BMJ Open

Dengue epidemic alert thresholds for surveillance and decision-making in Puerto Rico: development and prospective application of an early warning system using routine surveillance data

Maile B Thayer ,[1] Melissa Marzan-Rodriguez,[2] Jomil Torres Aponte,[2] Aidsa Rivera,[1] Dania M Rodriguez,[1] Zachary J Madewell ,[1] Kristyna Rysava ,[1] Gabriela Paz-Bailey,[1] Laura E Adams,[1] Michael A Johansson [3]

To cite: Thayer MB, Marzan-Rodriguez M, Torres Aponte J, *et al*. Dengue epidemic alert thresholds for surveillance and decision-making in Puerto Rico: development and prospective application of an early warning system using routine surveillance data. *BMJ Open* 2025;**15**:e106182. doi:10.1136/bmjopen-2025-106182

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2025-106182).

Received 06 June 2025
Accepted 04 September 2025

[1]Dengue Branch, Centers for Disease Control and Prevention, San Juan, Puerto Rico
[2]Puerto Rico Departamento de Salud, San Juan, Puerto Rico
[3]Northeastern University, Boston, Massachusetts, USA

Correspondence to
Dr Michael A Johansson;
m.johansson@northeastern.edu

ABSTRACT

Objectives The Puerto Rico Department of Health (PRDH) seeks to identify dengue epidemics as early as possible with high specificity.

Design Development and prospective application of an early warning system for dengue epidemics using routine historical surveillance data. A weekly intercept-only negative binomial regression model was fitted using historical probable and confirmed dengue data. A range of threshold definitions was explored using three model-estimated percentiles of weekly dengue case counts.

Setting Dengue is endemic in Puerto Rico with irregular occurrence of large epidemics with substantial impact on health burden and health systems. Probable and confirmed dengue data are routinely collected from all hospitals and private clinics.

Participants A total of 86 282 confirmed or probable dengue virus cases were reported from 1 January 1986 to 30 June 2024, with an annual mean of 2212 cases (median: 1533; range: 40–10 356).

Primary and secondary outcome measures The model was fitted retrospectively to mimic real-time epidemic detection and assessed based on sensitivity and specificity of epidemic detection.

Results The 75th percentile threshold aligned best with historical epidemic classifications, balancing false alarms and missed detections. This model provides a robust method for defining thresholds, accounting for skewed data, using all historical data and improving on traditional methods like endemic channels.

Conclusions In March 2024, PRDH declared a public health emergency due to an early, out-of-season surge in cases that exceeded the epidemic alert threshold developed in this study. This real-time application highlights the value of these thresholds to support dengue epidemic detection and public health response. Integrating thresholds with other tools and strategies can enhance epidemic preparedness and management.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This method accounts for data skewness and uses all available historical data in Puerto Rico, unlike other traditional early warning systems.
- ⇒ Weekly thresholds are smoothed to reduce noise, which may limit sensitivity to abrupt short-term changes.
- ⇒ This tool was validated and used by local health partners and aligns closely with past epidemic periods on the island.
- ⇒ This method relies on several years of consistent historical data to ensure robust model fitting, which can be a challenge in settings with limited or inconsistent surveillance data.
- ⇒ The effectiveness of this method depends on the availability of timely case reports.

INTRODUCTION

Epidemics can have rapid and major impacts on population health and health systems. While predicting epidemics remains challenging, timely detection enables public health officials to respond swiftly and effectively. In endemic areas, where disease is consistently present, epidemic warnings must be carefully developed and linked to actionable responses to optimise public health efforts. In places where diseases are endemic, it can be difficult to distinguish baseline levels from epidemic activity due to the constant presence of the disease. There is also the potential for underestimation of epidemic severity due to adaptation and partial immunity within the population, fluctuations in case reporting and possible changes in diagnostic practices. In non-endemic areas, developing accurate epidemic detection mechanisms is often more complex due to low baseline levels of the disease. The US Centers for Disease Control and Prevention (CDC) defines an epidemic as 'an increase, often sudden, in the number of cases of a disease above what is normally expected in that population in that area'.[1] For a more precise definition tailored to specific diseases, regions and time frames, factors such as frequency, duration, amplitude and overall burden of cases should be considered. These elements are essential for defining disease-specific and region-specific epidemic alert thresholds—levels of disease incidence that trigger a necessary response. These definitions must try to strike a balance between minimising false alarms (too many epidemic alerts) and failures to correctly identify an epidemic (not enough epidemic alerts). These elements are crucial for defining epidemic alert thresholds, the number of cases that must be exceeded to signal an alarm for an epidemic.

Dengue viruses (DENV) impose a considerable economic and health burden to the areas affected, particularly during epidemic years. DENV is a mosquitoborne virus that presents an increasing public health challenge in tropical and subtropical regions.[2 3] Roughly half of the global population currently lives in areas suitable for DENV transmission, with the majority of those affected residing in Asia, followed by Africa and the Americas.[4]

Puerto Rico, a US territory in the Caribbean, has been greatly affected by DENV in recent decades and accounts for over 95% of locally acquired cases in the USA. Transmission is endemic and seasonal, with large epidemics occurring roughly every 4–5 years.[5] A large DENV epidemic in Puerto Rico took place in 2012–2013, resulting in over 18000 suspected and 9200 confirmed cases.[6–8] This period was followed by several years of relatively lower case counts until March 2024, when a concerning surge prompted a public health emergency declaration by Puerto Rico's Department of Health due to an unseasonably early increase in cases.

