Predicting dengue fever outbreaks: a review of epidemiological models and early warning systems

Vanessa Racloz¹#: v.racloz@uq.edu.au
Rebecca Ramsey²#: rebecca.ramsey@qut.edu.au
Shilu Tong²: s.tong@qut.edu.au
Wenbiao Hu¹*: w.hu@sph.uq.edu.au

¹ University of Queensland, School of population Health, Herston Road, Brisbane QLD, Australia
² School of Public Health and Institute of Health and biomedical Innovation, Queensland University of Technology, Kelvin Grove Campus, Victoria Park Road, Kelvin Grove, QLD 4059 Australia

*Corresponding author
# Joint first authors

Key words: dengue fever, modelling, risk factors, early warning, climate change

Abstract

Dengue fever is one of the world’s most important vector-borne diseases. The transmission area of this disease continues to expand due to many factors including urban sprawl, increased travel and global warming. Current preventative techniques are primarily based on controlling mosquito vectors as other prophylactic measures, such as a tetravalent vaccine are unlikely to be available in the foreseeable future. However, the continually increasing dengue incidence suggests that this strategy alone is not sufficient. Epidemiological models attempt to predict future outbreaks using information on the risk factors of the disease. Through a systematic literature review, this paper aims at analyzing the different modeling methods and their outputs in terms of accurately predicting disease outbreaks. We found that many previous studies have not sufficiently accounted for the spatio-temporal features of the disease in the modeling process. Yet with advances in technology, the ability to incorporate such information as well as the socio-environmental aspect allowed for its use as an early warning system, albeit limited geographically to a local scale.

learning points in the review

- Dengue fever as well as other infectious vector borne diseases continues to spread and cause disease, despite surveillance and control efforts
- Integrated mathematical and mapping models aid in the prevention and prediction process of early warning systems
- Models should include multi-factorial components including environmental, disease dependent and socio-economic parameters.
- Increased transparency through information sharing and collaborations are needed to help fight emerging infectious diseases
Introduction

Dengue fever (DF) is one of the most common widespread vector borne diseases in the world [1,2,3,4]. There are currently 2.5 billion people living in areas at risk of DF transmission, with 100 million cases reported annually [5,6]. DF is a flaviviral disease caused by one of four serotypes of dengue virus (DEN 1-4) which are transmitted by mosquito vectors, in particular the peridomestic species Aedes aegypti [2,7], and Ae. albopictus, which has recently been expanding its geographic distribution as seen in several outbreaks [8]. Infection by one serotype will provide lifelong immunity to that particular strain but not to the remaining three [1,9]. Cross-strain infections are common and can have severe consequences, with extreme cases leading to death [10]. Over the past 40 years the incidence and geographic distribution of DF has increased in many countries, particularly in those with tropical and subtropical climates [6,11,12,13,14]. DF has strong spatial and temporal patterns which have been linked to climatic and environmental conditions [15]. Thus the inclusion of spatial and temporal data in analytic processes may potentially allow for the identification of DF characteristics linked to these parameters and have significant applications in the prevention and control of this disease. Additionally, as discussed in the Intergovernmental Panel on Climate Change report [16], with global temperatures likely to increase, it is predicted that the endemic range of DF will expand geographically [17,18,19,20,21]. Warmer temperatures will also allow for increased vector reproduction and activity and decreased incubation time of larvae, resulting in an increased capacity for producing offspring. Thus an increase in the transmission potential and prevalence of DF seems likely [18,22].
As with many infectious diseases, one of the success measures of a surveillance system depends on the ability to predict an imminent outbreak through an early warning system. The process of identifying a potential threat and targeting surveillance and control methods form part of an early warning system. Such an approach is categorized as a targeted surveillance system as opposed to random surveillance [23]. This is an important difference in order to increase the probability of detection of any first or repeated incursion of disease at the earliest time possible. The ability to create an early warning system through the combination of climate, environmental, host and vector based data through various processes such as mathematical modeling and Geographical Information System (GIS) mapping have been used in many ways to improve veterinary and public health surveillance systems [24]. The combination of different prediction, surveillance and control methods and the tools involved in each process present a great potential in the combat against a variety of disease as described in Eisen & Eisen 2011[25]. This paper aims at providing an insight into the current DF modelling processes and the implementation of their outputs reported in published studies.

