Improving early epidemiological assessment of emerging *Aedes*-transmitted epidemics using historical data

Julien Riou[1,2,*], Chiara Poletto[1], Pierre-Yves Boëlle[1]

1 Sorbonne Université, INSERM, Institut Pierre Louis d'épidémiologie et de Santé Publique, IPLESP UMR-S1136, F-75012 Paris, France
2 EHESP School of Public Health, Rennes, France

*julien.riou@iplesp.upmc.fr

Abstract

Model-based epidemiological assessment is useful to support decision-making at the beginning of an emerging *Aedes*-transmitted outbreak. However, early forecasts are generally unreliable as little information is available in the first few incidence data points. Here, we show how past *Aedes*-transmitted epidemics help improve these predictions. The approach was applied to the 2015-2017 Zika virus epidemics in three islands of the French West Indies, with historical data including other *Aedes*-transmitted diseases (Chikungunya and Zika) in the same and other locations. Hierarchical models were used to build informative *a priori* distributions on the reproduction ratio and the reporting rates. The accuracy and sharpness of forecasts improved substantially when these *a priori* distributions were used in models for prediction. For example, early forecasts of final epidemic size obtained without historical information were 3.3 times too high on average (range: 0.2 to 5.8) with respect to the eventual size, but were far closer (1.1 times the real value on average, range: 0.4 to 1.5) using information on past CHIKV epidemics in the same places. Likewise, the 97.5% upper bound for maximal incidence was 15.3 times (range: 2.0 to 63.1) the actual peak incidence, and became much sharper at 2.4 times (range: 1.3 to 3.9) the actual peak incidence with informative *a priori* distributions. Improvements were more limited for the date of peak incidence and the total duration of the epidemic. The framework can adapt to all forecasting models at the early stages of emerging *Aedes*-transmitted outbreaks.

Author summary

In December, 2015, *Aedes* mosquito-transmitted Zika outbreaks started in the French West Indies, about two years after Chikungunya epidemics, spread by the same mosquito, hit the same region. Building on the similarities between these epidemics – regarding the route of transmission, the surveillance system, the population and the location – we show that prior information available at the time could have improved the forecasting of relevant public health indicators (i.e. epidemic size, maximal incidence, peak date and epidemic duration) from a very early point. The method we describe, together with the compilation of past epidemics, improves epidemic forecasting.

Introduction

Model-based assessments must be done in real time for emerging outbreaks: this was the case in recent years for MERS-CoV in the Middle East [1–3], Ebola virus in West

Africa [4–10], Chikungunya virus (CHIKV) [11] and Zika virus (ZIKV) [12–14] in the Americas. These analyses often focused on transmissibility and reproduction numbers rather than on forecasting the future impact of the epidemic. Indeed, forecasting is difficult before the epidemic reaches its peak, all the more when information on natural history, transmissibility and under-reporting is limited [15]. Yet, it is precisely at the beginning of an outbreak that forecasts would help public health authorities decide on the best strategies for control or mitigation.

Several methods have been used to make epidemic predictions, including exponential growth models [5, 16], sigmoid-based extrapolations [17], SIR-type models [18] and more realistic model accounting for spatial and population structure [19]. But in addition to specifying a model, selecting good parameter values is also essential to obtain good predictions. In models for directly transmitted diseases, this can come from realistic demographic and behavioral characteristics, for example the contact frequency between individuals [20], mobility patterns [21–24], and from clinical and epidemiological characteristics like the duration of the serial interval [25]. Such information is less easily available and more limited for mosquito-transmitted diseases [26]. However, outbreaks of the same disease or diseases with similar routes of transmission may have occurred in the same or similar locations, so that relevant information may be recovered from the analysis of past outbreaks.

Here, we show that past outbreaks of *Aedes*-transmitted diseases can substantially improve the epidemiological assessment of diseases transmitted by the same vector from early surveillance data. We introduce a hierarchical statistical model to analyze and extract information from historical data and obtain *a priori* distributions for required epidemiological parameters [27]. The method is illustrated with ZIKV outbreaks in the French West Indies between December, 2015 and February, 2017, using historical data regarding CHIKV and ZIKV epidemics in French Polynesia and the French West Indies between 2013 and early 2015. We assess the improvement in predictability of several operational indicators using different choices of *a priori* distribution and according to epidemic progress.

Methods

Data

Surveillance data on the 2015-2017 ZIKV epidemics in Guadeloupe, Martinique and Saint-Martin were collected by local sentinel networks of general practitioners and reported weekly by the local health authorities (Fig. 1) [28]. Cases of ZIKV infection were defined as "a rash with or without fever and at least two signs among conjunctivitis, arthralgia or edema". We obtained numbers of suspected cases by week for each island, extrapolated from the number of active sentinel sites (dataset $\mathcal{D}1$). In the West Indies, local health authorities described the situation as "epidemic" when incidence was larger than 1 per 2,000 population per week (i.e. 200 cases in Guadeloupe and Martinique [29], and 20 cases in Saint-Martin). Following this description, we defined the "S"(-tart) of the epidemic as the first week above this threshold, the "P"(-eak) date when incidence was the highest, and the "E"(-nd) of the epidemic as the third consecutive week below the threshold (to ascertain the downwards trend). The time interval from "S" to "E" corresponds to the period of "high epidemic activity".

We then analyzed historical data on the spread of emerging *Aedes*-transmitted diseases in similar locations. CHIKV epidemics occured in the same three islands during 2013-2015. Both diseases were transmitted by the same vector (*Aedes*), circulated in the same immunologically naive populations within a period of two years, had the same kind of clinical signs (i.e., fever, rash and arthralgia) and were reported by the same

surveillance system. Surveillance data on CHIKV epidemics in the French West Indies was available from local health authorities (dataset $\mathcal{D}2$) [30]. Finally, we also selected the ZIKV and CHIKV epidemics that occured in six islands or archipelagoes of French Polynesia between 2013 and 2017, as they provided information on the differences between the two diseases. Surveillance data regarding the outbreaks in French Polynesia (dataset $\mathcal{D}3$) was collected following similar methods as in the French West Indies [31, 32].

Epidemic model

The ZIKV outbreaks in Guadeloupe, Martinique and Saint-Martin were modelled separately using a dynamic discrete-time SIR model within a Bayesian framework. Briefly, the two main components of the model were: (i) a mechanistic reconstruction of the distribution of the serial interval of the disease (the time interval between disease onset in a primary case and a secondary case) that allows bypassing vector compartments; (ii) a transmission model for the generation of observed secondary cases in the human host. The generation time distribution was reconstructed by estimating the durations of each part of the infection cycle using disease- and mosquito-specific data from the literature, and assuming a fixed local temperature of $28°C$, as described in more detail in [33]. This led to a gamma distributions with mean 2.5 weeks (standard deviation: 0.7) for ZIKV and with mean 1.6 weeks (sd: 0.6) for CHIKV.

