Temporal Modeling of Dengue Fever: A Comprehensive Literature Review

ASWI[1,2,a*], SUSANNA Cramb[1,3,b], GENTRY White[1,c], WENBIAO Hu[4,d] and KERRIE Mengersen[1,e]

[1]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

[2]Universitas Negeri Makassar, Indonesia

[3]Cancer Council Queensland, Brisbane, Australia

[4]School of Public Health and Social Work, Queensland University of Technology, Australia

[a]aswi@hdr.qut.edu.au, [b]SusannaCramb@cancerqld.org.au, [c]gentry.white@qut.edu.au, [d]w2.hu@qut.edu.au, [e]k.mengersen@qut.edu.au [*]aswi@hdr.qut.edu.au

Abstract. Dengue fever has become a major public health problem in several countries. This paper aims to review and compare a number of temporal modeling approaches that have been proposed for predicting or forecasting the occurrence of dengue fever. This review also examines influential covariates considered in these studies. A comprehensive literature search was undertaken in September 2018, using Medline (via Ebscohost), ProQuest, Scopus, and Web of Science electronic databases. The search was confined to articles in English, published in refereed journals between January 2000 and September 2018. The most popular approach to temporal modeling of dengue was found to be an autoregressive integrated moving average (ARIMA) model. A limited number of studies applied Bayesian hierarchical dynamic generalized linear models. Climatic variables were most commonly associated with dengue incidence for temporal modeling.

Introduction

Dengue fever (DF) is a major significant vector-borne diseases world-wide [1]. Several statistical models for investigating the relationship between outbreak of DF and climatic variables have been proposed. These models may focus on analyzing and modeling factors affecting the current number of dengue cases, or predicting future numbers of dengue fever cases. The objectives of this comprehensive review were to review published temporal models that have been applied to dengue fever, and to compare analytical methods.

Methods

Databases and Search terms. The methodology for the review included a search strategy, with inclusion and exclusion criteria following the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines. Medline (via Ebscohost), ProQuest, Scopus and Web of Science databases were searched electronically in September 2018. The comprehensive literature search was confined to refereed journal articles in English only published between January 2000 and September 2018. The keywords used for searching databases were kept consistent: dengue and time series; dengue and temporal; dengue and temporal and model. All results within the databases were combined first and any duplicates were deleted using EndNote. The titles and abstracts of articles were screened first and remaining papers were investigated by reading the full text. The final set of papers was selected based on the inclusion criteria.

Inclusion and exclusion criteria. Articles were included if they used temporal or time series models to model dengue fever and were published in a peer-reviewed journal. There were no geographical restrictions, but the articles had to be published in English. Models that were not applied to dengue fever, non-temporal models, review papers and conference proceedings were excluded.

Data extraction. The data were collected and extracted manually. The retained data included first author name, dengue data, year of publication, objectives, covariate data, analytical method and key findings.

Results

Literature search. Figure 1 displays a flow chart of the literature search. Based on keyword searches, 648 papers from Scopus, 219 from Medline (via Ebscohost), 440 papers from Web of Science, and 186 papers from ProQuest were obtained. From 1493 citations initially identified, there were 827 potentially relevant articles left after duplicates were removed (666 duplicate articles). An additional 692 articles were removed after screening of titles and abstract. A further 45 of the remining article were excluded after reviewing the full article because they did not meet the inclusion criteria. 90 papers were finally selected for inclusion in the review.
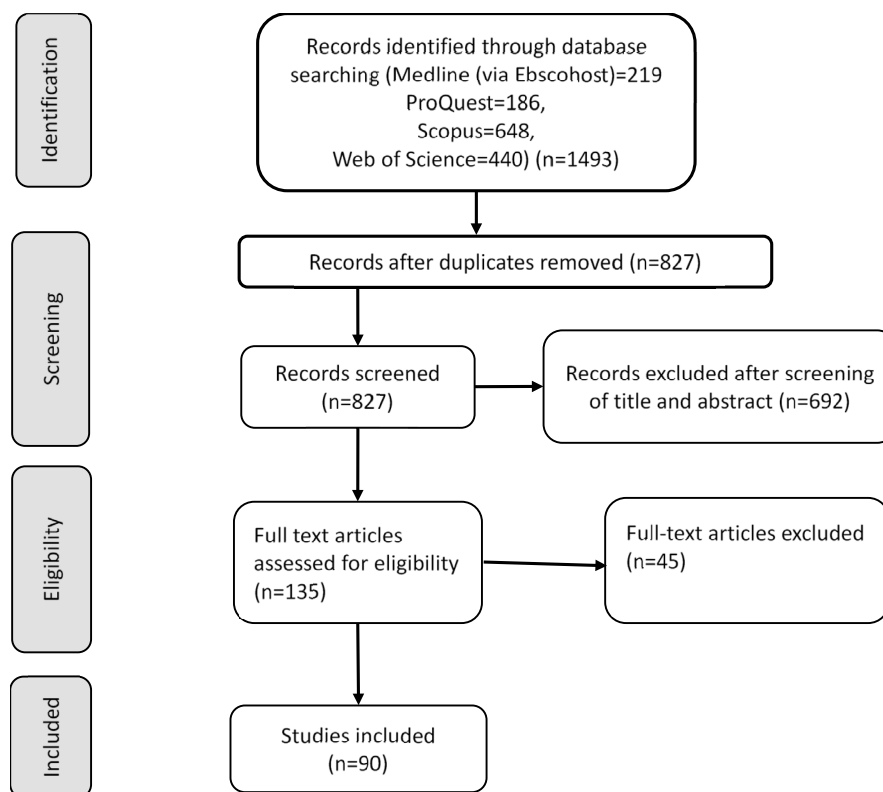

