Time series non-Gaussian Bayesian bivariate model applied to data on HMPV and RSV: a case of Dadaab in Kenya

Raymond Nyoka[1,5]* , Thomas N. O. Achia[1], Jimmy Omony[2], Samuel M. Musili[3], Anthony Gichangi[4] and Henry Mwambi[1]

Abstract

Background: Human metapneumovirus (HMPV) have similar symptoms to those caused by the respiratory syncytial virus (RSV). The modes of transmission and dynamics of time series data still remain poorly understood. Climatic factors have long been suspected to be implicated in impacting on the number of cases for these epidemics. Currently, only a few models satisfactorily capture the dynamics of time series data of these two viruses. Our objective was to assess the presence of influence of high incidences between the viruses and to ascertain whether higher incidences of one virus are influenced by the other.

Methods: In this study, we used a negative binomial model to investigate the relationship between RSV and HMPV while adjusting for climatic factors. We specifically aimed at establishing the heterogeneity in the autoregressive effect to account for the influence between these viruses.

Results: In this study, our findings showed that RSV incidence contributed to the severity of HMPV incidence. This was achieved through comparison of 12 models with different structures, including those with and without interaction between climatic factors. The models with climatic factors out-performed those without.

Conclusions: The study has improved our understanding of the dynamics of RSV and HMPV in relation to climatic cofactors thereby setting a platform to devise better intervention measures to combat the epidemics. We conclude that preventing and controlling RSV infection subsequently reduces the incidence of HMPV.

Keywords: Non-Gaussian bivariate Bayesian model, RSV, HMPV, Epidemic, Time series, Climatic factors

Background

Epidemiological knowledge of the respiratory system has been mostly related to developed countries, though the burden of respiratory virus infections (RVIs) is more manifested in developing countries with very high hospitalization and mortality rates [1]. Higher mortality is associated with increased displacement into overcrowded refugee camps [2]. The burden of RVIs is considerably high during crises times [3] and is more severe in infants [4]. Recently, Pastula et al. [5] highlighted that hospitalization for respiratory syncytial virus (RSV) is

not limited to infants but also includes adults. In 2001, HMPV was identified as a potential etiologic agent for respiratory infections [6]. A study at Queen Mary Hospital in Hong Kong showed that the peaks of HMPV and that of RSV activity occurred in spring and the early months of summer and viral diagnoses during the study period showed that RSV and HMPV had similar seasonality [7]. Guerrero et al. [8] indicate that RSV but not HMPV induces a productive infection in human monocyte-derived dendritic cells. Reinfection by RSV has a great impact on human health and may cause long-term effects on the host immune response [9]. Greensill et al. [10] detected HMPV in 21 out of 30 infants infected with severe RSV and were hospitalized requiring intensive-care unit ventilator support. Konig et al. [11] found out that 60% of the cases with HMPV had

* Correspondence: nyomus@gmail.com
[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, South Africa
[5]Nairobi, Kenya
Full list of author information is available at the end of the article

RSV. They also found that HMPV contributed to the severity of Lower respiratory tract infections (LRTIs) at a lower rate than RSV and coinfection was considered a cause of severe lower respiratory tract disease. The HMPV infections have similar symptoms to those caused by RSV [12, 13], they share similar risk factors [14] and simultaneous detection times [15]. The HMPV and RSV may cross-react directly or indirectly because they are both co-viruses to each other [16]. The correlation between RSV and HMPV in the refugee settings and even in the tropical region has not been studied. We specifically aimed at establishing the heterogeneity in the autoregressive effect to account for the influence between these viruses. The modelling of the time series events of these viruses will not only help in the prediction of their outbreaks but also in estimating which outbreaks precede each other. The results could be used by other countries in the tropical zone of Africa with similar settings to inform control measures to prevent outbreaks.

In Section 2, we show the data source and the statistical model fitting with and without climatic covariates to a bivariate time series. In Section 3, we show the applicability of the models illustrated with a real-world example and the results obtained. In Section 4, we discuss the results and finally conclude in Section 5.

Methods
Data
A surveillance system for viral respiratory illnesses that included RSV and HMPV was implemented in a refugee camp in Dadaab located in northeastern province of Kenya from September 2007 to August 2011. Both paediatric and adult patients presenting to a medical unit and who met the case definition for influenza-like illness or severe acute respiratory infection were enrolled in the surveillance. Laboratory confirmed test results for RSV and HMPV were obtained after adults and guardians of all minors filled a consent form. The number of laboratory-confirmed cases was recorded every day. In this analysis, only the monthly counts of RSV and HMPV cases among children younger than 5 years were considered. Local weather and climatic data from a neighboring weather station were obtained from the World Meteorological Organization's (WMO's), World Weather Watch Program, according to WMO Resolution 40 (Cg-XII) (available at http://www7.ncdc.noaa.gov/CDO/cdo). The meteorological dataset was recorded on a daily basis and aggregated monthly for the purpose of this analysis. The variables included the mean temperature, mean dew point for the day (both in ^{0}F), the mean sea level pressure for the day in millibars, the mean visibility for the day in miles, the mean wind speed for the day in knots, the minimum and maximum

temperature (°F) reported during the day and the total precipitation (in inches).