Then, we linked weekly observed incidence $O_{t,X}$ to past incidence with:

$$O_{t,X}|O_{0,\cdots,t-1,X}, \mathcal{R}_{0,X}, \rho_X, \phi_X \sim \text{NB}\left(\mathcal{R}_{0,X}\frac{S_{t,X}}{N}\sum_{n=1}^{5} w_{t,X,n}O_{t-n}, \phi_X\right) \quad (1)$$

where subscript X refers to disease ($X = C$ for CHIKV or $X = Z$ for ZIKV), $\mathcal{R}_{0,X}$ is the basic reproduction number, N the population size, and $S_{t,X} = N - \sum_{k=0}^{t-1} O_{k,X}/\rho_X$ the number of individuals susceptible to infection at time t where ρ_X is the reporting rate. The term $\sum_{n=1}^{5} w_{t,X,n}O_{t-n}$ accounts for exposure to infection at time t, where w_k is the discretized serial interval distribution. The variance is computed as the mean divided by the overdispersion parameter ϕ_X. The model was implemented in Stan version 2.15.1 within R version 3.4.0 [34–36]. More details regarding the epidemic model are available in Riou & al [33].

Informative and non-informative prior distributions

To analyze Zika epidemics, the reproduction ratio $\mathcal{R}_{0,Z}$, the reporting rate ρ_Z and the overdispersion parameter ϕ_Z must be estimated. For ϕ_Z, we used a non-informative prior in all cases [37]. For $\mathcal{R}_{0,Z}$ and ρ_Z, we designed three different prior distributions, labelled as "non-informative" (NI), "regional" (R), or "local" (L) and described below.

The NI prior distributions expressed vague characteristics of the parameters: $\mathcal{R}_{0,Z}$ will be positive and likely not greater than 20; and ρ_Z will range between 0 and 1.

The R and L priors were derived from the analysis of datasets $\mathcal{D}2$ and $\mathcal{D}3$ in three steps. We first analysed jointly the three CHIKV epidemics in dataset $\mathcal{D}2$ using model 1 and a two-level hierarchical structure for the island specific parameters. More precisely, we considered island specific reproduction numbers $\mathcal{R}_{0,C,i}$ sampled from a top-level island-averaged distribution $\mathcal{N}(\mu_{\mathcal{R}_0,C}, \sigma_{\mathcal{R}_0,C})$, and similarly $\text{logit}(\rho_{C,i}) \sim \mathcal{N}(\mu_{\rho_C}, \sigma_{\rho_C})$ for the logit of the reporting rate. We obtained the posterior distributions of the epidemiological parameters of the CHIKV outbreaks in each island $\pi(\mathcal{R}_{0,C,i}, \rho_{C,i}|\mathcal{D}2)$, as well as that of the hyperparameters $\pi(\mu_{\mathcal{R}_0,C}, \sigma_{\mathcal{R}_0,C}, \mu_{\rho,C}, \sigma_{\rho,C}|\mathcal{D}2)$.

We then analysed dataset $\mathcal{D}3$ using a hierarchical structure as above and introducing relative transmissibility and reporting of ZIKV with respect to CHIKV as follows:

Table 1. Prior distributions design and choices for modelling a Zika virus outbreak in a given island i of the French West Indies.

	Non-informative (NI)	Regional (R)*	Local (L)
$\mathcal{R}_{0,Z,i}$	Gamma$(1, 0.2)$	$\pi(\beta_{\mathcal{R}_0}\|\mathcal{D}3)\times$ $\int \varphi\left(\frac{\mathcal{R}_{0,C}-\mu_{\mathcal{R}_0,C}}{\sigma_{\mathcal{R}_0,C}}\right)\pi(\mu_{\mathcal{R}_0,C},\sigma_{\mathcal{R}_0,C}\|\mathcal{D}2)\,d\mu d\sigma$	$\pi(\beta_{\mathcal{R}_0}\|\mathcal{D}3)\times$ $\pi(\mathcal{R}_{0,C,i}\|\mathcal{D}2)$
$\rho_{Z,i}$	Beta$(1, 1)$	$\pi(\beta_\rho\|\mathcal{D}3)\times$ $\int \varphi\left(\frac{\mathcal{R}_{0,C}-\mu_{\mathcal{R}_0,C}}{\sigma_{\mathcal{R}_0,C}}\right)\pi(\mu_{\mathcal{R}_0,C},\sigma_{\mathcal{R}_0,C}\|\mathcal{D}2)\,d\mu d\sigma$	$\pi(\beta_\rho\|\mathcal{D}3)\times$ $\pi(\rho_{C,i}\|\mathcal{D}2)$
$\phi_{Z,i}$	half-Cauchy$(0, 2.5)$	half-Cauchy$(0, 2.5)$	half-Cauchy$(0, 2.5)$

*φ is the standard normal distribution pdf.

$\mathcal{R}_{0,Z} = \beta_{\mathcal{R}_0}\mathcal{R}_{0,C}$ and $\rho_Z = \beta_\rho\rho_C$. Here, we obtained the posterior distributions for the relative transmissibility $\pi(\beta_{\mathcal{R}_0}|\mathcal{D}3)$ and the relative reporting rate $\pi(\beta_\rho|\mathcal{D}3)$ of ZIKV with respect to CHIKV.

In a third step, these distributions were combined to obtain the R and L prior probability distributions as reported in Table 1. The L prior distributions summarized all information known for a specific island, computed as the product of island specific distributions and relative characteristics of ZIKV vs. CHIKV. The R prior distribution was based on the island-averaged distribution (i.e. the top level): this prior summarized information for a "typical" island of the French West Indies, rather than for a specific island, by averaging over the distribution of the hyperparameters.

Fitting and predicting ZIKV epidemics

We fitted model 1 to ZIKV data separately in Martinique, Guadeloupe and Saint-Martin using the K first weeks of data (varying K from 5 to 104) and each *a priori* distribution (NI, R or L), to obtain posterior distributions for parameters $\mathcal{R}_{0,Z}$, ρ_Z and ϕ_Z for every combination of island, K and choice of prior.

Then, using each set of posterior distributions, epidemics were simulated forward in each island, starting from week $K+1$, using the stochastic version of model 1 initialized with the observed data until week K. We used 16,000 replicates to compute the predictive distribution of the weekly number of future incident cases and a trajectory-wise 95% prediction band [38]. Using these simulated trajectories, we also computed the predictive distributions of four indicators of operational interest, for direct comparison with observed values:

- the final epidemic size, defined as the average total incidence across all simulated trajectories;
- the peak incidence, defined as the maximal value of the upper bound of the trajectory-wise 95% prediction band, as it corresponds to the capacity needed to ensure continuity of care [39];
- the date of peak incidence, defined as the average date of peak incidence across all trajectories;
- the epidemic duration, defined as the average duration between dates "S" and "E" across all trajectories.

The predictive distributions were compared using two measures of forecasting quality: (i) accuracy, i.e. the root-mean-square difference between predicted and observed values, and (ii) sharpness, i.e. the mean width of the 95% prediction band [40].

