

# Mathematical and statistical models for the control of mosquito-borne diseases: the experience of Costa Rica

Paola Vásquez,<sup>1</sup> Fabio Sanchez,<sup>1</sup> Luis A. Barboza,<sup>1</sup> Yury E. García,<sup>2</sup> Juan G. Calvo,<sup>1</sup> Shu-Wei Chou-Chen,<sup>1</sup> and Gustavo Mery<sup>3</sup>

**Suggested citation** Vásquez P, Sanchez F, Barboza LA, García YE, Calvo JG, Chou-Chen S-W, et al. Mathematical and statistical models for the control of mosquito-borne diseases: the experience of Costa Rica. *Rev Panam Salud Publica.* 2022;46:e113. <https://doi.org/10.26633/RPSP.2022.113>

## ABSTRACT

**Objective.** To summarize the results of research conducted in Costa Rica in which mathematical and statistical methods were implemented to study the transmission dynamics of mosquito-borne diseases.

**Methods.** Three articles with mathematical and statistical analysis on vector-borne diseases in Costa Rica were selected and reviewed. These papers show the value and relevance of using different quantitative methods to understand disease dynamics and support decision-making.

**Results.** The results of these investigations: 1) show the impact on dengue case reports when a second pathogen emerges, such as chikungunya; 2) recover key parameters in Zika dynamics using Bayesian inference; and 3) show the use of machine learning algorithms and climatic variables to forecast the dengue relative risk in five different locations.

**Conclusions.** Mathematical and statistical modeling enables the description of mosquito-borne disease transmission dynamics, providing quantitative information to support prevention/control methods and resource allocation planning.

## Keywords

Vector-borne diseases; models, theoretical; public health; Costa Rica.

In the efforts to reduce the burden and threat of mosquito-borne pathogens, the implementation of timely and effective prevention and mitigation measures represents a challenge to public health authorities worldwide. The past decades have been especially challenging, as increased urbanization, population mobility, deforestation, climate change, and insecticide resistance have created conditions for pathogens and vectors to emerge in new areas and to reemerge in regions where successful vector control campaigns were previously implemented. This highlights the need to explore innovative and cost-effective tools to strengthen prevention and mitigation programs in a variety of different settings and circumstances.

Vector-borne diseases involve complex transmission dynamics, which typically requires an interdisciplinary approach to

achieve effective prevention and control measures. The capabilities and tools used by such teams need urgent enhancement, to effectively couple with newly recognized modes of transmission and environmental conditions (1).

Furthermore, policymakers have increasingly recognized mathematical and statistical modeling techniques as valuable tools for planning vector control interventions (2). These models provide a simplified representation of a complex system that allows us to understand how a specific pathogen spreads, test mitigation actions, evaluate the impact of control strategies on disease incidence, and provide projections of possible transmission trends to optimize the allocation of available resources (3). For their development, multiple modeling techniques and data have been used (4). Results have shown the potential of such

<sup>1</sup> Universidad de Costa Rica, San José, Costa Rica ✉ Paola Vásquez, paola.vasquez@ucr.ac.cr

<sup>2</sup> University of California Davis, Davis, California, United States of America

<sup>3</sup> Pan American Health Organization, San José, Costa Rica

models to emphasize critical factors for planning and implementation of public health interventions and for guiding public policy, helping administer human, financial, and vector-control supplies more efficiently (2).

All things considered, every model has limitations that range from those intrinsic to the model itself to the lack of available data or user expertise. Therefore, several challenges arise when attempting to use models for public health purposes. Nevertheless, the increased knowledge in mosquito population biology, advances in technology, and access to real-time epidemiological, demographic, socioeconomic, and climate data have opened up new possibilities for these models to reinvigorate disease control (1).