Statistical modeling
In this paper, we used surveillance data aggregated by month in a time series model and the negative binomial distribution to address the issue of over-dispersion. We model the relationship between the two viruses namely, RSV and HMPV. Meteorological variables were included in the model to help assess for serial correlation. Held et al. [17] suggested that environmental factors can be incorporated into these models to improve model fit to data and predictions. These models help to assess the presence of influence of high incidences between the viruses and whether higher incidences of one virus are influenced by another. They also aid in evaluating if an epidemic component can be isolated within or between the viruses and how the autoregressive component captures the residual temporal dependence in the time series after adjusting for seasonal effects. Modeling count data is faced with many challenges since count outcomes do not meet the usual normality assumption required of many standard statistical tests. Typical log-transformation to induce normality does not often work or categorization of the outcome may lead to loss of information as described by O'Hara and Kotze [18]. The most commonly used models to study the dynamics of epidemics and predict future outbreaks using count data are the Poisson [19] and the negative binomial distributions [20]. We modelled the time-evolution of two epidemics using a bivariate approach suggested by Held et al. [17]. We assume that we have $i = 1, \ldots, m$ 'viruses' and denote with y_{it} the number of cases in virus i at time t. The general model for the multiple time series of count events $\{y_{it}, i = 1, \ldots, m; t = 1, \ldots, T\}$ for virus type i at time t assumes a Poisson distribution with conditional mean μ_{it} given by

$$\log(\ddot{E}_{it}) = \lambda_{i,t-1} y_{i,t-1} + \phi_{i,t-1} \sum_{j \neq i} \omega_{ij} y_{j,t-1} + \eta_{i,t} \nu_{it}.$$

(1)

It holds $\text{VAR}(y_{i,\,t}|y_{i,\,t-1}) = \text{E}(y_{i,\,t}|y_{i,\,t-1}) = \mu_{it}$. Hence, in the case of a conditional Poisson response model the conditional mean μ_{it}, is identical to the conditional variance δ of the observed process.

In model 1, $\lambda_{i,\,t-1}$ is the autoregressive parameter representing the proportion of epidemic cases from the total number of cases for virus type i at time t. When $\lambda_{i,\,t-1} \geq 1$ (an outbreak occurs) there is an influx of the endemic cases and $\lambda_{i,\,t-1} < 1$ means the process is stable (no outbreak occurs). The $\phi_{i,\,t-1}$ quantifies the influence of virus type j on i; $\eta_{i,\,t}$ represents the monthly varying population counts of virus type i at time t (treated as an

offset term in the model) and v_{it} is the endemic component that explains the baseline incidence rate of cases as subsequently shown in eq. (5). The variable $y_{j, t-1}$ denotes the number of cases observed in virus type j at time $t-1$. ω_{ij} is the weighting indicator and is equal to 1 if pathogens j and i have an autoregressive effect on each other and 0 otherwise.

This model is aggregation consistent where the aggregated counts $y_t = \sum_{i=1}^{m} y_{it}$ have the mean,

$$\log(\mu_t) = \lambda y_{t-1} + \phi_{t-1} Z_{t-1} + \eta_t v_t,$$

where, $\quad Z_{t-1} = \sum_{j \neq i} \omega_{ij} y_{j,t-1}, \eta_t = \sum_{i=1}^{m} \eta_{i,t}, \phi_t = \sum_{i=1}^{m} \phi_{i,t}, v_t = \sum_{i=1}^{m} v_{i,t}$. So, the parameter λ has the same interpretation for the aggregated counts similar to the counts y_{it}.

In the presence of over-dispersion, the Poisson model is replaced by a negative binomial model where the conditional mean remains unchanged but the variance δ is modified to $\mu_t(1 + \mu_t \psi)$ with over-dispersion parameter $\psi > 0$. The extent of over-dispersion is captured by how far the term ψ deviates from zero. An extensive discussion on handling over-dispersion can be found in the work of Ver Hoef and Boveng [21]. We are interested in two types of viruses transmitted through the same route, i.e. respiratory illness. Let $x_{k, t-1}$ denote climatic covariates with τ_k coefficients in the model and $k = 1, \ldots, K$ covariates. In the model, it is assumed that the cases follow a negative binomial distribution, $y_t \mid y_{t-1} \sim \text{NegBin}(\mu_t, \psi)$, with conditional mean

$$\log(\mu_t) = \lambda_{t-1} y_{t-1} + \tau_k x_{k,t-1} + \phi_{t-1} Z_{t-1} \\ + \exp(\eta_t) \qquad (2)$$

and conditional variance

$$\mu_t(1 + \mu_t \psi). \qquad (3)$$

The incidence of the disease μ_t was additively decomposed into two parts. The first part,

$$\xi_t = \lambda_{t-1} y_{t-1} + \tau_k x_{k,t-1} + \phi_{t-1} z_{t-1}, \qquad (4)$$

is the epidemic component explaining the outbreaks or irregularities in the data including the interaction between the viruses. The second part is $v_{i, t}$ which is expressed in log-scale as

$$\log(v_{it}) = \alpha_i + \sum_{s=1}^{S} \{\gamma_s \sin(\omega_s t) + \delta_s \cos(\omega_s t)\}. \qquad (5)$$

The endemic and epidemic components of the time series were explored and studied allowing for the separation of the regular pattern from irregular ones in estimating the epidemic peaks. The parameter α_i allows for different incidence levels of the viruses and S is the virus specific number of harmonic waves. The term in curly brackets captures seasonal variations. The γ_s and δ_s are the seasonal parameters while $\omega_s = 2\pi s/12$ for monthly data are the Fourier frequencies.