States and policy-makers define dengue epidemic alert thresholds in endemic regions using different methods, with varying results. One common approach, known as 'endemic channels', establishes thresholds by analysing recent data (often excluding 'epidemic years') using a measure of yearly variability (usually standard deviation, SD) and a measure of central tendency (usually the mean) to set an upper bound of expected case numbers (eg, 2 SDs above the mean).[9–12] Although widely used, this method has limitations, particularly the assumption that data (reported case counts of dengue) are normally distributed. In reality, epidemiological data are usually skewed, with many low counts and few very high counts. Assuming that the data are normally distributed, therefore, results in thresholds that do not accurately reflect the actual distribution of cases. This approach may even produce impossible values, such as the expectation of negative case counts in low transmission years. Most existing dengue threshold definitions depend on this assumption of normality, using variations of the mean as a measure of central tendency.[10 13] Similarly, endemic channel threshold implementations often consider only the last several (usually 5–8) years of data, excluding epidemic years or years with low case counts, such as during the COVID-19 pandemic.[9 10] This not only omits some available data but also overlooks valuable information from epidemics, the very events the thresholds are designed to detect.

Public health officials from the CDC and the Puerto Rico Department of Health (PRDH) have closely collaborated to develop and implement an improved method for defining dengue epidemic alert thresholds. This approach, currently used for classifying and detecting epidemics in Puerto Rico, allows for data heterogeneity and makes full use of historical data. This tool is currently used as a part of DENV surveillance, integrated in weekly epidemic assessments, and featured in the PRDH's weekly arboviral disease reports.

METHODS

Data

This thresholds analysis uses historical dengue surveillance data from Puerto Rico's Passive Arboviral Disease Surveillance System, including all probable and confirmed DENV cases (including private laboratory results) reported by healthcare providers from 1 January 1986 to 30 June 2024. In Puerto Rico, suspected dengue cases are first reported based on clinical suspicion to the surveillance system, jointly managed by PRDH and CDC, with laboratory confirmation in PRDH. Probable cases are defined as those with a positive immunoglobulin M result for DENV, and confirmed cases are defined as those with a positive reverse transcription Polymerase Chain Reaction (PCR) result for DENV. All analyses use probable and confirmed cases, aggregated on a weekly timescale.

Statistical analysis

Negative binomial model

We calculate real-time weekly dengue alert thresholds from 2001 onward using retrospective historical data by fitting a negative binomial intercept-only regression model to aggregated weekly probable and confirmed cases each week. This model calculates the thresholds in real time, that is, for each current weekly dengue epidemic threshold, it incorporates all prior probable and confirmed data points from the same week in previous years at the time of case reporting. For example, to calculate thresholds for the first week of 2023, we used all probable and confirmed cases from the first week in 1986–2022. We started to calculate thresholds in 2001 to allow for sufficient historical probable and confirmed dengue cases for the threshold calculations.

The negative binomial distribution models count data (in this case, dengue probable and confirmed case count data) while allowing for a flexible relationship between the mean and variance and accounting for skewed data, as is often seen with dengue case counts. Unlike traditional methods which assume normally distributed case counts, the negative binomial distribution effectively accounts for epidemic years and uses the full dataset. Note that a negative binomial intercept-only regression model is equivalent to a negative binomial distribution; however, we deliberately framed this approach as a regression model so that others can easily extend it to include additional predictors (environmental factors, demographic variables and lagged terms).

The negative binomial regression model allows us to extract percentile case count estimates that serve as dengue alert thresholds for comparison with current probable and confirmed weekly dengue case counts. For instance, the 50th percentile represents the historical median of probable and confirmed dengue case counts, which is not the actual median from historical counts but rather the median estimated by the model based on all historical data available at the time of epidemic threshold assessment. This approach provides an estimate of the 'typical' case count. Higher percentiles of case counts provide insights into epidemics, which are not expected annually. To illustrate the range of possible epidemic threshold definitions, we present three percentiles for epidemic detection: the 60th, 75th and 90th percentiles, representing progressively more stringent epidemic definitions. These percentiles roughly correspond to expected epidemics occurring once every 2.5, 4 and 10 years, respectively (note that $\frac{1}{1-0.60}=2.5$; $\frac{1}{1-0.75}=4$; and $\frac{1}{1-0.90}=10$). This range of percentile-defined thresholds demonstrates how threshold performance varies under more frequent vs rarer epidemic definitions, while keeping the example definitions concise and interpretable. Estimated thresholds are based on data for previous years and can be updated annually as new data become available. Each week's threshold is derived from its corresponding historical week. As data accumulates over time, the thresholds become progressively more stable.

After calculating the thresholds using the negative binomial regression model, we smooth the threshold trajectories to eliminate noise and random variation in the week-to-week threshold patterns. We apply centred linear filtering with convolution using a fast Fourier transform over 11 weeks (including the week of interest, plus 5 weeks before and 5 after each data point). This technique calculates a moving average of thresholds across the chosen window of 11 weeks. Centred linear filtering assigns weights to the data points within the window, with the current week receiving the highest weight and weights gradually decreasing for weeks further away. This approach minimises the influence of distant weeks while preserving overall trends. The fast Fourier transform efficiently performs the convolution, making it faster than calculating the weighted average directly. The 11-week window was chosen after performing a sensitivity analysis across a range of windows and selecting the one with the best balance between stability and reliability, that is, removing noise and maintaining the main signal in the data (online supplemental figure S2). With increasing smoothing windows, roughness of the signal decreased (the sum of squared differences between consecutive values), indicating greater stability of the threshold, and the variability in peak timing and magnitude increased, reflecting oversmoothing (online supplemental table S1). The window of 11 minimised roughness while maintaining consistent peak timing and magnitude. For windows less than 11, there were still some fluctuations due to noise, and for windows greater than 11, there was not much improvement on capturing the overall signal and the thresholds began to flatten.