In Costa Rica, a country with optimal conditions for vector survival, mosquito-borne diseases have long been a burden for public health authorities. Since the introduction of the dengue virus in 1993 up to 2020, more than 392 000 dengue cases have been reported (5). In 2014, the first cases of chikungunya were reported; by February 2016, the first cases of Zika virus disease appeared on the Pacific coast; and since 2016, the local transmission of malaria has slowly resumed its upward trend (5). Although there has been collaboration between scholars and public health authorities, in Costa Rica, the implementation of modeling techniques as active tools to inform and guide prevention and control measures is still in the early stages. In this process, the willingness of health officials to introduce flexible, evidence-based models into vector control planning has created opportunities for further collaboration. Recent challenges posed by the COVID-19 pandemic have highlighted the potential and urgent need for utilizing modeling tools in infectious disease prevention and control.

This article highlights the potential uses of three modeling techniques when developed and tailored to study the transmission dynamic of mosquito-borne disease and does so by reviewing previous work developed in Costa Rica using classic compartmental epidemic models, the Bayesian approach, and more complex computational machine learning tools. This work looks to analyze and highlight the contributions that each analysis can provide to public health officials and the implications that the effective use of these technologies can have in recognizing opportunities to enhance vector control and guide resource mobilization.

## MATERIALS AND METHODS

The three methodologies reviewed, despite a range of technical and computational complexity, were key to initiating collaboration between scholars and public health officials in Costa Rica. All were developed using publicly available data from institutions such as the Ministry of Health, the National Meteorological Institute, the National Institute of Statistics and Census (INEC), and the National Oceanic and Atmospheric Administration (NOAA).

### Classical compartmental epidemic models

In Costa Rica, classical methods were used to analyze the impact in terms of cases detected that the introduction of chikungunya could have in a country where dengue is endemic. In 2014, with the introduction of the chikungunya virus, Costa Rican health authorities faced the circulation of two pathogens

transmitted by the same vector with similar clinical manifestations and temporal and spatial distribution, given that dengue virus has been endemic since 1993 (5). Under this scenario, researchers implemented a single-outbreak deterministic model and a genetic algorithm using weighted least squares to calculate point estimates of key model parameters and initial conditions for dengue and chikungunya using weekly reported cases from May 2015 through May 2016 in Costa Rica (6).

The key parameters estimated were the transmission rate, diagnosis rate, average vector infectious period, and initial value of the susceptible population. The primary assumption was that the reported number of cases at week  $k$  follows a Poisson distribution whose rate is determined as a function of the expected number of cases within the  $k$ -th week. Authors used a least-squares procedure based on the normalized differences between observed and expected weekly reported cases to fit the whole set of parameters and initial populations.

The study also included a classical epidemic compartmental model. This approach is based on dividing the population into different compartments according to epidemiological status. The compartments range from the most basic division that includes susceptible (S), infectious (I), and recovered (R) individuals to those with additional partitions and more complex interactions between classes. Sisson et al. (7) implemented a deterministic version with five compartments for the host: susceptible (S), exposed (E), undiagnosed (U), diagnosed (D), and recovered (R); and three compartments for the mosquito: susceptible (S), exposed (E), and infected (I). Basically, the infected class was divided into two classes—diagnosed and undiagnosed—to consider the asymptomatic behavior of the diseases.

### Approximate Bayesian computation

In Costa Rica, authors used a Bayesian approach to study Zika transmission dynamics and possible implications for public health policy. Approximate Bayesian computation (ABC) is a computational technique that has its roots in the Bayesian statistics approach of data analysis and parameter estimation (7). The parameter estimation results give a distribution that provides a range of possible values instead of a single number. By January 2016, the first cases of Zika were detected in Costa Rica, adding to the complexity of pathogens transmitted by the vector, the *Aedes* mosquito.

With the combination of the ABC approach, the study described by Sanchez et al. (8) modelled the overall dynamics of Zika during the 2016–2017 outbreak in Costa Rica using a single-outbreak mathematical model with sexual transmission and host availability for vector-feeding. The aim of this analysis was to estimate the key parameters that fit the model to data and compute the basic reproductive number ( $R_0$ ) distribution. The estimated parameters were the average mosquito lifespan, mosquito biting rate, per-capita diagnosis rate, and the proportion of hosts available (males and females) for mosquito feeding. Weekly reported cases of Zika were used for parameter estimation and provided by the Ministry of Health.