Likelihood and posterior distribution

The counts y_t, conditional on the previous observation y_{t-1} (Only lag one was applied in our case because more than one lag did not fit the data well) are assumed to follow a Negative binomial distribution with mean

$$\mu_t \theta \equiv \mu_t = \xi + v, \qquad (6)$$

where $\theta = (\theta_1, \ldots, \theta_m, \psi_1, \ldots, \psi_m)^T$ The log-likelihood of the observation y_t is given as

$$l(\theta) = \sum_t l_t(\theta, \psi) \qquad (7)$$

and the likelihood as,

$$f(y_t | \theta) = \exp \left\{ \sum_t l_t(\theta, \psi) \right\}, \qquad (8)$$

where,

$$l_t(\theta, \psi) \propto \log \Gamma\left(y_t + \frac{1}{\psi}\right) - \log\Gamma\left(\frac{1}{\psi}\right) \\ + \frac{1}{\psi} \log\left(\frac{1}{1 + \psi\mu_t(\theta)}\right) \\ + y_t \log\left(\frac{\psi\mu_t(\theta)}{1 + \psi\mu_t(\theta)}\right), \qquad (9)$$

and $\Gamma(.)$ is the gamma function and ψ and τ are the dispersion parameters. The gamma priors are assumed for ψ and τ,

$$\psi \sim Ga\left(\alpha_\psi, \beta_\psi\right),$$

$$\tau \sim Ga\left(\alpha_\tau, \beta_\tau\right).$$

The virus dependent effects α_i are assumed to be independent and normally distributed with a large variance,

$$\alpha = (\alpha_1, \ldots, \alpha_I) \sim N\left(0, \sigma_\alpha^2 I\right), \sigma_\alpha^2 = 10^6,$$

where I is an identity matrix. All model parameters are non-negative and therefore we propose gamma prior distributions for them. The rate parameters λ_t assumes independent gamma priors with gamma hyperpriors on the second parameter,

$$\lambda_t \sim Ga\left(\alpha_\lambda, \beta_\lambda\right) \text{ and } \beta_\lambda \sim Ga(a, b).$$

Where we use $\alpha_\lambda = 1$, $a = 10$ and $b = 10$, with values for α_λ, a and b chosen arbitrarily.

Independent normal priors are assumed for γ and δ,

$$\gamma = (\gamma_1, \ldots, \gamma_I) \sim N\left(0, \sigma_\gamma^2 I\right), \sigma_\gamma^2 = 10^6,$$

Nyoka *et al. BMC Public Health*   (2019) 19:807

Page 4 of 14

Table 1 Simulation results including Parameter estimates, Standard errors and measure of model Goodness of Fit

Parameter	Model1 ($\psi=0\ \lambda=0$)	Model2 ($\psi=0\ \lambda\neq0$)	Model3 ($\psi\neq0\ \lambda=0$)	Model4 ($\psi\neq0\ \lambda\neq0$)
ψ_1	–	–	0.0000 (0.0000)	0.0000 (0.0000)
ψ_2	–	–	0.0000 (0.0000)	0.0000 (0.0001)
λ_1	–	0.1730 (0.3135)	–	0.1743 (0.3072)
λ_2	–	0.4337 (0.2010)	–	0.4482 (0.2115)
φ_1	0.4727 (0.2262)	0.4586 (0.2300)	0.4726 (0.3092)	0.4585 (0.2300)
φ_2	0.8123 (0.0420)	0.0963 (0.2204)	0.3034 (0.2204)	0.1485 (0.2424)
AIC	1644.14	1626.58	1636.92	1630.33

$$\boldsymbol{\delta}=(\delta_1,...,\delta_I)\sim N\left(0,\sigma_\delta^2 I\right),\sigma_\delta^2=10^6.$$

The parameter ϕ_t assumes gamma priors, $\phi_t\sim Ga(\alpha_\phi,\beta_\phi)$.

The posterior distribution is therefore given as,

$$f(\boldsymbol{\theta}|\boldsymbol{y}_t)\propto f(\boldsymbol{y}_t|\boldsymbol{\theta})f(\boldsymbol{\theta}),$$

which can be expressed as,

$$f(\boldsymbol{\theta}|\boldsymbol{y}_t)\propto exp\left\{\sum_t l_t(\boldsymbol{\theta},\boldsymbol{\psi})\right\}\times\prod_{s=1}^{S}e^{-\frac{1}{2}\sigma_\gamma^2}\times\prod_{s=1}^{S}e^{-\frac{1}{2}\sigma_\delta^2}\times\prod_{i=1}^{m}e^{-\frac{1}{2}\sigma_{\alpha_i}^2}$$
$$\times\prod_{i=1}^{m}\lambda_i^{\alpha_{\lambda_i}-1}e^{-\beta_{\lambda_i}^{\lambda_i}}\lambda_i^{a-1}e^{-b\lambda_i}\times\prod_{i=1}^{m}\psi_i^{\alpha_{\psi_i}-1}e^{-\beta_{\psi_i}^{\psi_i}}\times\prod_{i=1}^{m}\phi_i^{\alpha_{\phi_i}-1}e^{-\beta_{\phi_i}^{\phi_i}}\times\prod_{i=1}^{m}\tau_i^{\alpha_{\tau_i}-1}e^{-\beta_{\tau_i}^{\tau_i}}.$$
$$(10)$$