Results

ZIKV epidemics in the French West Indies

The timecourse of the ZIKV epidemics in Guadeloupe, Martinique and Saint-Martin between December, 2015 and February, 2017 differed markedly: the initial growth was early and sudden in Martinique, while it was delayed in Guadeloupe and Saint-Martin, starting only after four months of low-level transmission (Fig. 1). The epidemic showed a sharp peak in Guadeloupe, reaching a maximal weekly incidence of 6.9 cases per 1,000 inhabitants 9 weeks after the start of the period of high epidemic activity. In Martinique and Saint-Martin, weekly incidence reached a maximum of 4.8 cases per 1,000 inhabitants after a period of 10 and 21 weeks, respectively. Conversely, the period of high epidemic activity was longer in Martinique and Saint-Martin (37 and 48 weeks, respectively), than in Guadeloupe (27 weeks). In the end, a total of about 37,000 cases were observed in Martinique (97 cases per 1,000 inhabitants), more than in Saint-Martin (90 cases per 1,000 inhabitants) and Guadeloupe (77 cases per 1,000 inhabitants).

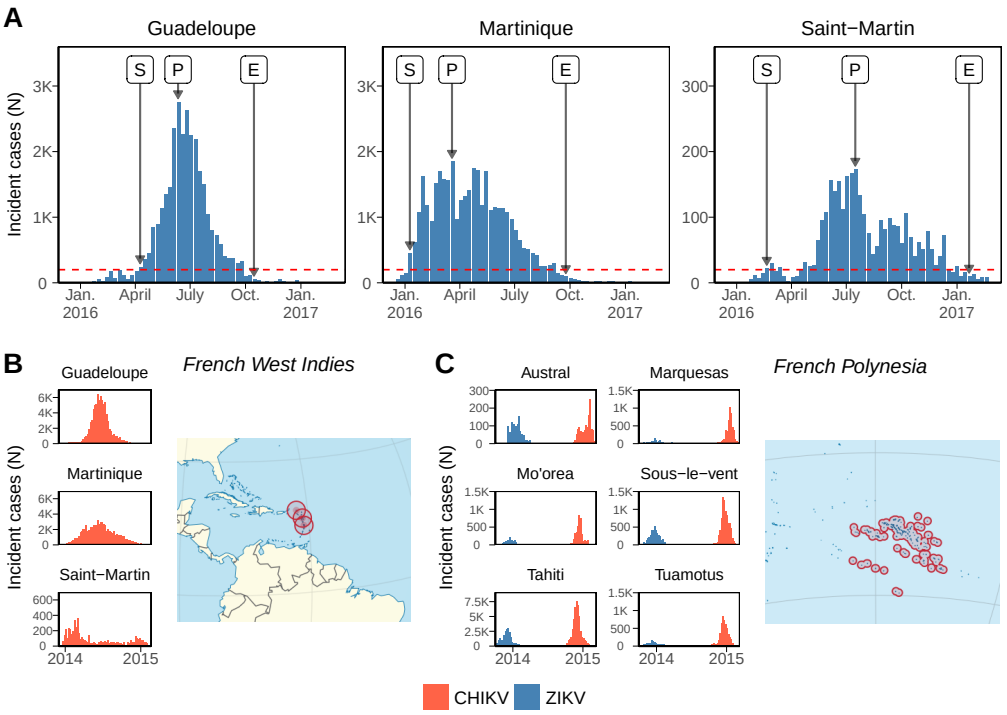

Fig 1. (A) Weekly number of Zika virus (ZIKV) cases reported by the surveillance systems in the French West Indies during 2016-2017 (dataset $\mathcal{D}1$). The dotted line shows the threshold defining high epidemic activity, "S" and "E" mark the start and the end of the period of high epidemic activity and "P" marks the date of peak incidence. (B-C) Weekly incidence (per 1,000 population) during the epidemics of Chikungunya virus (CHIKV) in the French West Indies in 2013-2015 (dataset $\mathcal{D}2$) and of ZIKV then CHIKV in French Polynesia in 2013-2015 (dataset $\mathcal{D}3$).

Prior information from past epidemics

The CHIKV epidemics observed in the same three islands of the French West Indies during 2013-2015 are shown in Fig. 1B, and the CHIKV and ZIKV epidemics observed in French Polynesia during 2013-2015 in Fig. 1C. The *a priori* distributions on the reproduction ratio and reporting rates for the ZIKV epidemics in the French West Indies defining the NI, R or L approaches are shown in Fig. 2. The R priors were wide, with 95% credible intervals between 0.5 and 2.5 for $\mathcal{R}_{0,Z}$ and between 0 and 0.30 for ρ_Z. On the contrary, the more specific L priors on $\mathcal{R}_{0,Z}$ were highly concentrated around 1.5 in Guadeloupe and 1.3 in Martinique, and ranged between 1.0 and 1.8 in Saint-Martin. Likewise, the island-specific priors on ρ_Z carried more information that their regional counterpart, peaking around 0.19 in Guadeloupe, and covering wider intervals in Martinique (0.20-0.40) and Saint-Martin (0.03-0.39).

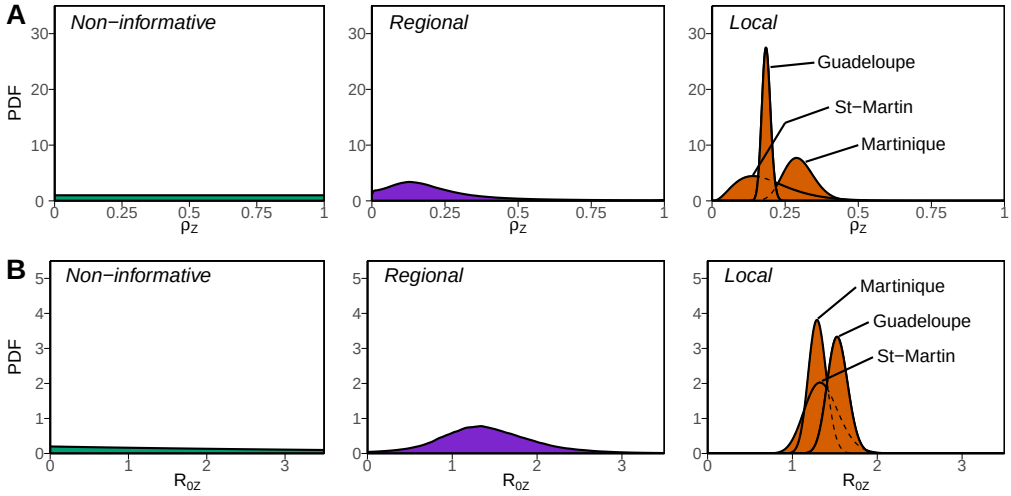

Fig 2. *A priori* distributions for the reporting rate ρ_Z (panel A) and the basic reproduction number $\mathcal{R}_{0,Z}$ (panel B) during the Zika virus epidemics in the French West Indies obtained from the analysis of historical outbreaks: (left) non-informative , island averaged (right) and (middle) island specific.

Epidemiological parameters and predictive distribution of future incidence

Fig. 3 shows the future course of the ZIKV epidemics predicted using data available two weeks after date "S" in each island, for the three choices of prior distributions. At this week, predictions with the NI priors were largely off-target, overestimating the future magnitude of the epidemic in all three islands. Using the R priors reduced the gap between forecasts and future observation. Major improvements in both *accuracy* and *sharpness* were obtained only with L priors. These results were typical of the initial phase of the epidemics, as shown in Fig. 4.Overall, the quality of the forecasts improved at each week K in all three islands and also as prior distributions brought more specific information.