Figure 1. Flow chart of literature search

Dengue Data Geography. There were 13 studies in Brazil and China respectively, 12 studies in Thailand, eight studies in Singapore, seven studies in Sri Lanka, six studies in Taiwan, four studies in Colombia, three studies in each of Indonesia, Vietnam, Bangladesh, and Mexico, two studies in each of Pakistan, India, Philippines, Malaysia, and Australia, while other areas (Saudi Arabia, France, Curacao and Venezuela) had one study each. One study covered three countries (Puerto Rico, Mexico, and Thailand).

Year of Publication. The highest number of studies were published in 2016 (13 studies), with 12 studies published in 2015, 11 studies in each of 2017 and 2018, nine studies in 2012, eight studies in 2013, six studies in each of 2011 and 2014, four studies in 2009, three studies in 2017, two studies in each of 2008 and 2006. One study was published in each of 2004, 2005, and 2010. No included study was published during 2000 to 2003.

Analytical Method

A variety of temporal approaches were used in modeling dengue fever. These models are explained as follows.

ARIMA/SARIMA/ARIMAX. Out of 90 studies, 26 studies used the Box-Jenkins methodology [2], non-seasonal autoregressive (AR) integrated (I) moving average (MA) or (ARIMA), seasonal ARIMA (SARIMA), and ARIMAX to forecast dengue incidence. The ARIMA model is one of the most powerful approaches for forecasting which has four steps: model identification, estimation, diagnostic checking, and forecasting [2]. The SARIMA model includes both seasonal and non-seasonal components and can be extended to incorporate exogenous input (SARIMA-X). Brunkard et al. [3] assessed the relationships between long-term El Niño Southern Oscillation (ENSO)-related weather and microclimate on dengue incidence by using ARIMA with exogenous input (ARIMAX). An ARIMAX model is a linear regression (X) with the residual term specified by AR or MA terms [3]. Stationarity is a fundamental property in time series models. However, not all time series are stationary.

Wavelet Analysis (WA) and Cross Wavelets. The association between diseases and climate can vary over time (including the mean and variance, i.e. non-stationary). In this case, a specialized type of time series analysis, wavelet analysis, is a more appropriate tool to detect periodicity and more powerful for longer time series. Thirteen studies used wavelet time series in modeling dengue fever disease.

Vincenti et al. [4] applied WA to investigate the effect of ENSO on dengue outbreaks in Venezuela and found that there was a significant effect of ENSO and local climatic changes at both seasonal and inter-annual scale on dengue cycles. Wavelet analysis can be employed to disclose insights into a non-stationary time series [5]. The cross wavelet revealed that rainfall was the most important factor for dengue cases [5]. Ehelepola et al. [6] investigated the relationship between diurnal temperature ranges (DTRs) and dengue incidence in Colombo using WA and found that dengue transmission was negatively correlated with large DTRs.

Time frequency Analysis. Another approach that is appropriate for non-stationary and/or nonlinear time series data is time-frequency analysis. Various approaches to time frequency analysis include Fourier transform (FT), Wigner distribution function, short time Fourier transform (STFT), Hilbert-Huang Transform (HHT), and continuous wavelet transform (WT) [7]. Tsai et al. [7] compared the most two well-known types of time frequency analysis, HHT and WT in order to investigate the relationship between DF incidence and hydrologic and climatic factors. The WT is an effective approach for analyzing the time series frequency distribution of non-stationary data, but it struggles to capture the nonlinear features, while HHT can capture the non-linear features. The HHT can also provide sharper time-frequency distribution. Using these approaches, dengue fever was strongly correlated with the occurrence of El Niño weather pattern.