### Statistical learning

In Vásquez et al. (9), researchers used statistical learning methods to retrospectively predict the relative risk of dengue by using weekly dengue incidence and climate variables.

Data included nine years of weekly dengue reported cases (2007–2016) provided by the Ministry of Health; temperature, precipitation, and relative humidity data provided by the National Meteorological Institute from five climatologically diverse cantons across the country (Limón, Buenos Aires, Alajuela, Santa Cruz, and Liberia); and sea surface temperature anomaly from NOAA. The information was used to train two different machine learning approaches—Generalized Additive Model (GAM) and Random Forest—to retrospectively predict the relative risk of dengue in 2017 (9). We included the lag periods of each of the covariates in the predictive model. The maximum allowed lag was taken as 30 weeks.

Machine learning techniques represent a comprehensive set of algorithms that allows solving different tasks by learning from past transmission trends. Algorithms under this approach require an initial process of defining the model and the variables, but its advantages in reduced execution time and good degree of precision are considerable. In the case of GAM, the model requires a choice of hypotheses to generate a very flexible form of regression model that allows complex, non-linear covariate effects that are driven by the data.

For all the methods presented in this article, comparison metrics were used both in the variable selection processes and the comparison of results. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) scores were used for model selection in the chikungunya and dengue 2015–2016 single-outbreak deterministic model. Simulation methods, such as ABC, that guaranteed convergence toward an optimal model were implemented in the Zika research study. Machine learning algorithms utilized Normalized Root-Mean-Square Error (NRMSE).

## RESULTS

For the study of the dengue and chikungunya outbreak in 2015–2016, a linear regression model adjusted to the reported number of log-cases evidenced an exponential growth for dengue ( $R^2 = 0.9294$ ) but not for chikungunya ( $R^2 = 0.6365$ ) during the 12-month study period. Furthermore, by using classical deterministic compartmental models, the estimation of the parameter resulted in a dengue transmission rate similar to values previously reported in the literature. However, the results suggested that people were diagnosed at a faster rate with dengue, whereas people with chikungunya were diagnosed at a lower rate. The best-fit curves for dengue showed that the expected cases between weeks 20 and 30 should have been lower than those reported. On the contrary, the best fit curve for chikungunya, in the same period, projected more cases than those informed. These results suggested that a misdiagnosis is likely in dengue and chikungunya cases. This may be due to the similarity of the symptoms of the two diseases in their initial state and the fact that chikungunya was a new disease in the country.

The ABC approach allowed us to estimate the basic reproductive number distribution from the 2016–2017 Zika outbreak in Costa Rica. The model demonstrated that in order to have an  $R_0 < 1$ , host availability needed to be very low (below 3%); given the limited resources in the country, this is not a viable public health strategy to be implemented. However, a key finding of the model relied on the elasticity and sensitivity analysis on the  $R_0$  parameters, which demonstrate a higher

sensitivity to the female host availability than to the male host availability (8). In Costa Rica, according to the national census on use of time, women spend a larger proportion of their time in their homes, coinciding with the breeding and feeding habits of the *Aedes* mosquito, which disperses relatively short distances from its development sites, prefers to rest indoors, and feeds during daylight hours whenever the host is available.

The machine learning algorithms showed an adequate performance in estimating the relative risk of dengue during 2017, while using the nine-year period of 2007–2016 as training periods. The predictive capacity of the models was corroborated through the NRMSE. This method was determined using the relative risk predicted by the model and the relative risk observed for each canton, which allowed comparing the dispersion obtained from the prediction with respect to its average behavior. Detailed results of the NRMSE are described by Vásquez et al. (9). Both the GAM and the Random Forest presented adequate and similar performance in predicting relative risk in the different cantons. The prediction for Alajuela and Limón was quite accurate, where both models managed to capture the peaks as well as the decrease in incidence with an accuracy of no more than two weeks. Buenos Aires presented a low number of cases for most of 2017 with a sudden increase during November, which was captured by the models. For the cantons located in the North Pacific—Liberia and Santa Cruz—the prediction was not as accurate, as they overestimated the reported relative risk in most weeks. However, they could still capture the tendency of cases in selected weeks (9).