The posterior distributions of the parameters built up differently as data accrued for $\mathcal{R}_{0,Z}$ and ρ_Z. For all 3 choices of *prior* distributions, the posterior distributions of $\mathcal{R}_{0,Z}$ quickly overlaid after a few points of incidence data were observed (Fig. 5A). In sharp contrast, the posterior distributions of ρ_Z could remain affected by the choice of *prior*

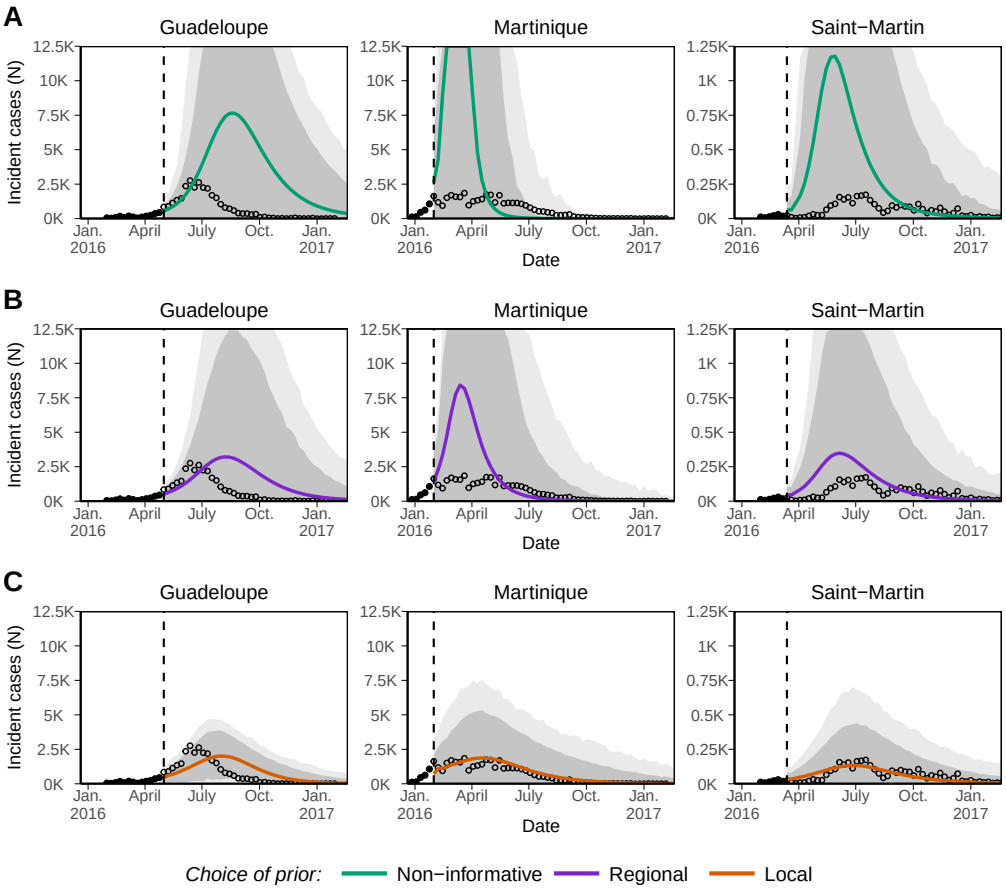

Fig 3. Predictive distribution of weekly incidence of Zika virus infections in Guadeloupe, Martinique and Saint-Martin using on either non-informative (NI, panel A), informative regional (R, panel B) or informative local (L, panel C) priors, and calibrated using data available up to the vertical dashed line (here chosen two weeks after date "S"). Continuous lines correspond to mean prediction of future incidence, dark and light grey areas to 50% and 95% prediction intervals, respectively, and circles to observed incidence.

distributions (Fig. 5B). In Martinique and Saint-Martin, ρ_Z remained essentially unidentified with the NI priors for the entire duration of the epidemic, with 95% credible intervals ranging approximately from 20 to 80%, even though the posterior mean was close to the estimates obtained with the more informative priors (around 25% at the end). In Guadeloupe, all posterior distributions for ρ_Z were similar after the peak, irrespective of the choice of priors. In all cases, informative priors allowed for a more precise estimation of ρ_Z, and this remained the case over the whole course of the epidemics.

Operational indicators

The forecasts of the four indicators of operational interest produced before peak incidence were contrasted. With NI priors, forecasts of total epidemic size overestimated the final counts by on average 3.3 times (range: 0.2 to 5.8) (Fig. 6A), with substantial

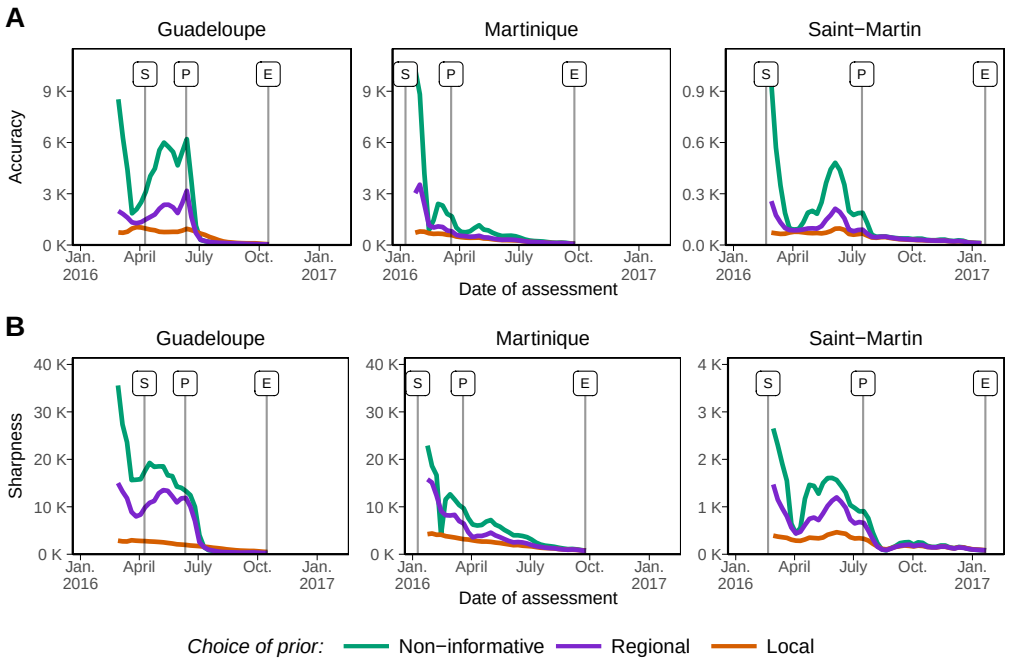

Fig 4. Accuracy (panel A, lower values indicate better accuracy) and sharpness (panel B, lower values indicate better sharpness) of the predictive distribution of future incidence based on epidemiological assessments conducted each week. Colours correspond to different *a priori* distributions on the parameters: non-informative priors or informative priors based on historical data, either considered at the regional or the local level.

variations in the forecasts from one week to the next. For instance in Martinique, projections varied from 105,000 total observed cases (95% prediction interval [95%PI]: 5,300-340,000) on February 7th to 8,100 (95%PI: 5,700-13,700) on February 14th. For the same indicator, forecasts using the R priors were only 1.7 times too high (range: 0.4 to 3.3) and those produced using L priors were only 1.1 times too high (range: 0.4 to 1.5). As a comparison, with the L priors on February 7th, the forecast of epidemic size was 47,200 (95%PI: 20,900-71,100) in Martinique, much closer to the final count of 37,400 observed cases at the end of the epidemic in this island.