Simulations

We investigated the proposed model performance on simulated data. We simulated bivariate data using a frequentist approach in R software using the package "Surveillance" previously used by Held et al. [22, 23]. We used the function "hhh4" with the class "disprog" to simulate two disease pathogen counts replicated 10,000 times. We then applied the Bayesian approach to compare different models based on varied scenarios. We considered a situation where there is the presence of overdispersion with the parameter $\psi_i\neq0$ assuming the negative binomial distribution and where $\psi_i=0$ assumes the Poisson distribution. We also considered the presence and absence of the parameter λ_i (the 'epidemic' component) to evaluate temporal dependence. In this simulation, we disregarded the linear trend. It is evident from Table 1 that the simulation results show that $\psi_i=0$ and therefore the best performing model is the Poisson (model 2) with the presence of the epidemic component having the least AIC = 1626.58. This implies that in the simulated data there was no overdispersion but rather temporal dependence.

Application on data

Let $\{y_{it}, i=1,2; t=1,\dots,48\}$ be the time series of virus counts for RSV (y_{1t}) and HMPV (y_{2t}) over the 48 months study time-frame. There were only two oscillations in a year for each of the two viruses to complete a cycle hence two harmonic waves (s = 2) were included in the model. The bivariate model for the two time series is therefore:

Table 2 Models of the epidemic part $\xi_{i,\,t}$ with assumptions made on interactions between the viruses with and without the climatic factors

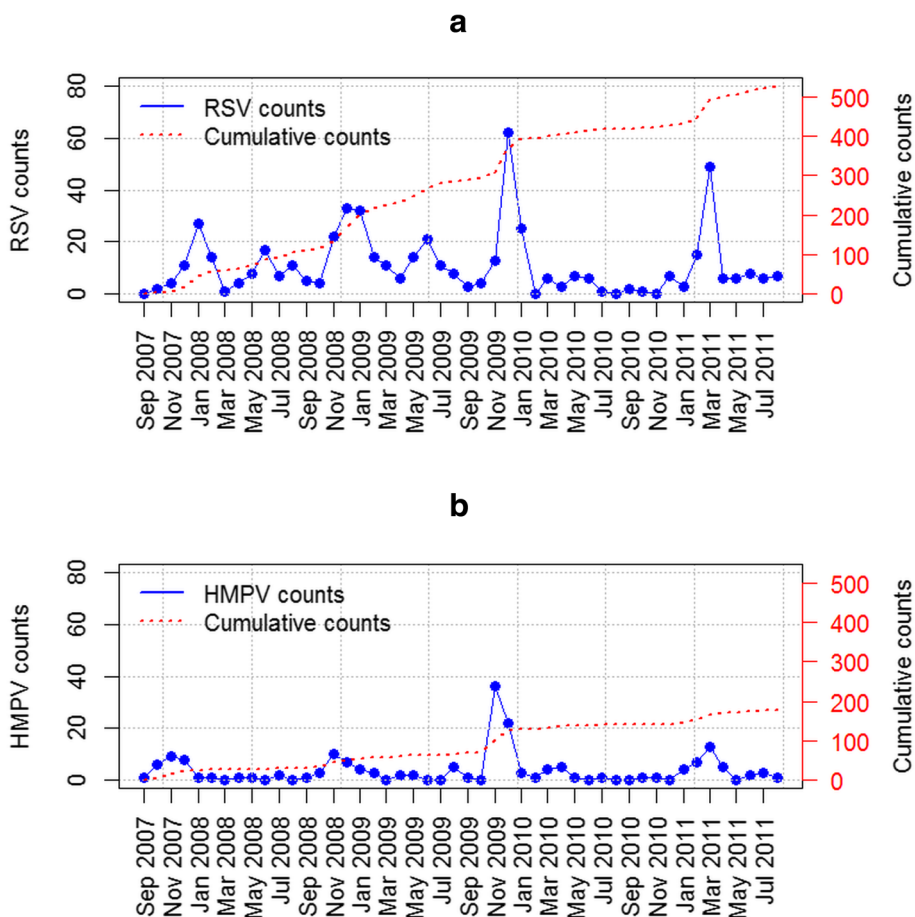| Model	$\xi_{i,\,t}$ (with climatic factors)	$\xi_{i,\,t}$ (without climatic factors)
1	$\lambda y_{i,\,t-1}+\tau_{i,\,k}x_{k,\,t-1}$	$\lambda y_{i,\,t-1}$
2	$\lambda y_{i,t-1}+\varphi\sum_{j\neq i}w_{ji}y_{j,t-1}+\tau_{i,k}x_{k,t-1}$	$\lambda y_{i,t-1}+\varphi\sum_{j\neq i}w_{ji}y_{j,t-1}$
3	$\lambda_i y_{i,\,t-1}+\tau_{i,\,k}x_{k,\,t-1}$	$\lambda_i y_{i,\,t-1}$
4	$\lambda_i y_{i,t-1}+\sum_{j\neq i}w_{ji}\varphi_i y_{j,t-1}+\tau_{i,k}x_{k,t-1}$	$\lambda_i y_{i,t-1}+\sum_{j\neq i}w_{ji}\varphi_i y_{j,t-1}$
5	$\lambda_{i,\,t-1}y_{i,\,t-1}+\tau_{i,\,k}x_{k,\,t-1}$	$\lambda_{i,\,t-1}y_{i,\,t-1}$
6	$\lambda_{i,t-1}y_{i,t-1}+\sum_{j\neq i}w_{ji}\varphi_{i,t-1}y_{j,t-1}+\tau_{i,k}x_{k,t-1}$	$\lambda_{i,t-1}y_{i,t-1}+\sum_{j\neq i}w_{ji}\varphi_{i,t-1}y_{j,t-1}$

