|                          | (0.01)    |          | (0.02)    |            |            |
| Socio economic index 2   | -0.03     |          | -0.03     |            |            |
|                          | (0.02)    |          | (0.03)    |            |            |
| Demographic index 1      | 0.00      |          | 0.01      |            |            |
|                          | (0.01)    |          | (0.02)    |            |            |
| Demographic index 2      | -0.03     |          | -0.03     |            |            |
|                          | (0.03)    |          | (0.03)    |            |            |
| Demographic index norm 3 | -0.01     |          | -0.03*    |            |            |
|                          | (0.01)    |          | (0.01)    |            |            |
| Mean temperature of warmest quarter | 0.45*** | 0.48*** |
|                          | (0.08)    | (0.09)   |          |            |            |
| Precipitation of warmest quarter | 0.07 | 0.01 |
|                          | (0.17)    | (0.17)   |          |            |            |
| Year                     |           |          |           |           |            |
|                          | (0.19)    | (0.20)   | (0.21)    | (0.23)     | (0.26)     |
| Regions                  | 27.70***  | 26.78***  | 26.06***  | 25.53***   | 21.77***   |
|                          | (30.00)   | (31.00)  | (31.00)   | (31.00)    | (31.00)    |
| AIC                      | 1587.14   | 1587.12  | 1588.66   | 1582.70    | 1578.34    |
| BIC                      | 1672.01   | 1672.26  | 1672.35   | 1664.96    | 1663.71    |
| Log Likelihood           | -760.47   | -760.36  | -759.92   | -759.28    | -755.88    |
| Deviance                 | 73.07     | 72.99    | 72.72     | 73.71      | 72.14      |
| Deviance explained       | 0.81      | 0.81     | 0.81      | 0.81       | 0.83       |
| Dispersion               | 1.00      | 1.00     | 1.00      | 1.00       | 1.00       |
| R²                       | 0.56      | 0.55     | 0.56      | 0.46       | 0.43       |
| GCV score                | 825.14    | 826.78   | 830.25    | 811.98     | 824.00     |
| Num. obs.                | 96        | 96       | 96        | 96         | 96         |
| Num. smooth terms        | 1         | 1        | 1         | 1          | 1          |

*** p < 0.001, ** p < 0.01, * p < 0.05
6.0.1 Vector and Environmental Relationships: regression analysis methods

Species distribution modelling methods to estimate population at risk. To assess the relationships between vector and environment, techniques were adapted from [53]. Since our species distribution data set only consisted of presence data, absence data was substituted with background data. Background data points were sampled randomly from the study area (Mex/US). Using background data allows us to characterize environments in the study region, which establishes the environmental domain of the study, whilst presence data should represent the conditions a species is more likely to be present than on average. Since vector presence is indicated by a binary variable equal to 1 and absence equal to 0, a relationship between environmental and vector presence was estimated with a logit model by maximum likelihood. Logit estimation techniques are based on the assumption that there is a latent variable $y$ and that this latent variable is a linear function of all the explanatory variables. Climate data for the species distribution prediction modelling were sourced from MERRAclim [47]. This data set was built using 2 m air temperature (Kelvin degrees) and 2 m specific humidity (kg of water/kg of air) hourly data derived from satellite observations from the Modern Era Retrospective Analysis for Research and Applications Reanalysis. Tables S1 and S2 provide summary statistics for these data sets.

In order to predict if a vector was present in a given location $i$, the following equations were used:

$$Pr(Aedes.aegypti = 1) = \beta_1 Tvar_1 + \beta_2 Tvar_2 + \beta_3 Pvar_1 + \beta_4 Pvar_2 + \epsilon$$

$$Pr(Aedes.albopictus = 1) = \beta_1 Tvar_1 + \beta_2 Tvar_2 + \beta_3 Pvar_1 + \beta_4 Pvar_2 + \epsilon$$

where $X_i$ is a vector of regressors/independent variables (in this case: $Tvar_1$ represents temperature annual range; $Tvar_2$ represents mean temperature of the coldest quarter; $Pvar_1$ represents precipitation of the driest quarter; and $Pvar_2$ represents precipitation of the warmest quarter) in each location and $\epsilon$ is the error term.

6.0.2 Vector and Environmental Relationships: regression analysis results

The impact of climate on the probability of a vector being present was assessed by running a logit regression. Results are reported in Table S3.

The strongest predictors of $A. aegypti$ are the mean temperature of coldest quarter", "temperature of annual range", and "precipitation of the driest quarter". The strongest predictors for the presence of $A. albopictus$ are "precipitation of warmest quarter" and "temperature annual range".

The results from the logistic regressions the study confirmed a positive and highly significant association between some climatic factors and vector presence; results were consistent with previous studies [28,45,55]:

$A. albopictus$ seems to be distributed in environments that are warmer and wetter during the hotter months, $A. aegypti$ seems to be sensitive to colder temperatures and temperature range.

For both species minimum temperature is a major limiting factor affecting distribution and this is conclusive with our results. We could expect that, as minimum winter temperatures rise due to climate change, this will benefit both $Aedes$ species allowing them to exploit new habitats, and in turn, potentially increasing the distribution of dengue.

| Statistic                  | N   | Min  | Max  | Mean  | St. Dev |
|----------------------------|-----|------|------|-------|---------|
| Temp.Annual.Range           | 2,019 | 10.600 | 55.400 | 40.789 | 6.780   |
| Mean.Temp.Coldest.Quarter   | 2,019 | −1.100 | 29.000 | 10.994 | 5.230   |
| Precip.Driest Quarter      | 2,019 | 420.000 | 1,985.000 | 1,397.858 | 287.921 |
| Precip.Warmest.Quarter     | 2,019 | 1,055.000 | 2,440.000 | 2,070.442 | 281.595 |

Table S1  $Aedes$ aegypti climate variable summary

| Statistic                  | N   | Min  | Max  | Mean  | St. Dev |
|----------------------------|-----|------|------|-------|---------|
| Temp.Annual.Range           | 1,825 | 13.400 | 58.800 | 40.789 | 6.705   |
| Mean.Temp.Coldest.Quarter   | 1,825 | −1.100 | 29.000 | 10.994 | 5.230   |
| Precip.Driest Quarter      | 1,825 | 465.000 | 1,985.000 | 1,246.870 | 247.645 |
| Precip.Warmest.Quarter     | 1,825 | 1,062.000 | 2,407.000 | 2,070.442 | 178.508 |

Table S2  $Aedes$ albopictus climate variable summary
### Table S3  
*Aedes* Climate Regression Analysis

|                  | A. *aegypti* (1) | *pb* | A. *albopictus* (2) |
|------------------|------------------|------|---------------------|
| Mean temperature of coldest quarter | $0.016^{***}$ (0.002) | $-0.008^{***}$ (0.002) |
| Precipitation of driest quarter | $0.004^{***}$ (0.001) | $0.004^{***}$ (0.001) |
| Precipitation of warmest quarter | $-0.002^{***}$ (0.0004) | $0.002^{***}$ (0.0004) |
| Temperature annual range | $-0.002$ (0.002) | $-0.001$ (0.002) |
| Constant         | $-0.968$ (1.185) | $-5.712^{***}$ (1.064) |

**Note:**  
* p<0.1; ** p<0.05; *** p<0.01
Additional file 2
Final data-set US/Mex

Additional file 3
Final data-set Mex