## DISCUSSION

With rapidly evolving demographic, environmental, and epidemiological conditions worldwide, public health systems must be prepared to detect and respond quickly and effectively to changes in the transmission dynamics of infectious disease, potential drivers, geographical distribution, and emergence of novel pathogens. The successful implementation of this response requires not only an evolution from traditional control approaches but also the introduction of novel tools that optimize the allocation of available resources (1).

This article summarizes research conducted in Costa Rica involving mathematical and statistical techniques to study mosquito-borne diseases. Using classical, deterministic compartmental models to fit the tendencies of dengue and chikungunya, we were able to warn public health authorities of potential issues with detection of cases, particularly over-reporting, which may have an impact on improving detection protocols. By analyzing the transmission dynamics of Zika using Bayesian statistical methods, we detected a higher risk of infection for individuals who spent more time at home, which can serve as a guide to health officials to reinforce bottom-up communication approaches at the community level. Public health campaigns targeting individuals that spend long hours at home could have translated in a greater impact of the recommended mosquito control strategies in those communities.

The use of machine learning algorithms and climate information showed that, by using historical epidemiological and meteorological data, the GAM and the Random Forest approach could be used as a tool in the effort to develop early

warning system models for Costa Rica and other countries in the region (9).

In this sense, having mathematical models as an additional tool in decision-making offers the opportunity to improve the implementation of strategies, program management, and optimization of available resources. Furthermore, the emergency experienced due to SARS-CoV-2 highlighted the value of and expanded opportunities for collaboration between public health authorities and academia to develop modeling tools and utilize them in decision-making for public health measures in Costa Rica. During 2020 and 2021, health authorities, in collaboration with academic institutions and the Pan American Health Organization, implemented a multilayer network model to study the transmission dynamics of SARS-CoV-2 and analyze hypothetical scenarios (10). The projections provided by this model, added to other inputs, helped make decisions in public policy and manage the sanitary emergency in the country. The structure of this model also offers the possibility to study the geographical distribution of vector-borne diseases and their correlation with human mobility, climatic, and social variables. Therefore, future work for a better understanding of vector-borne diseases in Costa Rica is oriented using this methodology.

The predictive capabilities of modeling techniques and the greater use of non-traditional information in the development of vector-borne disease control and prevention strategies can translate into more comprehensive enhanced surveillance at all levels. At the central government level, benefits may include efficient allocation of human, technical, and administrative resources. At the regional and local levels, benefits may include early identification of possible lack of resources, shortages, and improved design of interventions tailored to specific scenarios, such as more effective approaches to community risk communication and surveillance.

**Author contributions.** All authors conceived the original idea, analyzed the results, and wrote the paper. All authors reviewed and approved the final version.

**Conflict of interest.** The authors report no conflict of interest.

**Disclaimer.** Authors hold sole responsibility for the views expressed in the manuscript, which may not necessarily reflect the opinion or policy of the *Revista Panamericana de Salud Pública/Pan American Journal of Public Health* and/or those of the Pan American Health Organization.