Fig. 1 The monthly counts of epidemics (**a**) RSV and (**b**) HMPV plotted against time. The cumulative counts of HMPV cases were approximately 2.5 times less than the RSV counts for the same time-frame

$$\log\begin{pmatrix} \ddot{E}_{1,t} \\ \ddot{E}_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_{1,t-1} & \phi_{1,t-1} \\ \phi_{2,t-1} & \lambda_{2,t-1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix}$$

$$+ \begin{pmatrix} \tau_{1,1} & \tau_{1,2} & \tau_{1,3} & \tau_{1,4} \\ \tau_{2,1} & \tau_{2,2} & \tau_{2,3} & \tau_{2,4} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \\ x_{4,t-1} \end{pmatrix}$$

$$+ \eta_t \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix},$$

where

$v_{1,t} = \alpha_1 + r_{1,1}\sin(\omega_1 t) + \delta_{1,1}\cos(\omega_1 t) + r_{1,2}\sin(\omega_2 t) + \delta_{1,2}\cos(\omega_2 t),$

$v_{2,t} = \alpha_2 + r_{2,1}\sin(\omega_1 t) + \delta_{2,1}\cos(\omega_1 t) + r_{2,2}\sin(\omega_2 t) + \delta_{2,2}\cos(\omega_2 t)$

and $x_{1,\,t-1}$, $x_{2,\,t-1}$, $x_{3,\,t-1}$ and $x_{4,\,t-1}$ are the climatic factors representing rainfall, wind speed, mean dew point and visibility respectively. The term η_t corresponds to an offset term in the model (the monthly varying population counts at time t).

The models were compared for their fit to the epidemic data. Naturally, models are compared for their performance based on the ability to fit well on the data and their reliability in predicting future epidemic outbreaks. Fundamentally, in our model fitting to data we searched for the model that provided the best trade-off between the fit to data and the model structure complexity. Often, approaches such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are sufficient for ranking and selecting the best performing models. However, when the data is non-Gaussian and the model is Bayesian, like in our case, then the deviance information criterion (DIC) is more appropriate. For the comparison of our models we used the DIC proposed by Spiegelhalter et al. [24], specifically for Bayesian-based models and it is a Bayesian generalization of the AIC and BIC. The model with the smallest DIC value gives the better trade-off between model fit and complexity; therefore, it is considered as the model that best predicts a replication of a data set with a similar structure as that which was observed currently [25].

Fig. 2 The monthly counts of RSV and HMPV plotted against time. Overall, the epidemics coincide in timing of their occurrence peaks, especially in March 2011

To further assess the model performance with regards to the parameters, sensitivity analysis to alternative prior assumptions was performed because there are no true priors in the Bayesian analysis. In order to ensure reliable and robust results from our best model, it was crucial to verify how sensitive the resulting posteriors were for each prior input for the epidemic parameter λ_{it} and ϕ_{it}, the parameter that quantifies the influence of one virus on the other. Therefore, we assumed independent gamma priors with uniform hyper-priors on the second parameter, $\lambda_{it} \sim Ga(\alpha_\lambda \beta_\lambda)$ and $\beta_\lambda \sim Beta(a, b)$ using $\alpha_\lambda = 1$, $a = 0.5$ and $b = 0.5$. Similarly, for the influential parameter, we used the Beta distribution prior, $\phi_{it} \sim Beta(\alpha_\phi \beta_\phi)$. To our understanding, this comparison of models has not yet been done using RSV and HMPV time series data. All the models in our work were run and tested in the statistical software WinBUGS version 14. The models differed on the epidemic part $\xi_{i,\,t}$ by the assumptions made on the interactions between the viruses. We used 6 models depending on the assumptions applied as explained below with each model with a corresponding inclusion of climatic factors giving rise to a total of 12 models (Table 2). In model 1, it is assumed that the incidence rate is the same in every virus; hence, no interactions between the viruses. Model 2 assumes that there is the interaction between viruses where the sum of related viruses at the same time point has an equal rate. Models 3 and 4 are generalizations of models 1 and 2 respectively with a different rate for each virus. Models 5 and 6 generalize model 3 and 4 respectively with a different rate for each virus per time point.

The best model was then evaluated on whether; there were substantial interactions between cases of RSV and HMPV (alternatively stated as $\phi_{RSV} \neq \phi_{HMPV} \neq 0$), the existence of the influence of RSV on HMPV ($\phi_{RSV} = 0$, $\phi_{HMPV} \neq 0$), the existence of the influence of HMPV on RSV ($\phi_{HMPV} = 0$, $\phi_{RSV} \neq 0$) or there were no interactions at all ($\phi_{RSV} = \phi_{HMPV} = 0$).