Similarly, forecasts of maximal weekly (observed) incidence produced before date "P" were generally too large when using NI priors, on average 15.3 times higher than the actual maximal weekly incidence observed thereafter (range: 2.0 to 63.1) and with large fluctuations (Fig. 6B). Using informative *prior* distributions improved the forecasts, reducing the maximum predicted incidence to 7.5 times higher (range: 3.0 to 25.5) with the R priors and 2.4 times higher (range: 1.3 to 3.9) with the L priors. In all cases, forecasts of maximal incidence were never smaller than actual incidence.

Forecasting the dates of interest in the epidemics showed mixed results, with less differences depending on the choice of priors. The forecasts of the date of peak incidence were too late by on average 1.0 month (range: -3.4 to +3.9) using the NI priors, with large variations (Fig. 6C). Forecasts were only slightly better with the R priors (+0.7 months, range: -2.1 to +2.8) and the L priors (+0.9 months, range: -1.1 to +3.2), but sharper. Better forecasts were obtained for Martinique than for the other islands.

The forecasts of the total duration of the period of high epidemic activity were

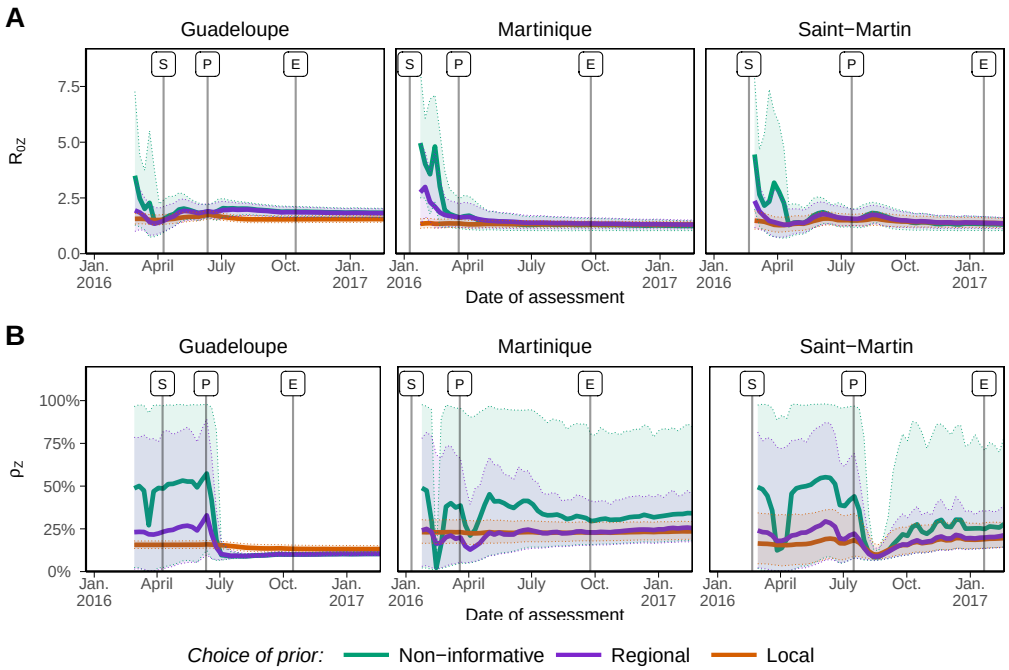

Fig 5. Posterior distributions (mean and 95% credible intervals) of the basic reproduction number $\mathcal{R}_{0,Z}$ (panel A) and the reporting rate ρ_Z (panel B) as more data was used. Colours correspond to different *a priori* distributions on the parameters: non-informative priors or informative priors based on historical data considered either at the regional or the local level.

overestimated by a factor 1.3 (range: 0.1 to 2.6) with NI priors, again with high variability from week to week (Fig. 6D). Informative priors brought a small improvement, in particular regarding the stability of the forecasts over time, overestimating the actual duration by a factor 1.2 (range: 0.7 to 2.0) with R priors and by a factor 1.1 (range: 0.9 to 1.7) with L priors.

Discussion

Obtaining reliable model-based forecasts in real time at the beginning of an epidemic is a difficult endeavour. It is however precisely during these periods that forecasts may have the most impact to guide interventions. Here, we compared several approaches to provide forecasts for ZIKV epidemics from an early point in a retrospective analysis of the outbreaks that occurred in the French West Indies in 2015-2017. We found that the accuracy and sharpness of the forecasts before peak incidence were substantially improved when *a priori* information based on historical data on past epidemics was used.

The three ZIKV outbreaks in the French West Indies provided an ideal situation to look for ways to improve prediction of *Aedes*-transmitted diseases using historical data. Indeed, CHIKV outbreaks had been observed in the same locations about two years before ZIKV. CHIKV is also transmitted by *Aedes* mosquitoes, and both viruses spread in a region where the populations were immunologically naive at first. Furthermore, all epidemics were observed by the same routinely operating GP-based surveillance networks, and the three locations benefit from a mature public health system, with easy access to medical consultation and individual means of protection. Last, pest control is

done in routine, suggesting that additional interventions for pest control would bring little effect. All the more than traditional insecticide-based measures have shown limited efficacy in this context [41, 42]. This motivated our decision to use constant parameters for transmsission and reporting over time in the modelling.

Having established the many similarities between ZIKV and CHIKV epidemics regarding transmission and reporting, we assumed that information could be transported (as defined in [43]) between diseases and between places. Bayesian approaches allow such transportation using informative *a priori* distributions on model parameters [27]. Historical data on CHIKV epidemics was therefore used to build informative *a priori* distributions on the two key parameters $\mathcal{R}_{0,Z}$ and ρ_Z. Hierarchical models are particularly adapted to this task, as they naturally pool information among several past epidemics and capture both within- and between-location variability [44]. We used separate hierarchical models to obtain information about (i) transmissibility and reporting during CHIKV outbreaks in the French West Indies and (ii) relative transmissibility and reporting between ZIKV and CHIKV in French Polynesia, rather than from a global joint model (as in [33]). This choice was made to show that prior information can be combined from separate sources in a modular way. We finally considered two versions of informative priors to capture different degrees of knowledge. The "local" priors corresponded with estimates for the previous CHIKV epidemic in the same island, therefore including island-specific epidemic drivers, such as population structure and distribution, socio-economic circumstances and environmental conditions. On the other hand, the "regional" priors were computed from the *island-averaged* hyperparameters. These summarize the diversity of past observations, making priors valid for any island typical of the West Indies, especially when no other epidemic has been observed previously.