Dengue epidemic alert threshold outcome

Based on the analyses described above, the final criterion used to detect epidemics was exceeding the 75th percentile epidemic alert threshold for two or more consecutive weeks. The primary outcome of this method is a real-time epidemic alert threshold by week, that is, a percentile-defined benchmark estimate for current weekly levels of dengue incidence that serves as a decision marker for whether a population (in this case, Puerto Rico) is in or at risk for an epidemic. We applied this method across all weeks from 1986 to mid-2024 in real-time (ie, dengue cases reported at the time of assessment, reflecting what would have been known at the time).

Application of thresholds in real-time

To illustrate how these epidemic alert thresholds can be applied in practice, we highlight two examples from Puerto Rico using the final 75th percentile definition of thresholds in comparison to real-time reported dengue cases. In both instances, real-time reported probable and confirmed cases from the surveillance system were high and the epidemic alert thresholds helped to guide decision-making on whether to declare an epidemic.

Threshold performance

Model performance for each threshold definition carried out in real-time (ie, calculating each weekly threshold only with data available up to that date) was evaluated in comparison to expected epidemic years from public health partners at PRDH, using sensitivity and specificity as the primary performance metrics. Sensitivity in this case is a measure of the proportion of true positive epidemic years among all dengue epidemic years, while specificity is the proportion of true non-epidemic years among all non-epidemic years. The optimal definition balances high sensitivity and specificity. We further evaluated the performance of each threshold definition using positive predictive value and negative predictive value and plotted each threshold in the receiver operating characteristic (ROC) plane for visualisation. We calculated conservative 95% confidence intervals (CIs) for all performance metrics using Clopper-Pearson exact binomial confidence intervals due to the relatively small sample size of epidemic and non-epidemic years available.[14]

All analyses were performed using R V.4.4.0.[15] We used the 'MASS' package for fitting the negative binomial model

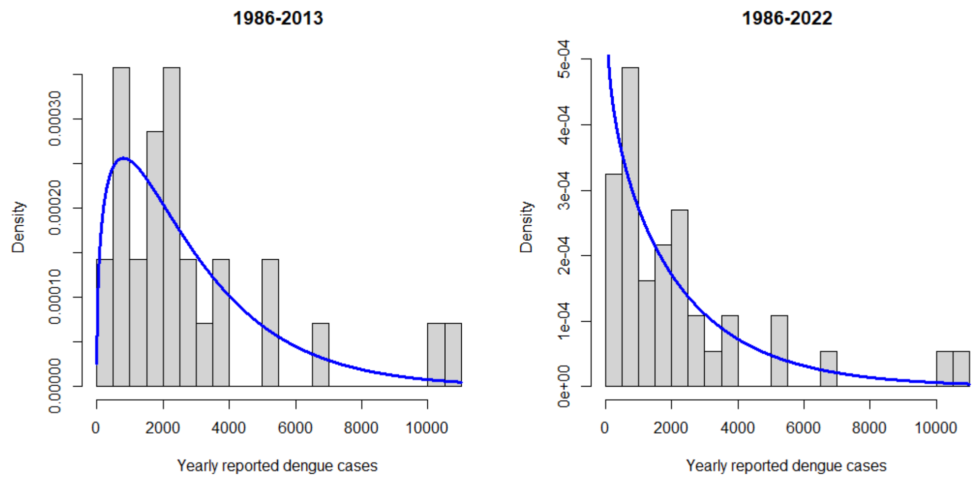

Figure 1 Histograms of total yearly cases from 1986 to 2013 and from 1986 to 2023 in Puerto Rico, fit with negative binomial distributions. The densities of annual reported probable and confirmed island-wide dengue cases are shown (grey bars), with the corresponding fitted negative binomial distribution in blue. On the left are pre-Zika and chikungunya data (1986–2013) and on the right are all available historical data as of 1 January 2024 (1986–2023).

(function 'glm.nb') and package 'stats' for smoothing (function 'filter').[16 17] Code with example data is available on GitHub: https://github.com/CDCgov/dengue_epidemic_thresholds.[18]

RESULTS

From January 1986 to June 2024, a total of 86 282 confirmed or probable DENV cases were reported, with an annual mean of 2212 cases (median: 1533; range: 40–10 356). Histograms of yearly cases demonstrate right skewness, with more pronounced skewness in recent data due to more frequent low case counts since 2013 (figure 1, online supplemental figure S1). Cases typically peaked around October and reached a yearly low in April. Annually fitted thresholds from 2001 showed general stability over time, with slight increases following the 2010 and 2012–2013 epidemics, and a gradual decrease during subsequent low transmission years (figure 2).

To evaluate the performance of the proposed thresholds in identifying epidemic years, we compared them with data from 2001 to 2023. According to public health partners, the largest epidemics during this period occurred in 2007, 2009, 2010, 2012 and 2013. The three threshold definitions mostly agreed on identifying epidemic years, detecting epidemics in 2001, 2005, 2007, 2010, 2012 and 2013. However, there were discrepancies, resulting in variations in numbers of identified outbreaks and different timing of when the epidemics would have been detected (figure 3). With a threshold at the 60th percentile, the alert threshold would have been exceeded in 14 of the 23 years, including 6 years that would not have been detected using higher threshold definitions.