Methods

Through a comprehensive literature review, major databases including Blackwell synergy, CSA Illumina, Web of Science, Academic Search Elite, CINAHL with full text, Georef, medline, Professional Development Collection, Informaworld, InformitSearch, Proquest, Springerlink and Wiley Interscience and Pubmed (http://www.ncbi.nlm.nih.gov/pubmed) were searched. The key words used in this literature search were Dengue, Dengue fever, climate change, Dengue haemorrhagic fever, Climate anomalies, Dengue fever and climate anomalies, Risk factors and dengue fever, Dengue fever and modelling, vector borne diseases, Dengue fever and Aedes aegypti, Dengue fever and Aedes aegypti, vector borne disease modelling, regression analysis, spatio-temporal models, infectious disease surveillance and early warning systems. Studies were included if the use of one or more epidemiological models were reported. During the initial search, studies were selected based on a review of titles and abstracts. Full studies were retrieved and reviewed for all relevant studies as seen in Figure 1.

In order to analyse the DF models, it was important to review the background information as well as the method used in output generation. Due to the difference in output objectives, biological factors, spatio-temporal parameters, geographical scales and mathematical equations used in more current models, the comparison of efficacy between models is complex. A synopsis of the different pathways and risk factors found in the literature review is shown in Figure 2.
Results

Although many articles were identified in the field of DF, those containing a current mathematical model in combination with a predictive application were selected. As mentioned, the models include different transmission mechanisms, clinical manifestation data, current disease and vector control methods, treatment options, risk factors of DF and the potential of developing into dengue hemorrhagic fever. The risk factors for developing DF included biological, human, vector, environmental, socio-demographic data as well as climate and parameters linked to climate change (Figure 2).
As seen in Table 1, different categories for the analysis of the models existed, such as spatial scale, data collection time frame, model type and finally the incorporated risk factors. Although the main countries in the study were Australia, Brazil, China, Cuba, India, Indonesia, Mexico, Puerto Rico, Singapore, Thailand and the USA, the actual spatial scale used in the models varied from community level to multi-country. Collection time points spanned from daily measures to biannual analysis. Although the mathematical basis of many of the models shared a common regression point, these varied from logistic, autoregressive, spatio-temporal or Poisson equations. Finally, one of the most encompassing and diverse parameters were the risk factors used in the dengue model creation such as temperature, precipitation, vegetation indices, wind velocity or even hygienic markers.
There were a number of different models capable of producing prediction equations for the transmission of dengue fever. The type of model selected was dependent on the type of data collected and the nature of the variables (Figure 3), and due to the subtle differences involved in each outbreak, no universal models existed for analysis and prediction.

Figure 3: Flow chart process for data incorporation in dengue fever outbreak modelling

Traditionally, the data usually consisted of serologic and environmental or socioeconomic variables. Recently, socio-environmental changes have been identified as important determinants in the transmission of DF, and spatial and temporal aspects of these changes have been increasingly incorporated into studies [26]. The inclusion of spatial data allows for the identification of spatial patterns of occurrence and the ability to identify areas at high risk of disease. The majority of previous studies in the past decade have implemented logistic or multiple regression models to identify possible risk factors. A drawback of these models is that they are not capable of accounting for autocorrelation in time-series data, which may limit the predictive capabilities of the resultant model.
In terms of choice of mathematical methods, the AutoRegressive Integrated Moving Average (ARIMA) and Seasonal AutoRegressive Integrated Moving Average (SARIMA) models, which have the ability to cope with stochastic dependence of consecutive data, have become well established in the commercial and industrial fields [27,28]. A DF study in Queensland, Australia used ARIMA modelling to examine the relationship between weather variables and the disease [29]. The implementation of SARIMA accounts for auto-correlations in time-series as well as seasonality, long-term trends and lags. Consequently, SARIMA has higher predictive capabilities than other models described above. However, this approach requires the input of a large amount of data meaning that SARIMA may not be suitable for studies with a small sample sizes. SARIMA is also based on the assumption of normality. For diseases that are rare or occur less frequently, the assumption of normality may not be met and thus SARIMA might not be an appropriate choice. To account for the possibility of non-normally distributed data, SARIMA could be combined with another ecological model such as a Poisson likelihood or Bayesian spatiotemporal model. Unfortunately a software package combining these approaches is not yet available. Further studies into combined ecological models should be undertaken in order to determine the socio-environmental impact of DF in a systematic way.

As seen in Table 1, Figures 2 and 3, there are several different data collection and analysis pathways used to model DF transmission and intervention strategies. In terms of using the outputs of these models, two main objectives were identified: the use of DF models as a retrospective and validating method, and as an early warning tool to predict potential epidemics. Retrospective models use data as a validation method as seen in [30] for DF in Peru where data from 1994 to 2006 was analysed, and validated in latter epidemics throughout the region by evaluating the degree of association with demographic and geographic variables. Such techniques also allow for intervention and control strategies to be tested firstly on a hypothetical level, and then applied in the field, as seen in Luz et al.,2011 [31]. In this study, epidemiological and economic assessments of different vector control strategies were tested in the city of Rio de Janeiro in Brazil in relation to DF.