Distributed Lag Non-linear Model (DLNM). A DLNM model can be used to explain the effect on factors simultaneously through both non-linear and delayed effects, originally proposed by Gasparrini et al. [8]. Six studies have used DLNM to investigate the correlation between climatic variables and dengue incidence. DLNM combined with a quasi-Poisson regression have also been used [9, 10]. Xu et al. [10] used a Poisson regression combined with DLNM to investigate the effect of Absolute Humidity (AH) on weekly bases and other climatic factors. They found that AH had better predictive capacity for modeling dengue incidence.

An Integer-valued Autoregressive conditional heteroscedastic (INARCH(p)). Temporal or time series count data often have overdispersion present, which may cause difficulties in modeling. For handling overdispersion, an integer-valued generalized ARCH (INGARCH) model has been used. INARCH(p) or INGARCH (0, p) are also called the purely autoregressive model. Xu et al. [11] have proposed a novel model for handling the conditional equidispersion, under dispersion and overdispersion in integer-value pure time series. In their model, the number of weekly dengue cases was the response variable and covariates were both climate variables and lagged terms. Conditional overdispersion was present in their dengue case data. A new model, DINARCH (a new dispersed

INARCH model) can handle this overdispersion. They found that their model performed better than the INARCH(p) model [11].

Markov Prediction Model (MPM). MPM is a stochastic model that assumes that the probability of each event depends only on the preceding event. Ahdika et al. [12] compared two models, MPM and an INAR(1)-Poisson model for predicting counts of dengue hemorrhagic fever (DHF) cases, and concluded that MPM was the superior model.

Spearman's rank correlation coefficient (SRCC). Wang et al. [13] investigated the correlation between dengue incidence and some climatic variables as well as the delayed effects by using SRCC. The lags of time between climatic variables and dengue incidence were calculated by using time-lag correlation coefficients. They found that dengue incidence and weekly mean temperate were strongly correlated at a 17-18 weeks lag time in 2005. They also found that both the short and long-term time series data can give complementary perspectives on the association between climatic variables and dengue incidence in Singapore. Wongkoon et al. [14] used a time-series adjusted Poisson regression and Spearman correlation analysis to investigate the impact of weather on the dengue transmission in Thailand. They concluded that temperature (both minimum and maximum) at lag zero months, relative humidity and total rainfall at lag two months respectively were important influences on dengue incidence.

Standard multiple regression. Phung et al. [15] compared three different regression models to predict dengue incidence in Can Tho city, Vietnam and to investigate the relationships between dengue incidence and climatic factors. These models were a standard multiple regression model (SMR), Poisson distributed lag model (PDLM) and SARIMA. The results showed that standard multiple regression lacked accuracy in predicting dengue incidence, while SARIMA and PDLM models had the best forecasting of dengue incidence for a 3-month period and for a 6, 9, and 12-month period, respectively. They further found that dengue incidence is significantly correlated with temperature and relative humidity but not cumulative rainfall. Withanage et al. [16] studied how to develop and validate an accuracy forecasting model by applying three time-series regression using various lag times of number of rainy days, monthly rainfall, temperature (both minimum and maximum), wind speed, relative humidity (both minimum and maximum) and retrospective dengue incidence on a monthly basis. The results showed that the current month's dengue incidence was related to the preceding month's dengue cases.

Time series regression models

Generalized linear model (GLM) with Poisson, quasi-Poisson and Negative Binomial distribution. Time series regression models are very common approach in environmental epidemiology studies. Commonly this is a GLM with a Poisson distribution, also well known as Poisson regression [17]. Eleven studies used Poisson regression. Earnest et al. [18] determined the relationship between climatic data and dengue incidence after considering lag periods. Hii et al. [19] used time series Poisson regression with the inclusion of time trend, lag of dengue cases as AR terms, lagged terms of weather predictors and modelled the predictors as smooth cubic spline functions. They concluded that dengue incidence rose linearly at time lags of 5-20 and 5-16 weeks with increases in precipitation and temperature, respectively. Among the studies, seven studies used other different distribution models, i.e. quasi-Poisson (2 studies) [9, 10] and negative binomial (5 studies).

Generalized Additive Model (GAM) with Poisson distribution. GAM is a non-parametric regression approach with multiple predictors that applies a "smoothing function" [20]. Two studies used Poisson GAM [21, 22]. Cheong et al. [21] used Poisson GAM with a natural cubic spline for estimating climatic effects on dengue while accounting for a non-linear temporal effect.