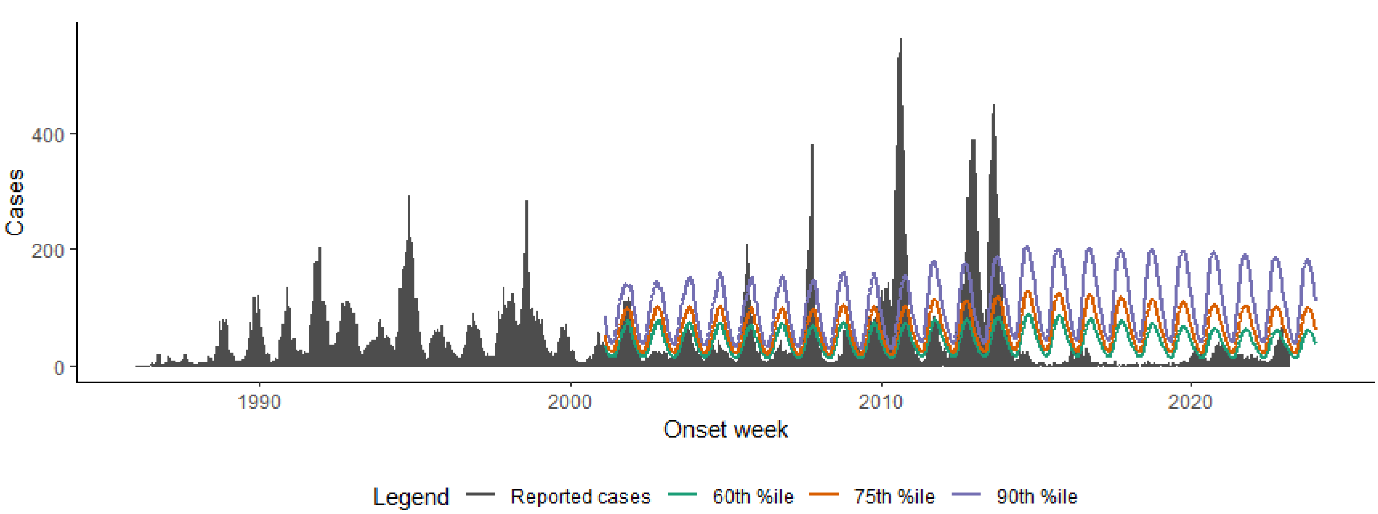

Figure 2 Reported probable and confirmed dengue virus cases and three estimated outbreak thresholds—Puerto Rico, 1986–2024. The weekly reported probable and confirmed dengue cases (≥1 positive PCR or IgM result) are represented by the grey bars. Weekly thresholds were calculated at the beginning of each new year starting in 2001 and are represented by the coloured lines (60th percentile in green; 75th percentile in orange and 90th percentile in purple).

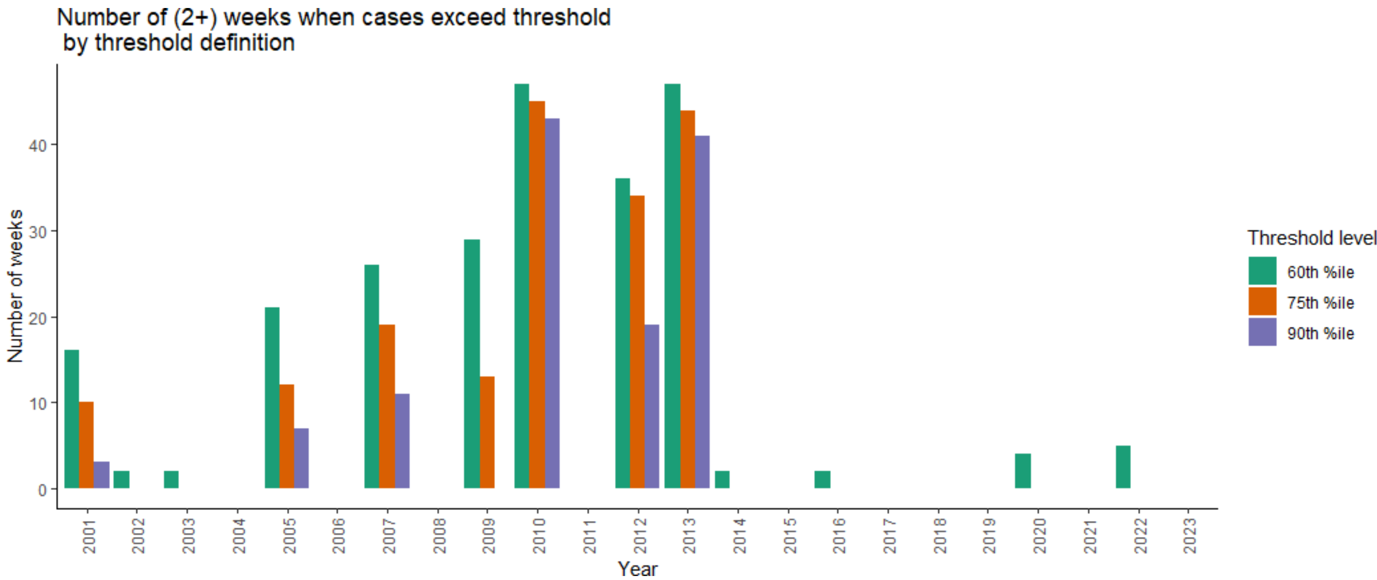

Figure 3 Number of weeks (2 or more) per year when dengue cases exceed threshold, with differing definitions: Puerto Rico January 2001–June 2024. The number of weeks when the reported probable or confirmed cases are above the dengue threshold is tallied for each year using different threshold definitions, represented by the different-coloured bars. A threshold using the 60th percentile is shown in green, the 75th percentile (currently used) is shown in orange and the 90th percentile is shown in purple. Two or more consecutive weeks above the threshold are considered an epidemic. Note that 2024 is shown, but data are only included up to the end of June.