Although many models discussed above include risk factors involving basic climate and household information, the calculations are mainly based on human and vector borne parameters. Through the advancement and access to technology, various software programs and improvements in database infrastructure allow for the use of multidimensional values to be included in models in order to progress from a purely applied mathematically based theme to more a dynamic one.

Discussion

As seen in the tables and the studies reviewed, there are a large number of environmentally related as well as disease based parameters which influence the intensity, frequency, location and spread of a DF outbreak. Several limitations exist when using models as predictive tools in DF outbreaks. One of the main limitations for such models as mentioned is the geographic restriction due to data sources, often meteorological stations which might affect the availability of data as well as the spatial applicability. In order to be less constraint on such static datasets, Fuller et al., 2009 [32], included vegetation indices data, as well as sea surface temperatures in relation to El Niño Southern Oscillation (ENSO). Using this model, the authors suggested that a DF outbreak could be predicted with a 40 week advance in Costa Rica. Secondly, the differences in input parameters vary due to both natural and artificial factors. Biologically, differences in egg survival time, extrinsic incubation periods, median of lag phase are all directly or indirectly affected by external factors including temperature, humidity or even the immune system on an individual level. [15]. The effects of socio-environmental factors on mosquito vectors and transmission of DF are often not immediate, which involves a lag time.
between exposure to a risk factor and the development of the disease. The inclusion of temporal data allows for the identification of lag times and patterns of transmission over time [15]. Spatial data alone cannot provide the analysis of the temporal kinetics of an outbreak whereas the use of temporal data does not allow for the identification of high risk areas [15]. Ideally, studies should include both spatial and temporal aspects in the analysis to maximise the ability of the resultant equation to predict future outbreaks. Through the literature review, several interesting parameters were indentified which affect the predictive ability of models, such as the range of transmission of DF being at temperatures of 18 - 33.2 °C, with females feeding more frequently when temperatures are higher [4,33]. Directly affecting the biology of the vector, temperature also plays an important role on pathogen replication, maturation and period of infectivity. Transmission was also higher in areas where two or more serotypes were found to circulate simultaneously [7]. On the clinical level, facts such as the range in viraemic phase of DF which lasts from 2 to 12 days [34] will affect the precision of the model output. On an artificial level, models can vary due to the choice of regression analysis, the choice of map in terms of digital charting such as raster or vector outputs, although current methods favour vector maps due to their more flexible nature. Within these also lie the choice of geometric factors such as point or polygonal data which in turn will affect the predictive power of the models. [35]

Although modelling studies promote the need for a DF vaccine [36], a suitable vaccine that accounts for all four serotypes of DF is yet to be developed, hence the most effective means of controlling DF is through prevention via vector control. However, many vector control programs deteriorate as the economic condition of most high risk countries is unfavourable [37]. The identification of areas most at risk of DF transmission is essential to ensure the most efficient and effective use of resources for the continuation of vector control and eradication programs. With the predicted socio-environmental changes brought by urbanisation, climate change and globalisation, the regions at risk of transmission along with the economic impact of DF are set to increase. The analysis of previous outbreaks of DF may provide a means of predicting future epidemics in order to establish early warning systems and allocate resources more efficiently [38,39,40].

The use of models as a prediction method or part of a surveillance system in terms of early warning have been done for other vector borne diseases such as malaria [41,42], Rift Valley Fever [43] and bluetongue virus [44], which after determining the basic transmission pattern in a mathematical model, could then apply climatic events to predict potential outbreaks through Geographical Information Systems (GIS). Such models have been created on a local scale to predict DF outbreaks based on climatic factors as seen in Brazil [45] which used thermal, hydroclimatic, wind, atmospheric pressure, and humidity data as well as in and Puerto Rico [46] where climatic water budget indicators were used to create an early warning system, with the latter study being able to predict a DF outbreak with a three week warning period.

Few studies have been able to collect the necessary amount of spatio and temporal data as well as epidemiological information to analyse the correlation between all these factors. Bayesian spatio-temporal modelling takes into account the effect of covariates and correlations as well as being able to correct for possible errors arising from median estimates of random effects as seen in Yang et al., 2005 [47] for schistosomiasis, whereby conditional autoregressive models (CAR) were used in the Bayesian smoothing process. Another study addressing the advantages of this modelling technique is seen for dengue in Brazil [48], but as mentioned, data constraints, in this case the lack of socio-economic and meteorological covariates affect the predictive power of the model. Through the analysis of various dengue models and the ability to include varying levels of qualitative and quantitative data, the CAR method seems to have the most potential for developing a robust climate-based epidemic forecasting model.