Results

The monthly observed number of RSV and HMPV cases in Dadaab from September 2007 to August 2011 that were collected in the surveillance system was plotted (Fig. 1). The HMPV data shows a strong seasonality pattern as indicated by the four peaks during November of the years 2007, 2008 and 2009 while a fourth peak appears in March 2011 (Fig. 2). These HMPV peaks coincide with the RSV peaks.

Table 3 Comparison DIC values for different models

Model	1	2	3	4	5	6
DIC (with climatic factors)	490.43	558.30	559.46	558.45	502.17	173.52
DIC (without climatic factors)	549.82	541.11	548.44	536.09	571.72	744.22

Table 4 Four sub-models from the best model. The symbols "−" and "√" mean the absence and presence of interactions, respectively. Model 6 (i) no interactions between HMPV and RSV ($\varphi_{HMPV} = \varphi_{RSV} = 0$); Model 6 (ii) influence of HMPV on RSV ($\varphi_{RSV} \neq 0$, $\varphi_{HMPV} = 0$), Model 6 (iii) influence of RSV on HMPV ($\varphi_{RSV} = 0$, $\varphi_{HMPV} \neq 0$) and Model 6 (iv) interactions between HMPV and RSV ($\varphi_{HMPV} \neq \varphi_{RSV} \neq 0$)

Model	HMPV→RSV	RSV → HMPV	DIC
6(i)	−	−	543.68
6(ii)	√	−	457.61
6(iii)	−	√	112.14
6(iv)	√	√	173.52

We compared 12 models with various structures (Table 2) and the results for the DIC values are given in Table 3. Models 6 and 1 with climatic factors clearly out-perform the other models since, overall, they have lower DIC values. Model 6 with climatic factors had the least DIC value (173.52) and provided the best fit and explanation for the variation observed in the data. The models showed that the inclusion of climatic factors play an important role in the estimation of the number of cases for the two epidemics (RSV and HMPV). We further considered different scenarios on the best model with four sub-models (results are shown in Table 4).

Model 6(i) in Table 4 does not allow for interactions between HMPV and RSV ($\phi_{HMPV} = \phi_{RSV} = 0$) and its
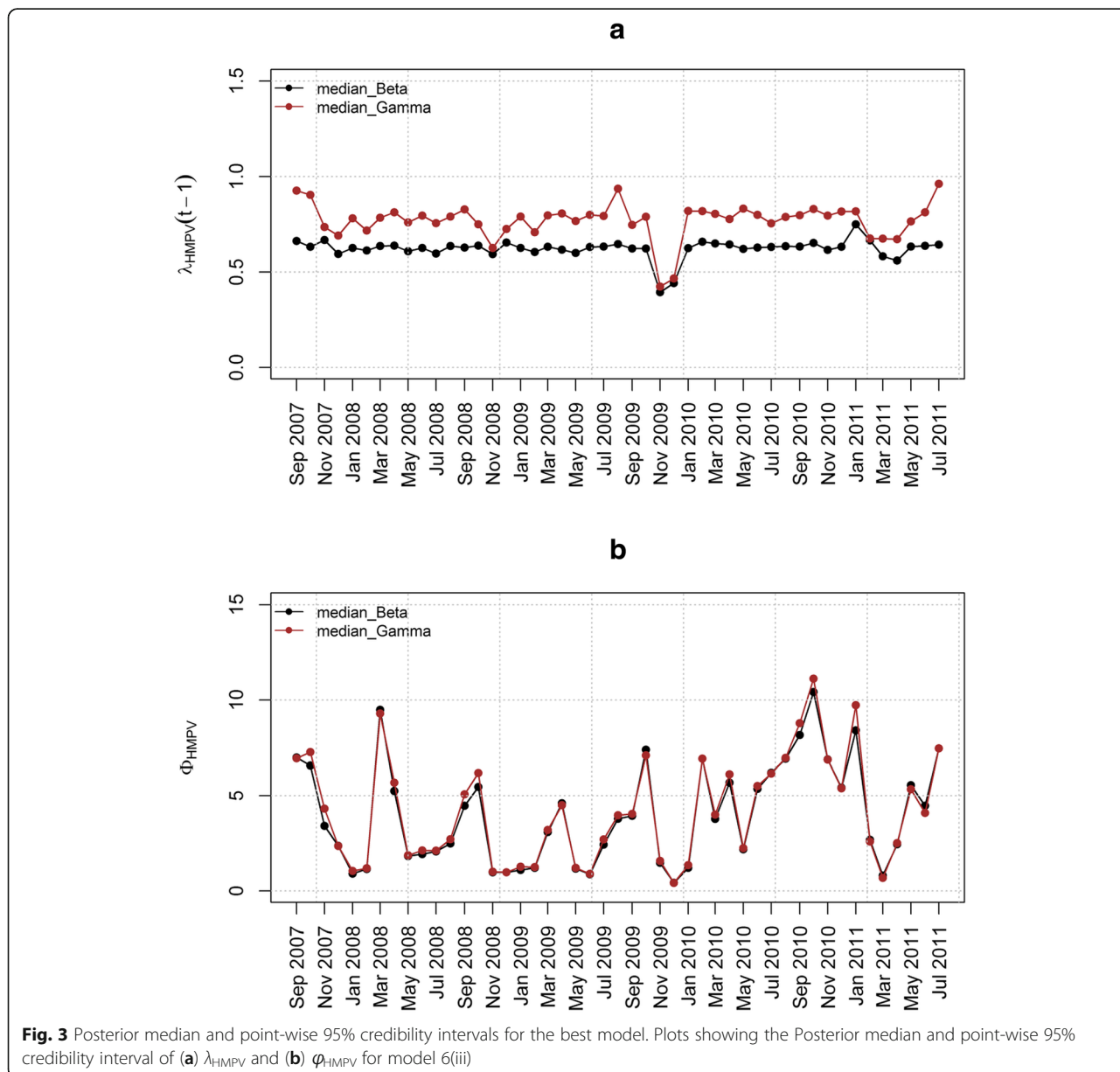