In 2009, cases exceeded the 60th percentile threshold and then dropped below it six different times. In 2020, cases exceeded the 60th percentile two separate times. With a threshold at the 90% percentile, the threshold would have been exceeded in only six of the 24 years. In two of those 6 years identified, cases exceeded and then dropped below the 90th percentile threshold twice in 2010 and then 16 times in 2012. Comparing these thresholds with alternative classifications revealed that 60% was too low, detecting epidemics in years with manageable case numbers, yet 90% was too high; for example, using 90% missed 2009 when case numbers were high and exceeded and then dropped below the threshold multiple times in 2010 and 2012. After consultation with PRDH, the 75th percentile was found to align best with their experience and management response required for previous epidemics. An epidemic is then considered by PRDH as two or more consecutive weeks when reported probable and confirmed cases exceed the weekly (75th percentile) epidemic alert threshold.

Using the 75th percentile alert threshold definition along with the requirement that reported probable and confirmed cases exceed this threshold for at least two consecutive weeks, there were 6 years with outbreaks detected, with long periods above the threshold occurring in 2010 and 2013 (45 weeks each, figure 3). This threshold definition accurately identified the 6 years (2001, 2005, 2007, 2010, 2012 and 2013) that PRDH identified as epidemic years. The 75th percentile definition yielded the best performance across all metrics, with a sensitivity of 100% and specificity of 89% when comparing threshold-defined epidemic years to PRDH-expected epidemic years (online supplemental figure S3 and table S2). Moreover, the seasonality that is naturally incorporated into these weekly thresholds detected not only whether an epidemic is occurring, but also which time of the year, how long and by how much the reported probable and confirmed cases exceeded the threshold. Weekly alert 75th percentile thresholds ranged from a low of 18 cases during the low season in spring to a high of 126 during peak transmission months in fall (mean: 62; median: 59) from 2001 to June 2024. For other settings, it is important to collaborate with local health departments and policy-makers to define the threshold level and decide on the number of consecutive weeks needed to define epidemics.

Two real-time uses of 75th percentile thresholds: 2022 and 2024 dengue in Puerto Rico

At the end of 2022, there was a rapid increase of DENV cases in Puerto Rico. The epidemic alert thresholds were already in use in real-time, and the official alert criteria were exceeding the 75th epidemic alert thresholds for two or more consecutive weeks. Although the reported probable and confirmed case numbers were higher than usual (exceeding the historical median) starting in October 2022, they approached but never surpassed the (75th percentile) epidemic alert threshold (figure 4A). With the weekly thresholds available, PRDH and the CDC monitored the status of the case counts each week in comparison to the thresholds and concluded that while cases were above the historical median, they never exceeded the thresholds and hence did not trigger an epidemic alert. Additionally, trends in the historical median and threshold helped predict that dengue case numbers would decrease at the beginning of the year,