Analyses conducted at the early stage of an epidemic using non-informative *a priori* distributions – as is often done – led to poor forecasts before the peak of incidence was reached, an observation already made in other studies [45]. Indeed, early forecasts of epidemic size were largely off-target and unstable, varying between 0.2 and 5.8 times the eventually observed total incidence. Worst case projections on maximal incidence were very imprecise, ranging between 2 and 63 times the eventually reached maximal weekly incidence. Using historical data led to a substantial increase of the quality of these forecasts from the very early stages of the epidemics. Using "local" priors, the ratio between forecasts and reality ranged between 0.4 and 1.5 for epidemic size and between 1.3 and 3.9 for maximal incidence. The less specific "regional" priors increased accuracy and sharpness as well, though to a lesser extent. However, the date of peak incidence and the date of the end of the period of high epidemic activity were only slightly improved by integrating historical data.

The posterior distributions of all forecasted quantities changed as more data was included (Fig. 4). The posterior estimates of $\mathcal{R}_{0,Z}$ were quickly similar and in good agreement with prior information. On the contrary, the reporting rate ρ_Z remained essentially unidentifiable until after the incidence peak in Guadeloupe, and to the end of the outbreak in Martinique and Saint-Martin. This suggests that prior information is essentially required for the reporting rate, a difficult to estimate quantity as already noted [15].

Predicting the future course of epidemics from an early point is increasingly seen as a problem of interest [8, 9, 46], and forecasting challenges have been set up for influenza [47], for Ebola [45] and for Chikungunya [48]. Comparing and systematically evaluating models' forecasting performances is still at the beginning. As of now, comparisons targeted the merits of different models including exponential growth models, sigmoid models, or more detailed epidemic mechanistic models [45]. Our work provides a complementary approach where information is derived from past epidemics

using hierarchical models to obtain accurate parameter ranges, bringing along increased reliability of early forecasts. Only few works used hierarchical models for modelling multiple epidemics, for instance with the joint analysis of six smallpox epidemics [49], for transmissibility and duration of carriage in the analysis of multistrain pneumococcus carriage [50], or again for forecasting seasonal influenza [51]. The approach can easily be extended to several parameters, as we illustrated, making its use possible for all parametric models used for prediction.

The choice of appropriate historical data is the cornerstone of any such attempt. Yet little is known regarding the comparative epidemiology of diseases in the same or similar places and on the condition where transportability can be assumed. For influenza, it has been reported that the reproduction ratios in two successive flu pandemics (1889 and 1918) showed substantial correlation ($r = 0.62$) in the same US cities, even years apart [52]. For *Aedes*-transmitted diseases, comparisons of ZIKV and Dengue virus outbreaks [53] and of ZIKV with CHIKV outbreaks [33] in the same locations have highlighted similarities in the epidemic dynamics. In any case, careful consideration of all the factors that may influence transmission and reporting is needed. For example, some cases of sexual transmission have been reported for ZIKV [54, 55]. Yet, no epidemics were seen in locations without enough *Aedes* mosquitoes, as for example in metropolitan France, despite the introductions of several hundreds ZIKV infected cases [56]. This justifies that information on CHIKV can provide *prior* information of interest, while further research may precise the relative role of the different transmission modes [57]. More generally, documenting, analyzing and comparing more systematically past epidemics [58], is necessary to provide the data required to derive prior information. In particular, our approach could be applied to potential future emergences of other *Aedes*-transmitted diseases such as Mayaro virus [59], Ross River virus [60] or Usutu virus [61] in areas where ZIKV or CHIKV have been observed.

Supporting information

Fig. S1: Forecasted mean weekly number of clinical cases from analyses conducted at every point of the epidemics. Extended version of Fig. 3 (animated GIF file).

Acknowledgments

We thank both the *CIRE Antilles-Guyane* and the *Centre d'Hygiène et de Salubrité Publique de Polynésie française* for collecting the data and making it publicly available.

References

1. Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. The Lancet. 2013;382(9893):694–699. doi:10.1016/S0140-6736(13)61492-0.

2. Poletto C, Pelat C, Lévy-Bruhl D, Yazdanpanah Y, Boëlle P, Colizza V. Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. Eurosurveillance. 2014;19(23):20824. doi:10.2807/1560-7917.ES2014.19.23.20824.

3. Cauchemez S, Fraser C, Van Kerkhove MD, Donnelly CA, Riley S, Rambaut A, et al. Middle East respiratory syndrome coronavirus: quantification of the extent

of the epidemic, surveillance biases, and transmissibility. The Lancet infectious diseases. 2014;14(1):50–56. doi:10.1016/S1473-3099(13)70304-9.

4. WHO Ebola Response Team. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. N Engl J Med. 2014;371(16):1481–1495. doi:10.1056/NEJMoa1411100.

5. Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS currents. 2014;6. doi:10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288.

6. Camacho A, Kucharski A, Aki-Sawyerr Y, White MA, Flasche S, Baguelin M, et al. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. PLoS currents. 2015;7. doi:10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2.

7. Lewnard JA, Mbah MLN, Alfaro-Murillo JA, Altice FL, Bawo L, Nyenswah TG, et al. Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis. The Lancet Infectious Diseases. 2014;14(12):1189–1195. doi:10.1016/S1473-3099(14)70995-8.

8. Chowell G, Viboud C, Simonsen L, Merler S, Vespignani A. Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: lessons and the way forward. BMC Medicine. 2017;15:42. doi:10.1186/s12916-017-0811-y.

9. Chretien JP, Riley S, George DB. Mathematical modeling of the West Africa Ebola epidemic. eLife. 2015;4:e09186. doi:10.7554/eLife.09186.

10. Gomes MFC, Pastore y Piontti A, Rossi L, Chao D, Longini I, Halloran ME, et al. Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. PLoS Currents. 2014;6. doi:10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5.

11. Cauchemez S, Ledrans M, Poletto C, Quenel P, De Valk H, Colizza V, et al. Local and regional spread of chikungunya fever in the Americas. Eurosurveillance. 2014;19(28):20854. doi:10.2807/1560-7917.ES2014.19.28.20854.

12. Perkins AT, Siraj AS, Ruktanonchai CW, Kraemer MUG, Tatem AJ. Model-based projections of Zika virus infections in childbearing women in the Americas. Nature Microbiology. 2016;1:16126. doi:10.1038/nmicrobiol.2016.126.

13. Zhang Q, Sun K, Chinazzi M, Piontti APy, Dean NE, Rojas DP, et al. Spread of Zika virus in the Americas. Proceedings of the National Academy of Sciences. 2017;114(22):E4334–E4343. doi:10.1073/pnas.1620161114.

14. Nishiura H, Mizumoto K, Villamil-Gómez WE, Rodríguez-Morales AJ. Preliminary estimation of the basic reproduction number of Zika virus infection during Colombia epidemic, 2015–2016. Travel medicine and infectious disease. 2016;14(3):274–276. doi:10.1016/j.tmaid.2016.03.016.

15. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015;347(6227):aaa4339. doi:10.1126/science.aaa4339.