Fig. 3 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (**a**) λ_{HMPV} and (**b**) φ_{HMPV} for model 6(iii)
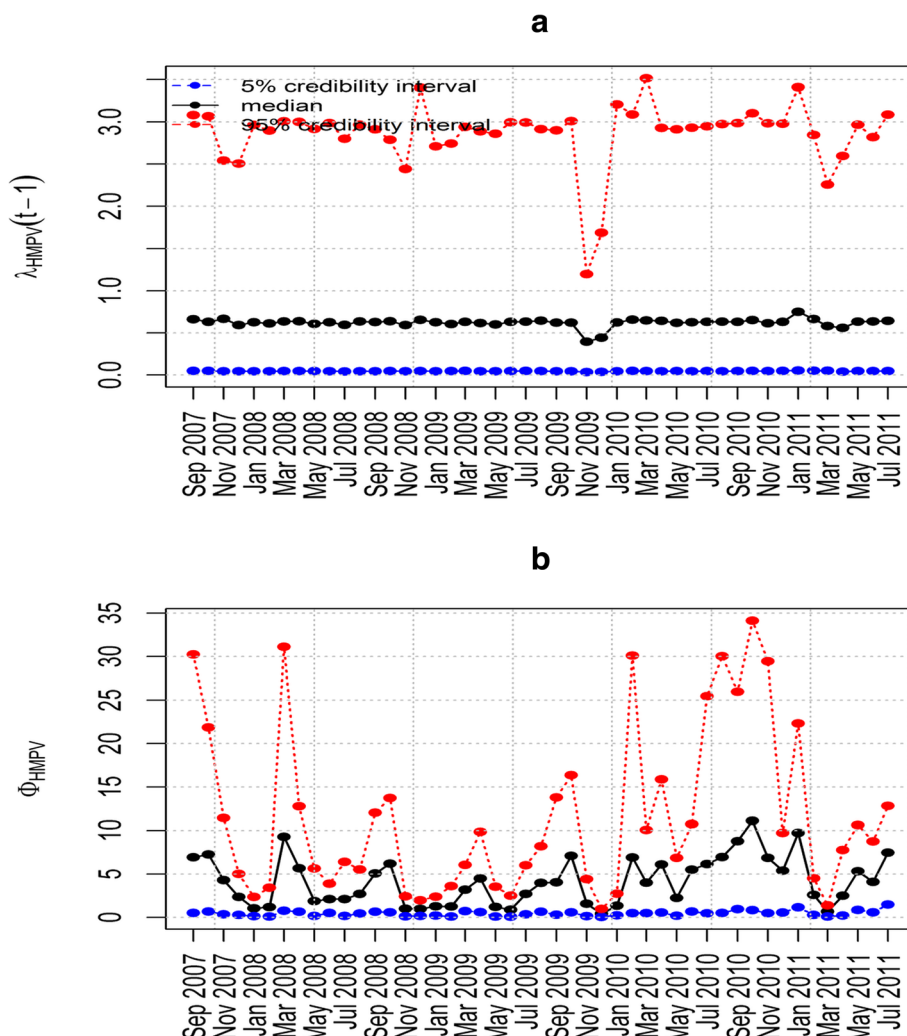
Fig. 4 Posterior median values for the priors with Gamma and Beta distributions for the best model. Plots showing the Posterior median values of (**a**) λ_{HMPV} and (**b**) φ_{HMPV} for model 6(iii). Median_Beta and median_Gamma are the posterior medians from the Beta distribution and the Gamma distribution priors respectively

Table 5 Posterior median and point-wise 95% credibility intervals for the best model

Parameter	5.0%	Median	95%
alpha1	−4.283	−3.998	−3.683
alpha2	−3.765	−3.765	−3.481
delta11	−2.564	−2.564	−1.979
delta21	−4.783	−4.783	−4.023
gamma11	−6.303	−5.653	−4.812
gamma21	−9.209	−7.965	−6.934
psi1	0.238	7.762	116.1
psi2	0.090	4.688	97.33