## REFERENCES

1. World Health Organization. Global vector control response 2017–2030. Geneva: WHO; 2017. Available from: <https://www.who.int/publications/i/item/9789241512978>.
2. Egger M, Johnson L, Althaus C, Schöni A, Salanti G, Low N, et al. Developing WHO guidelines: time to formally include evidence from mathematical modelling studies. *F1000Res*. 2017;6:1584.
3. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton, NJ: Princeton University Press; 2011:5–14.
4. Reiner Jr RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *J R Soc Interface*. 2013;10(81):20120921.
5. Costa Rica, Ministerio de Salud. Análisis de Situación de Salud, 2019. Available from: <https://www.ministeriodesalud.go.cr/>.
6. Sanchez F, Barboza LA, Burton D, Cintrón-Arias A. Comparative analysis of dengue versus chikungunya outbreaks in Costa Rica. *Ric Mat*. 2018;7(1):163–74.
7. Sisson SA, Fan Y, Beaumont M, editors. Handbook of approximate Bayesian computation. London: CRC Press; 2018.
8. Sanchez F, Barboza L, Vásquez P. Parameter estimates of the 2016–2017 Zika outbreak in Costa Rica: An Approximate Bayesian Computation (ABC) approach. *Math Biosci Eng*. 2019;16(4):2738–55.
9. Vásquez P, Loria A, Sanchez F, Barboza LA. Climate-driven Statistical Models as effective predictors of local Dengue incidence in Costa Rica: A Generalized Additive Model and Random Forest Approach. *Rev Mat*. 2020;27(1):1–21.
10. García YE, Mery G, Vásquez P, Calvo JG, Barboza L, Rivas T, et al. Projecting the impact of Covid-19 variants and vaccination strategies in disease transmission using a multilayer network model in Costa Rica. *Sci Rep*. 2022;12(1):2279.

---

Manuscript submitted on 24 January 2022. Revised version accepted for publication on 21 April 2022.

---

## Modelos matemáticos y estadísticos para el control de enfermedades transmitidas por mosquitos: la experiencia de Costa Rica

### RESUMEN

**Objetivo.** Resumir los resultados de las investigaciones realizadas en Costa Rica en las que se aplicaron métodos matemáticos y estadísticos para estudiar la dinámica de transmisión de las enfermedades transmitidas por mosquitos.

**Métodos.** Se seleccionaron y analizaron tres artículos con análisis matemáticos y estadísticos sobre enfermedades transmitidas por vectores en Costa Rica. En estos artículos se muestra el valor y la pertinencia de emplear diferentes métodos cuantitativos para comprender la dinámica de la enfermedad y brindar apoyo a la toma de decisiones.

**Resultados.** Los resultados de estas investigaciones: 1) muestran la repercusión en los informes de casos de dengue cuando surge un segundo agente patógeno, como el chikunguña; 2) recuperan parámetros clave en la dinámica del Zika mediante la inferencia bayesiana; y 3) muestran el uso de los algoritmos de aprendizaje automático y las variables climáticas para pronosticar el riesgo relativo de dengue en cinco lugares diferentes.

**Conclusiones.** Los modelos matemáticos y estadísticos permiten describir la dinámica de transmisión de las enfermedades transmitidas por mosquitos, mediante la provisión de información cuantitativa para brindar apoyo a los métodos de prevención y control y a la planificación de la asignación de recursos.

### Palabras clave

Enfermedades transmitidas por vectores; modelos teóricos; salud pública; Costa Rica.

---

---

## Modelos matemáticos e estatísticos para o controle de doenças transmitidas por mosquitos: a experiência da Costa Rica

### RESUMO

**Objetivo.** Resumir os resultados de estudos realizados na Costa Rica em que foram aplicados métodos matemáticos e estatísticos para estudar a dinâmica de transmissão de doenças transmitidas por mosquitos.

**Métodos.** Foram selecionados e revisados três artigos com análises matemáticas e estatísticas sobre doenças transmitidas por vetores na Costa Rica. Esses artigos mostram o valor e a pertinência do uso de diferentes métodos quantitativos para compreender a dinâmica das doenças e apoiar a tomada de decisões.

**Resultados.** Os resultados dessas investigações: 1) mostram o impacto nas notificações de casos de dengue quando surge um segundo patógeno, como o chikungunya; 2) recuperam parâmetros-chave na dinâmica do zika, usando a inferência bayesiana; e 3) mostram o uso de algoritmos de aprendizagem por máquina e variáveis climáticas para prever o risco relativo da dengue em cinco locais diferentes.

**Conclusões.** A modelagem matemática e estatística permite a descrição da dinâmica de transmissão de doenças transmitidas por mosquitos ao oferecer informações quantitativas para apoiar métodos de prevenção e/ou controle e o planejamento da alocação de recursos.

### Palavras-chave

Doenças transmitidas por vetores; modelos teóricos; saúde pública; Costa Rica.

---