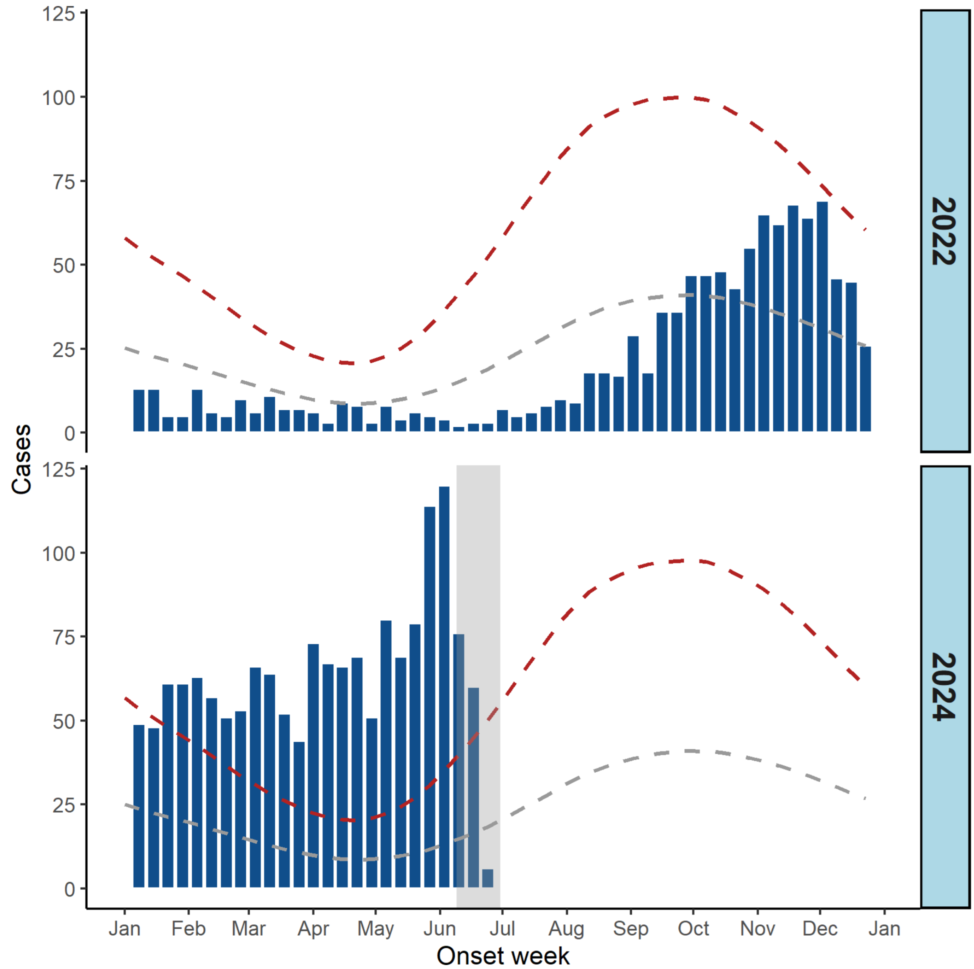

Figure 4 Reported dengue cases compared with historical weekly dengue alert thresholds in Puerto Rico, 2022 and 2024. The weekly reported probable and confirmed cases (≥1 positive PCR or IgM result) are represented by the navy bars, while the weekly historical median and threshold are represented by the dashed red and grey lines, respectively. The last 3 weeks of data in 2024 are highlighted in the light grey rectangle to denote that the most recent weeks are likely subject to reporting delays.

which was accurately reflected in the downturn in early 2023.

On the other hand, DENV cases began to increase in Puerto Rico at the beginning of 2024. The numbers exceeded the historical threshold in January 2024, a period when cases are generally low and were expected to be decreasing (figure 4B). The dengue alert threshold, combined with a review of years when cases similarly exceeded the threshold early in the year (2009, 2010, 2013), supported the decision to declare an epidemic in Puerto Rico. Cases remained above the threshold and correctly predicted in the first dengue epidemic in over a decade.

DISCUSSION

Epidemic alert thresholds serve as critical tools for public health decision-making for preparedness and response to combat dengue epidemics. A well-defined and validated threshold provides an early yet reliable indication of an impending epidemic, allowing for timely public messaging, resource allocation planning, faster implementation of interventions like vector control and risk communication strategies. Earlier response and interventions can help mitigate health risks and reduce healthcare costs or economic losses associated with epidemics.

The method presented here for classifying epidemics offers several advantages over traditional approaches used for epidemic detection. These traditional approaches use various tools, each with its own strengths and limitations. Early warning systems based on exceeding thresholds, for example, use predetermined thresholds based on specific indicators (eg, case numbers) to trigger alerts.[19] While simple and interpretable, these systems often rely on static thresholds that might not adapt to changing trends or skewed data distributions. Statistical anomaly detection, another approach, uses statistical techniques to identify unusual patterns in surveillance data that might signal an outbreak.[20 21] However, these methods can be sensitive to random fluctuations and might struggle to differentiate between true outbreaks and seasonal variations, potentially leading to missed detections or false alarms. Our approach acknowledges the skewed nature of dengue case counts and employs the negative binomial distribution, a more suitable statistical model for such data compared with methods assuming normality. Additionally, our method leverages all available historical data, strengthening the model's robustness as more information is incorporated each year. This approach integrates outliers, including past epidemics, as valuable data points for building more informed thresholds. By using a negative binomial regression model, our method allows model-estimated medians and quantiles to naturally incorporate these outliers, leading to more comprehensive thresholds for epidemic detection.

Similar to the 2 SD thresholds used in most endemic channels to assess annual variability,[9] our method requires choosing a percentile value to define the epidemic threshold. Two key factors can guide this choice, both related to the impact of large epidemics that would warrant additional response actions. First, historical data on epidemic occurrence in a specific region can inform the threshold selection. For instance, if major dengue epidemics in Puerto Rico typically occur every 4 or 5 years, thresholds set at the 75th or 80th percentile might be appropriate. These higher percentiles capture a substantial increase in cases compared with typical years, potentially signalling an emerging epidemic. Second, we can gain valuable insights by fitting the thresholds retrospectively to historical data. By comparing these thresholds with known epidemic years (established by public health authorities), we can determine which percentile level aligns most effectively with past epidemic detections. This approach ensures the threshold accurately identifies periods requiring a heightened public health response.

We did not use a data-driven metric like ROC curves to determine the epidemic alert threshold cut-off; rather, we deliberately prioritised a public health decision-making-driven approach. Prior to these epidemic alert thresholds, no gold standard to prospectively identify an outbreak or epidemic existed. Instead of a purely statistical approach, our engagement with public health partners and officials ensured that the thresholds were credible, acceptable and implementable in the context of decision-making in Puerto Rico. This public health-driven approach is a strength as it increases the likelihood of the adoption and sustainable use of these thresholds. Indeed, the epidemic alert thresholds are already incorporated into routine weekly reports of dengue on the island.