Time series regression tree. Four studies used time series regression tree models [23-26] to identify the possible threshold for predicting dengue fever outbreak. Time-series classification and regression tree (CART) models allow a flexible nonparametric approach for predicting the association between dependent and independent variables [25]. Liu et al. [25] found that mosquito density, diurnal temperature fluctuation, and timeliness of diagnosis play critical roles in dengue fever

outbreaks. Gu et al. [26] investigated the delayed effect of climatic factor and the risk of dengue fever by BRT (boosted regression tree).

Random Forest. Carvajal et al. [20] compared four models: general additive model (GAM), random forest (RF), SARIMA, and gradient boosting (GB) with respect to their predictive accuracy of the time series pattern of dengue incidence. They included climatic factors (rainfall, Southern Oscillation Index, temperature, flood, wind direction, wind speed and relative humidity) into the models to determine the factors that highly correlated to dengue incidence. The results showed that an RF model with lagged relative humidity gave the most accurate predictions.

Reserved Moving Approximate Entropy (RMApEn). Chen et al. [27] developed a novel method, namely, the RMApEn algorithm on the time series for predicting dengue fever. They concluded that combining this algorithm with the pattern recognition technique is appealing for predicting outbreak of dengue fever.

Bayesian Paradigm. Within the Bayesian paradigm, model parameters are considered to arise from a distribution. The key advantages of the Bayesian approach include determining uncertainties in the estimations and predictions, and straightforward incorporation of time series structure as a prior [28].

Bayesian Neural Network Autoregressive (NN(AR)). The Bayesian neural network autoregressive (NN(AR)) temporal models were introduced by Koh et al. [29] for predicting dengue fever incidence in Singapore by incorporating rainfall data. They compared NN (AR (2)) and AR (2) models on a dengue dataset and found that NN (AR (2)) provided more accurate forecasts with fewer data points, while AR (2) was preferred for longer data points.

Bayesian hierarchical dynamic Poisson generalized linear models (DGLMs). Martinez et al. [30] applied Bayesian hierarchical DGLMs to model the relationship between climatic variables (rainfall, temperature, solar radiation and relative humidity) and dengue counts on weekly bases in Colombia. Calendar effects using constant coefficients, RW1, or RW2 time-varying coefficients (TVCs) and climatic variables (using a constant or RW1 time-varying coefficients) were included in their model. They found that the model which included a RW1 time-varying coefficient for both calendar trend and climatic variables performed the best. Also, DLGMs can effectively forecast the number of dengue cases for one or two weeks ahead. Schmidt [31] also used Bayesian generalized dynamic models to model weekly dengue fever by considering Poisson gamma, Poisson -log-normal and a zero inflated model. They found that the zero inflated model was worse than the other models. Poisson and Poisson log-normal models provided similar values in terms of the predictive likelihood.

Summary

A variety of temporal modeling approaches for dengue fever, with or without explanatory variables, have been reviewed. Most studies used Box-Jenkins methodology: ARIMA/SARIMA models. Other common approaches included wavelet analysis, time series Poisson regression analysis, a negative binomial regression model, and DLNM model. Only one study compared the predictive accuracy of temporal patterns. A limited number of studies applied Bayesian approaches. Only one study used each of the Bayesian hierarchical dynamic generalized linear models and Bayesian neural network time series. Climatic variables were most commonly associated with dengue incidence count data for temporal modeling. Despite the popularity of Bayesian approaches and their appeal in epidemiology, only a limited number of studies applied Bayesian approaches to modeling temporal dengue fever. To significantly enhance the current ability to predict the dynamic of dengue fever incidence, future models need to consider Bayesian approaches.

Acknowledgement

References

[1]   Racloz V, et al. Surveillance of Dengue Fever Virus: A Review of Epidemiological Models and Early Warning Systems (Early Warning Modeling in Dengue Fever). *PLoS Neglected Tropical Diseases* 2012; 6(5): e1648.

[2]   Box GEP. *Time series analysis forecasting and control*. 4th ed. ed. Hoboken, N.J.: John Wiley, 2008.

[3]   Brunkard 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): 227-234.

[4]   Vincenti-Gonzalez MF, et al. ENSO-driven climate variability promotes periodic major outbreaks of dengue in Venezuela. *Scientific Reports* 2018; 8(1): 5727-5727.

[5]   Wickramaarachchi WPTM, Perera SSN, Jayasinghe S. Modelling and analysis of dengue disease transmission in urban Colombo: A wavelets and cross wavelets approach. *Journal of the National Science Foundation of Sri Lanka* 2016; 43(4): 337-345.