16. Meltzer MI. Modeling in Real Time During the Ebola Response. MMWR Supplements. 2016;65(Suppl-3):85–89. doi:10.15585/mmwr.su6503a12.

17. Hsieh Y, Cheng Y. Real-time Forecast of Multiphase Outbreak. Emerg Infect Dis. 2006;12(1):122–127. doi:10.3201/eid1201.050396.

18. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review. Physics of Life Reviews. 2016;18:66–97. doi:10.1016/j.plrev.2016.07.005.

19. Willem L, Verelst F, Bilcke J, Hens N, Beutels P. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). BMC Infectious Diseases. 2017;17:612. doi:10.1186/s12879-017-2699-8.

20. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74. doi:10.1371/journal.pmed.0050074.

21. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(51):21484–21489. doi:10.1073/pnas.0906910106.

22. Tizzoni M, Bajardi P, Poletto C, Ramasco JJ, Balcan D, Gonçalves B, et al. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. BMC Medicine. 2012;10(1):165. doi:10.1186/1741-7015-10-165.

23. Wesolowski A, Erbach-Schoenberg Ez, Tatem AJ, Lourenço C, Viboud C, Charu V, et al. Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. Nature Communications. 2017;8(1):2069. doi:10.1038/s41467-017-02064-4.

24. Charu V, Zeger S, Gog J, Bjørnstad ON, Kissler S, Simonsen L, et al. Human mobility and the spatial transmission of influenza in the United States. PLOS Computational Biology. 2017;13(2):e1005382. doi:10.1371/journal.pcbi.1005382.

25. Vink MA, Bootsma MC, Wallinga J. Serial intervals of respiratory infectious diseases: a systematic review and analysis. Am J Epidemiol. 2014;180(9):865–875. doi:10.1093/aje/kwu209.

26. Lessler JT, Ott CT, Carcelen AC, Konikoff JM, Williamson J, Bi Q, et al. Times to key events in the course of Zika infection and their implications: a systematic review and pooled analysis. Bull World Health Organ. 2016;94(11):841–849. doi:10.2471/BLT.16.174540.

27. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA; 2014.

28. CIRE Antilles Guyane. Point épidémiologique N°2; 2017. Available from: `http://invs.santepubliquefrance.fr/fr/content/download/134689/483872/version/140/file/pe_zika_antilles_guyane_230216.pdf`.

29. Ledrans M, Subissi L, Cassadou S, Adelaide Y, Aubert L, Barrau M, et al. Dynamique et ampleur des épidémies de Zika en Martinique et en Guadeloupe de décembre 2015 à septembre 2016. Bulletin de Veille Sanitaire Antilles-Guyane. 2016;4:23–28.

30. CIRE Antilles Guyane. Point épidémiologique N °2; 2015. Available from: `http://www.invs.sante.fr/fr/content/download/104810/376847/version/85/file/pe_chikungunya_antilles_060315.pdf`.

31. Direction de la santé, Bureau de veille sanitaire. Surveillance et veille sanitaire en Polynésie Française; 2015. Available from: http://www.hygiene-publique.gov.pf/IMG/pdf/bss_sem_9-10-2015-version_mars_2015.pdf.

32. Centre d'hygiène et de salubrité publique de Polynésie française. Surveillance de la dengue et du zika en Polynésie française; 2014. Available from: http://www.hygiene-publique.gov.pf/IMG/pdf/bulletin_dengue_28-03-14.pdf.

33. Riou J, Poletto C, Boëlle PY. A comparative analysis of Chikungunya and Zika transmission. Epidemics. 2017;19:43–52. doi:10.1016/j.epidem.2017.01.001.

34. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. Journal of Statistical Software. 2015;doi:10.18637/jss.v076.i01.

35. Stan Development Team. RStan: the R interface to Stan; 2016. Available from: http://mc-stan.org/.

36. R Core Team. R: A Language and Environment for Statistical Computing; 2017. Available from: https://www.R-project.org/.

37. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics. 2008;2(4):1360–1383. doi:10.1214/08-AOAS191.

38. Kolsrud D. Time-simultaneous prediction band for a time series. Journal of Forecasting. 2007;26(3):171–188. doi:10.1002/for.1020.

39. Andronico A, Dorléans F, Fergé JL, Salje H, Ghawché F, Signate A, et al. Real-Time Assessment of Health-Care Requirements During the Zika Virus Epidemic in Martinique. American Journal of Epidemiology. 2017;186(10):1194–1203. doi:10.1093/aje/kwx008.

40. Gneiting T, Katzfuss M. Probabilistic forecasting. Annual Review of Statistics and Its Application. 2014;1:125–151. doi:10.1146/annurev-statistics-062713-085831.

41. Ferguson NM, Cucunubá ZM, Dorigatti I, Nedjati-Gilani GL, Donnelly CA, Basáñez MG, et al. Countering the Zika epidemic in Latin America. Science. 2016;353(6297):353–354. doi:10.1126/science.aag0219.

42. Marcombe S, Darriet F, Tolosa M, Agnew P, Duchon S, Etienne M, et al. Pyrethroid Resistance Reduces the Efficacy of Space Sprays for Dengue Control on the Island of Martinique (Caribbean). PLoS Negl Trop Dis. 2011;5(6):e1202. doi:10.1371/journal.pntd.0001202.

43. Pearl J, Bareinboim E. External validity: From do-calculus to transportability across populations. Statistical Science. 2014;29(4):579–595. doi:10.1214/14-STS486.

44. Gelman A. Multilevel (Hierarchical) modelling: What it Can and Cannot Do. Technometrics. 2006;48(3):432–435. doi:10.1198/004017005000000661.

45. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, et al. The RAPIDD Ebola Forecasting Challenge: Synthesis and Lessons Learnt. Epidemics. 2017;(in press). doi:10.1016/j.epidem.2017.08.002.

46. Chowell G, Hincapie-Palacio D, Ospina J, Pell B, Tariq A, Dahal S, et al. Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. PLoS Currents. 2016;8. doi:10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583.

47. CDC. New Flu Activity Forecasts Available for 2016-17 Season; CDC Names Most Accurate Forecaster for 2015-16; 2016. Available from: `https://www.cdc.gov/flu/spotlights/flu-activity-forecasts-2016-2017.htm`.

48. DARPA. Forecasting Chikungunya Challenge, Challenge Description on InnoCentive Site; 2014. Available from: `https://www.innocentive.com/ar/challenge/9933617`.

49. Elderd BD, Dwyer G, Dukic V. Population-level differences in disease transmission: A Bayesian analysis of multiple smallpox epidemics. Epidemics. 2013;5(3):146–156. doi:10.1016/j.epidem.2013.07.001.

50. Cauchemez S, Temime L, Guillemot D, Varon E, Valleron AJ, Thomas G, et al. Investigating heterogeneity in pneumococcal transmission: a Bayesian MCMC approach applied to a follow-up of schools. Journal of the American Statistical Association. 2006;101(475):946–958. doi:10.1198/016214506000000230.

51. Osthus D, Gattiker J, Priedhorsky R, Del Valle SY. Dynamic Bayesian Influenza Forecasting in the United States with Hierarchical Discrepancy. arXiv preprint arXiv:170809481. 2017;.