Although temporal changes in dengue cases such as under-reporting or population migration over time are a recognised limitation in surveillance data, surveillance has been largely consistent in Puerto Rico and we did not adjust for it in the analysis. The decision to restrict data to only probable and confirmed dengue cases reduced potential variability in reporting practices. This subset of data is the most constant and reliable source of data for dengue trends in Puerto Rico. Clinically suspected cases are confounded by the chikungunya and Zika virus outbreaks on the island, hospitalisations due to dengue are not well captured in the passive surveillance system, and fatalities are also underreported and sparse.[22] Additionally, we used almost 40 years of weekly data to estimate the epidemic alert thresholds, which can incorporate any potential temporal changes over time as they are updated every year. As a result, under-reporting is unlikely to systematically bias the thresholds.

Additionally, dengue dynamics evolve over time and changing dengue incidence requires regular updates of these epidemic alert thresholds. Variations of this approach could restrict to smaller time series of current data, enabling dynamic updating. However, restricting data in this way could create less stable threshold estimates because of the smaller data sample. The inclusion of almost four decades of data in the analysis presented here, updated to incorporate new data each year, naturally incorporates possible temporal changes over time. Moreover, our approach applies smoothing of weekly thresholds to reduce noise and random fluctuations. This smoothing improves interpretability but may reduce sensitivity to abrupt short-term changes, which should be considered when applying the tool for real-time decision-making. However, periodic reassessments of these thresholds are essential to maintain the accuracy and relevance of these benchmark estimates.

While the negative binomial model offers a relatively straightforward approach to fitting epidemic alert thresholds, several limitations require consideration. First, the model relies on several years of consistent historical data to ensure robust model fitting. This can be a challenge in settings with limited or inconsistent surveillance data.[23] Second, the use of all historical data assumes that it is representative of current patterns. If there is a reason to believe that trends have changed, the historical data could be restricted to exclude certain years. Additionally, the effectiveness of this method hinges on the availability of timely case reports for comparison with the thresholds. During outbreaks, reporting delays can be exacerbated due to overwhelmed surveillance systems.[24] Nowcasting methods that estimate unreported cases and validated forecasts that assess the likelihood of exceeding the threshold in the near future could help mitigate these delays, as demonstrated for mpox and COVID-19.[25] Furthermore, the method can be less effective when dealing with extended periods of low or zero weekly case counts. This could occur during periods of minimal transmission or in settings with small populations. We can improve the model's performance in sparse data scenarios by incorporating techniques like temporal bootstrapping or sparse approximate inference for spatiotemporal point processes.[26] We used an intercept-only negative binomial regression framework to provide a simple and interpretable threshold method that indirectly incorporates seasonality from the historical data, while remaining flexible for future extensions with predictor data. Future work could explore more complex variations of the thresholds model, for example, differentiating between higher and lower dengue case incidence years or adding additional predictors to the regression model such as environmental or demographic factors.

Finally, while the thresholds presented here use data on probable and confirmed dengue cases, laboratory confirmation might not be readily available in all locations. However, the model can also use suspected case data for threshold development. In Puerto Rico, suspected case data were used before the introduction of chikungunya and Zika viruses, which share symptoms with dengue. The emergence of these viruses led to a rise in 'suspected' dengue cases during the chikungunya outbreak in 2014 and the Zika outbreak in 2016, likely reflecting these other infections. Using confirmed and probable data addressed this issue. While the model can use suspected case data in settings with limited laboratory resources, this approach may include false positives, potentially leading to less accurate thresholds. In Puerto Rico, with its robust dengue surveillance system and routine testing for chikungunya and Zika, confirmed and probable cases provided the most reliable data for threshold development. The thresholds shown illustrate Puerto Rico as an example, but the thresholds are easily adaptable to other settings with different data types and patterns.

This evidence-based approach provided valuable real-time data to support communication with media and public partners, contextualising the increase in dengue cases within historical norms in both 2022 and 2024. In 2024, this early identification of an epidemic via these epidemic alert thresholds allowed public health officials to respond quickly and appropriately before cases began to increase sharply above expected levels.

The utility of this tool lies in its real-time capability to identify and classify an epidemic alert. In Puerto Rico, the weekly historical median and epidemic alert thresholds are established annually in January for the upcoming year. Each week, as reported cases become available, the number of cases is compared with the weekly historical median and epidemic threshold to assess the current status of DENV transmission. If case counts exceed the threshold for two consecutive weeks, an epidemic alert is declared. This evidence-based, timely and straightforward comparison and classification provides an ideal tool for guiding responses to emerging epidemics before high levels of DENV transmission are apparent. However, to maximise planning and response efforts, this method should be used in combination with other strategies and considered within the broader epidemiological context. Recognising high levels of DENV transmission is an important step, but it is also crucial to examine other criteria such as geographical regions and subpopulations where high levels are occurring. Using and evaluating these thresholds in conjunction with other tools, and as part of a larger discussion with key partners, can help support transparency and build trust between public health agencies and community members.

This threshold tool serves as an early warning system for dengue outbreaks, offering a crucial component within the broader public health surveillance and response framework. It can be combined with other tools for earlier identification and detection, such as integrating forecasts with thresholds or using methods with higher sensitivity and lower specificity. These methods can provide early indicators of increasing incidence. Although we have described these threshold methods in the context of dengue in Puerto Rico, they are adaptable to other geographical scales, locations and diseases.