[6]   Ehelepola NDB, Ariyaratne K. The correlation between dengue incidence and diurnal ranges of temperature of Colombo district, Sri Lanka 2005-2014. *Global Health Action* 2016; 9: 32267-32267.

[7]   Tsai CW, Yeh TG, Hsiao YR. Evaluation of hydrologic and meteorological impacts on dengue fever incidences in southern Taiwan using time-frequency analysis methods. *Ecological Informatics* 2018; 46: 166-178.

[8]   Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Statistics in Medicine* 2010; 29(21): 2224-2234.

[9]   Talagala T. Distributed lag nonlinear modelling approach to identify relationship between climatic factors and dengue incidence in Colombo district, Sri Lanka. *Epidemiology Biostatistics and Public Health* 2015; 12(4).

[10]   Xu HY, et al. Statistical Modeling Reveals the Effect of Absolute Humidity on Dengue in Singapore. *Plos Neglected Tropical Diseases* 2014; 8(5).

[11]   Xu HY, et al. A model for integer-valued time series with conditional overdispersion. *Computational Statistics and Data Analysis* 2012; 56(12): 4229-4242.

[12]   Ahdika A, Lusiyana N, Iop. Comparison of INAR(1)-Poisson model and Markov prediction model in forecasting the number of DHF patients in west java Indonesia. In: *7th International Conference on Applied Physics and Mathematics*, 2017. (Journal of Physics Conference Series).

[13]   Wang ZX, et al. Delayed Effects of Climate Variables on Incidence of Dengue in Singapore during 2000-2010. In: Dan Y, ed. *International Conference on Environmental Science and Development*, 2012: pp. 22-26. (APCBEE Procedia).

[14]   Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Weather factors influencing the occurrence of dengue fever in Nakhon Si Thammarat, Thailand. *Tropical Biomedicine* 2013; 30(4): 631-641.

[15]   Phung D, et al. Identification of the prediction model for dengue incidence in Can Tho city, a Mekong Delta area in Vietnam. *Acta Tropica* 2015; 141: 88-96.

[16]   Withanage GP, et al. A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. *Parasites & Vectors* 2018; 11.

[17]   Imai C, Hashizume M. A Systematic Review of Methodology: Time Series Regression Analysis for Environmental Factors and Infectious Diseases. *Tropical Medicine and Health* 2015; 43(1): 1-9.

[18]   Earnest A, Tan SB, Wilder-Smith A. Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections. *Epidemiology And Infection* 2012; 140(7): 1244-1251.

[19]   Hii YL, et al. Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Global Health Action* 2009; 2: 124-132.

[20]   Carvajal TM, et al. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infectious Diseases* 2018; 18(1): 183-183.

[21]   Cheong YL, et al. Assessing Weather Effects on Dengue Disease in Malaysia. *International Journal of Environmental Research and Public Health* 2013; 10(12): 6319-6334.

[22]   Li C, et al. Modeling and projection of dengue fever cases in Guangzhou based on variation of weather factors. *The Science Of The Total Environment* 2017; 605-606: 867-873.

[23]   Huang X, et al. A threshold analysis of dengue transmission in terms of weather variables and imported dengue cases in Australia. *Emerging Microbes & Infections* 2013; 2(12): e87-e87.

[24]   Liu K, et al. Using Baidu Search Index to Predict Dengue Outbreak in China. *Scientific Reports* 2016; 6: 38040-38040.

[25]   Liu KK, et al. Risk assessment of dengue fever in Zhongshan, China: a time-series regression tree analysis. *Epidemiology And Infection* 2017; 145(3): 451-461.

[26]   Gu HG, et al. Meteorological Factors for Dengue Fever Control and Prevention in South China. *International Journal of Environmental Research and Public Health* 2016; 13(9).

[27]   Chen CC, Chang HC**.** Predicting dengue outbreaks using approximate entropy algorithm and pattern recognition. *Journal of Infection* 2013; 67(1): 65-71.

[28]   Blangiardo M. *Spatial and spatio-temporal Bayesian models with R-INLA*. Chichester, West Sussex: John Wiley and Sons, Inc., 2015.

[29]   Koh YM, et al. A model comparison algorithm for increased forecast accuracy of dengue fever incidence in Singapore and the auxiliary role of total precipitation information. *International Journal of Environmental Health Research* 2018; 28(5): 535-552.

[30]   Martinez-Bello DA, Lopez-Quilez A, Torres-Prieto A**.** Bayesian dynamic modeling of time series of dengue disease case counts. *Plos Neglected Tropical Diseases* 2017; 11(7).

[31]   Schmidt AM, Pereira JBM**.** Modelling Time Series of Counts in Epidemiology. *International Statistical Review* 2011; 79(1): 48-69.