The identification of high risk areas and trigger factors such as humidity, precipitation, temperature or even travel related disease could allow for early implementation of such
interventions so that DF can be effectively and efficiently controlled and prevented. Through the modelling, eventual intervention strategies have been analysed such as the effect of vaccination and the reduction in the number of susceptible individuals [14,49]. Although this review primarily focuses on mathematical models, the implementation of tools which use a combination of various climatic, environmental, epidemiological and socio economic factors to create an early warning system are showing encouraging results, as seen in the Chinese Infectious Disease Automated alert and Response System (CIDARS), which uses a combination of a fixed threshold, spatial and temporal detection methods for real time warning of many infectious diseases on a national scale[50]. Information gathering and sharing platforms [51] as seen in the multi-disease data management system interface are promising tools for infectious disease surveillance.

Conclusion

Interventions based on early warning systems aimed at preventing DF transmission require significant financial resources and human input, thus it is desirable to target areas and populations at high risk of DF. Modelling processes have shown their potential in identifying such high risk areas. The authors encourage the collection of information on both a spatial and temporal level, along with climatic and socio-environmental variables during future outbreaks of DF, as this will allow for the development of models with maximum predictive capabilities. Multiple and logistic regression models are most often used for analyses, yet as mentioned they are limited due to their inability at accounting for possible confounding factors, auto-correlations, trends and lags in a sufficient manner thus limiting their predictive performance. Recently, the use of spatial and temporal data has enhanced the ability of models to predict outbreaks of DF by allowing for the spatial identification of high risk areas whilst taking into account the temporal kinetics of DF transmission [15].

Certain factors will have to be taken into consideration when modelling DF in light of climate change and travel trends as well as vector habitat alterations. Due to the emerging spread of *Aedes albopictus*[52], models will have to be able to accommodate for the slightly different biology of these mosquitoes, as seen in the spatial modeling using socio environmental indicators in Brazil which had different breteau indices for both *Ae. aegypti* and *Ae. albopictus*[53]. Similarly to other vector borne disease, models vary in their complexity, methodology and area of study which can be very specific and not easily applied to other geographical areas, hence the comparison of less traditional mathematical techniques is more problematic. Transparency is a key factor which will allow for the improved accuracy and performance of models, not only for DF but for many other vector borne diseases which have complex transmission cycles.

Acknowledgments

This study was supported by Grant of the National Health and Medical Research Council, Australia (No. 1002608) and the Grants of Queensland University of Technology (2008BAI56B02, 2009ZX10004-201). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure statement

There were no conflicts of interest involved in this study.

Legend for Figures

**Figure 1:** Graphical summary of the literature search process.
**Figure 2:** Transmission pathway and risk factors involved in dengue fever outbreaks.
**Figure 3:** Flow chart process for data incorporation in dengue fever outbreak modelling.
Tables

Table 1: Setting and parameters used in predictive dengue model creation

<table>
<thead>
<tr>
<th>Spatial scale</th>
<th>Collection time frame</th>
<th>Model</th>
<th>Risk factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community</td>
<td>Daily</td>
<td>Poisson</td>
<td>Temperature</td>
</tr>
<tr>
<td>Parish</td>
<td>Weekly</td>
<td>Time-series</td>
<td>Precipitation</td>
</tr>
<tr>
<td>District</td>
<td>Monthly</td>
<td>Autoregressive</td>
<td>Wind velocity</td>
</tr>
<tr>
<td>Municipality</td>
<td>Bi-monthly</td>
<td>Multiple regression</td>
<td>Sea surface temperature</td>
</tr>
<tr>
<td>Province</td>
<td>Annually</td>
<td>Logistic regression</td>
<td>Humidity</td>
</tr>
<tr>
<td>City</td>
<td>Bi-annually</td>
<td>Autoregressive Integrated Moving Average (ARIMA)</td>
<td>Geographical settings</td>
</tr>
<tr>
<td>State</td>
<td></td>
<td>Step-wise regression</td>
<td>Hygienic parameters</td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td>Classification &amp; Regression Tree (CART)</td>
<td>Average pan evapotranspiration (APET)</td>
</tr>
<tr>
<td>Multi- country</td>
<td></td>
<td>Spatio-temporal regression</td>
<td>Proximity to potential artificial breeding sights</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vegetation dynamics</td>
</tr>
</tbody>
</table>
Predicting dengue fever outbreaks

References


Predicting dengue fever outbreaks