Contributors All authors have met the International Committee of Medical Journal Editors (ICMJE) criteria for authorship of this article and gave their approval for this version to be published. MBT: study conceptualisation, methodology, software, formal analyses, visualisation, writing and guarantor (original draft preparation, reviewing and editing). MM-R: data curation, study conceptualisation, supervision and writing (reviewing and editing). JTA: data curation, study conceptualisation and writing (reviewing and editing). AR: study conceptualisation and writing (reviewing and editing). DMR: data curation and writing (reviewing and editing). ZJM: visualisation and writing (reviewing and editing). KR: methodology and writing (reviewing and editing). GP-B: study conceptualisation, supervision and writing (reviewing and editing). LEA: study conceptualisation, supervision, methodology and writing (reviewing and editing). MJ: study conceptualisation, supervision, methodology, formal analyses, writing and guarantor (reviewing and editing).

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study did not require institutional ethical approval nor patient consent for publication, as it involved secondary analysis of routinely collected, deidentified surveillance data.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. Code with example data is available on GitHub: https://github.com/maile-thayer/epidemic_thresholds. More detailed data can be requested from the Puerto Rico Department of Health: https://forms.office.com/pages/responsepage.aspx?id=WgYG6T7wrUekxGsTmghEXBHzWo_iDtJOjtyx8LYE1wtUQTY3VjBIVVU3M1pSWDNJMUJGQk8zWDNNViQlQCN0PWcu&origin=QRCode&route=shorturl.

ORCID iDs
Maile B Thayer http://orcid.org/0000-0001-6952-0440
Zachary J Madewell http://orcid.org/0000-0002-5667-9660
Kristyna Rysava http://orcid.org/0000-0002-3475-931X
Michael A Johansson http://orcid.org/0000-0002-5090-7722

REFERENCES

1 U.S. Department of Health and Human Services. *Principles of epidemiology in public health practice*. 3rd edn. Atlanta: Centers for Disease Control and Prevention, 2012.
2 Shepard DS, Undurraga EA, Halasa YA, *et al*. The global economic burden of dengue: a systematic analysis. *Lancet Infect Dis* 2016;16:935–41.

3 Stanaway JD, Shepard DS, Undurraga EA, *et al*. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infect Dis* 2016;16:712–23.
4 Messina JP, Brady OJ, Golding N, *et al*. The current and future global distribution and population at risk of dengue. *Nat Microbiol* 2019;4:1508–15.
5 Méndez-Lázaro P, Muller-Karger FE, Otis D, *et al*. Assessing climate variability effects on dengue incidence in San Juan, Puerto Rico. *Int J Environ Res Public Health* 2014;11:9409–28.
6 Noyd DS. Recent Advances in Dengue: Relevance to Puerto Rico. *P R Health Sci* 2015;34:65–70.
7 Sharp TM, Hunsperger E, Muñoz-Jordán JL, *et al*. Sequential episodes of dengue--Puerto Rico, 2005-2010. *Am J Trop Med Hyg* 2014;91:235–9.
8 Centers for Disease Control and Prevention. Informe Semanal de Vigilancia del Dengue: 1 al 7 de enero de. 2014.
9 Bortman M. Elaboración de corredores o canales endémicos mediante planillas de cálculo. *Rev Panam Salud Publica* 1999;5:1–8.
10 Brady OJ, Smith DL, Scott TW, *et al*. Dengue disease outbreak definitions are implicitly variable. *Epidemics* 2015;11:92–102.
11 Hernández M, Arboleda D, Arce S, *et al*. Metodología para la elaboración de canales endémicos y tendencia de la notificación del dengue, Valle del Cauca, Colombia, 2009-2013. *biomedica* 2009;36:98.
12 Badurdeen S, Valladares DB, Farrar J, *et al*. Sharing experiences: towards an evidence based model of dengue surveillance and outbreak response in Latin America and Asia. *BMC Public Health* 2013;13:607.
13 World Health Organization. *Operational guide using the web-based dashboard early warning and response system (Ewars) for dengue outbreaks*. Geneva, 2020.
14 Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* 1934;26:404–13.
15 R Foundation for Statistical Computing. R: A language and environment for statistical computing [program]. Vienna, Austria. 2021.
16 The r stats package [program]. 4.4.0 version.
17 Venables WR. *Modern applied statistics with S*. 4th edn. New York: Springer, 2002.
18 Dengue Epidemic Alert Thresholds [program]. 2025.
19 Lowe R, Bailey TC, Stephenson DB, *et al*. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Comput Geosci* 2011;37:371–81.
20 Eze PU, Geard N, Mueller I, *et al*. Anomaly Detection in Endemic Disease Surveillance Data Using Machine Learning Techniques. *Healthcare (Basel)* 2023;11:1896.
21 Chen H, Zeng D, Yan P, *et al*. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Comput Geosci* 2010;49–72.
22 Tomashek KM, Rivera A, Torres-Velasquez B, *et al*. Enhanced Surveillance for Fatal Dengue-Like Acute Febrile Illness in Puerto Rico, 2010-2012. *PLoS Negl Trop Dis* 2016;10:e0005025.
23 Shadbolt N, Brett A, Chen M, *et al*. The challenges of data in future pandemics. *Epidemics* 2022;40:S1755-4365(22)00054-8.
24 McGough SF, Johansson MA, Lipsitch M, *et al*. Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Comput Biol* 2020;16:e1007735.
25 Charniga K, Madewell ZJ, Masters NB, *et al*. Nowcasting and forecasting the 2022 U.S. mpox outbreak: Support for public health decision making and lessons learned. *Epidemics* 2024;47:S1755-4365(24)00016-1.
26 Cseke B, Zammit-Mangion A, Heskes T, *et al*. Sparse Approximate Inference for Spatio-Temporal Point Process Models. *J Am Stat Assoc* 2016;111:1746–63.