52. Valleron AJ, Cori A, Valtat S, Meurisse S, Carrat F, Boëlle PY. Transmissibility and geographic spread of the 1889 influenza pandemic. Proceedings of the National Academy of Sciences. 2010;107(19):8778–8781. doi:10.1073/pnas.1000886107.

53. Funk S, Kucharski AJ, Camacho A, Eggo RM, Yakob L, Edmunds WJ. Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. PLoS Negl Trop Dis. 2016;10(2):e0005173. doi:10.1371/journal.pntd.0005173.

54. Althaus CL, Low N. How Relevant Is Sexual Transmission of Zika Virus? PLOS Medicine. 2016;13(10):e1002157. doi:10.1371/journal.pmed.1002157.

55. Allard A, Althouse BM, Hébert-Dufresne L, Scarpino SV. The risk of sustained sexual transmission of Zika is underestimated. PLOS Pathogens. 2017;13(9):e1006633. doi:10.1371/journal.ppat.1006633.

56. Septfons A, Leparc-Goffart I, Couturier E, Franke F, Deniau J, Balestier A, et al. Travel-associated and autochthonous Zika virus infection in mainland France, 1 January to 15 July 2016. Eurosurveillance. 2016;21(32):30315. doi:10.2807/1560-7917.ES.2016.21.32.30315.

57. Kim CR, Counotte M, Bernstein K, Deal C, Mayaud P, Low N, et al. Investigating the sexual transmission of Zika virus. The Lancet Global Health. 2018;6(1):e24–e25. doi:10.1016/S2214-109X(17)30419-9.

58. van Panhuis WG, Grefenstette J, Jung SY, Chok NS, Cross A, Eng H, et al. Contagious Diseases in the United States from 1888 to the Present. N Engl J Med. 2013;369(22):2152–2158. doi:10.1056/NEJMms1215400.

59. Long KC, Ziegler SA, Thangamani S, Hausser NL, Kochel TJ, Higgs S, et al. Experimental transmission of Mayaro virus by Aedes aegypti. The American journal of tropical medicine and hygiene. 2011;85(4):750–757. doi:10.4269/ajtmh.2011.11-0359.

60. Harley D, Sleigh A, Ritchie S. Ross River virus transmission, infection, and disease: a cross-disciplinary review. Clinical microbiology reviews. 2001;14(4):909–932. doi:10.1128/CMR.14.4.909-932.2001.

61. Calzolari M, Gaibani P, Bellini R, Defilippo F, Pierro A, Albieri A, et al. Mosquito, bird and human surveillance of West Nile and Usutu viruses in Emilia-Romagna Region (Italy) in 2010. PLoS One. 2012;7(5):e38058. doi:10.1371/journal.pone.0038058.

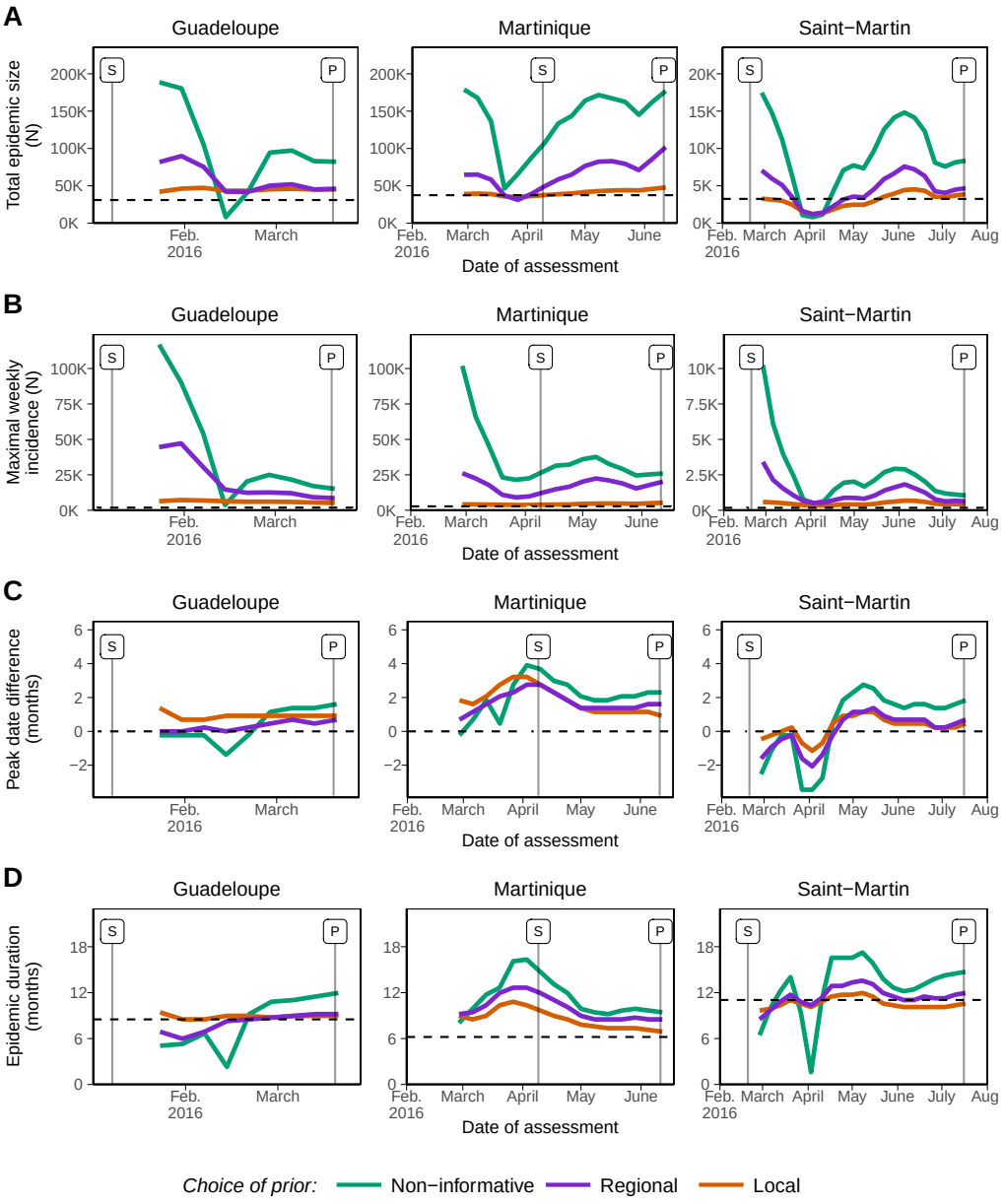

Fig 6. Early forecasts regarding four indicators of operational interest: (A) final epidemic size (total observed cases); (B) maximal weekly observed incidence; (C) date of peak incidence (difference with the date observed thereafter, in months); and (D) duration of the period of high epidemic activity (from date "S" to date "E", in months). The dashed lines represent the values values after the end of the epidemic. Colours correspond to different *a priori* distributions on the parameters: non-informative priors or informative priors based on historical data, either considered at the regional or the local level.