DIC value is 543.68. Model 6(ii) includes the influence of HMPV on RSV with the influence of RSV on HMPV equal to zero. This model yielded a DIC value of 457.61. Model 6(iii) includes the influence of RSV on HMPV where the influence of HMPV on RSV is zero. Compared to the others, this model yielded the smallest DIC value of 112.14 (Table 4). This implies that the two viruses can present as a co-infection where HMPV incidence is increased by increases in RSV. The results from sensitivity analysis shown in Fig. 3, indicates that this model is robust and insensitive to the prior distribution since its posterior distribution did not dramatically change upon altering the base prior parameter values. Model 6(iv) has both the influence of RSV on HMPV and the influence of HMPV on RSV which is the full model with a DIC value of 173.52 (Table 4). This indicates that the additional parameter (i.e., the influence of
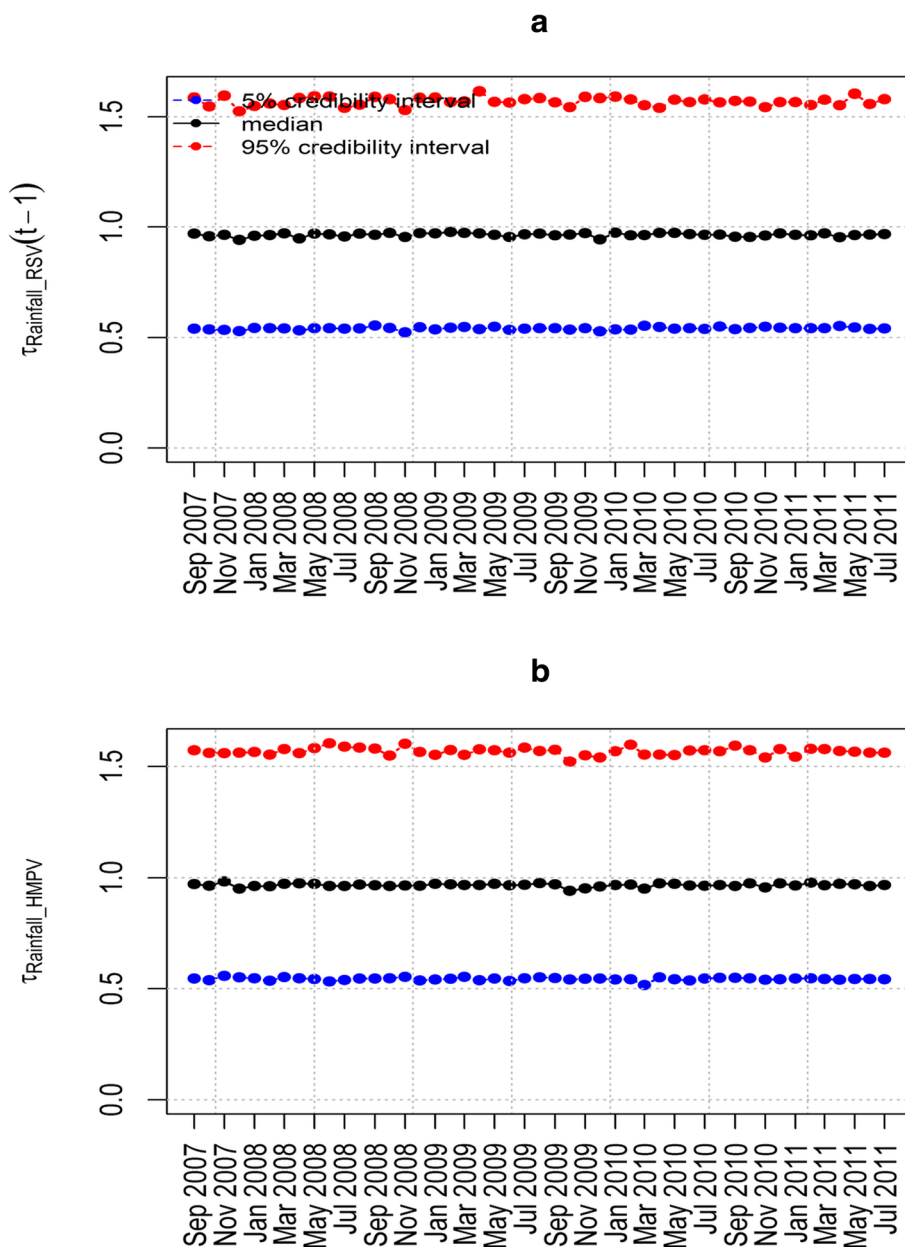
Fig. 5 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (**a**) $\tau_{Rainfall\_RSV}$ and (**b**) $\tau_{Rainfall\_HMPV}$ for model 6(iii)

HMPV on RSV) into model 6(iii) does not significantly improve the model fit to data.

The epidemic parameter λ_{HMPV} for model 6(iii) in Fig. 4(a) does not exceed the value 1. This implies that the time series is stable without a detection of an outbreak of HMPV due to the influence of RSV. Figure 4(b) shows the influence of RSV on HMPV with biannual peaks noted over the study period. The other parameters estimated in this model are shown in Table 5 that includes the posterior median and point-wise 95% credibility intervals. In particular, from

Table 5, the posterior median and the point-wise 95% credibility intervals for the over-dispersion parameters ψ_{HMPV} and ψ_{RSV} were 7.762(0.238, 116.1) and 4.688(0.090, 97.33) respectively. This indicates the existence of over-dispersion because the values for the parameters ψ_{HMPV} and ψ_{RSV} are greater than zero which relaxes our adoption of the negative-binomial modelling, despite in the simulation data there was over-dispersion detected. Figures 5, 6, 7 and 8 show the posterior median and point-wise 95% credibility intervals for the climatic factors.
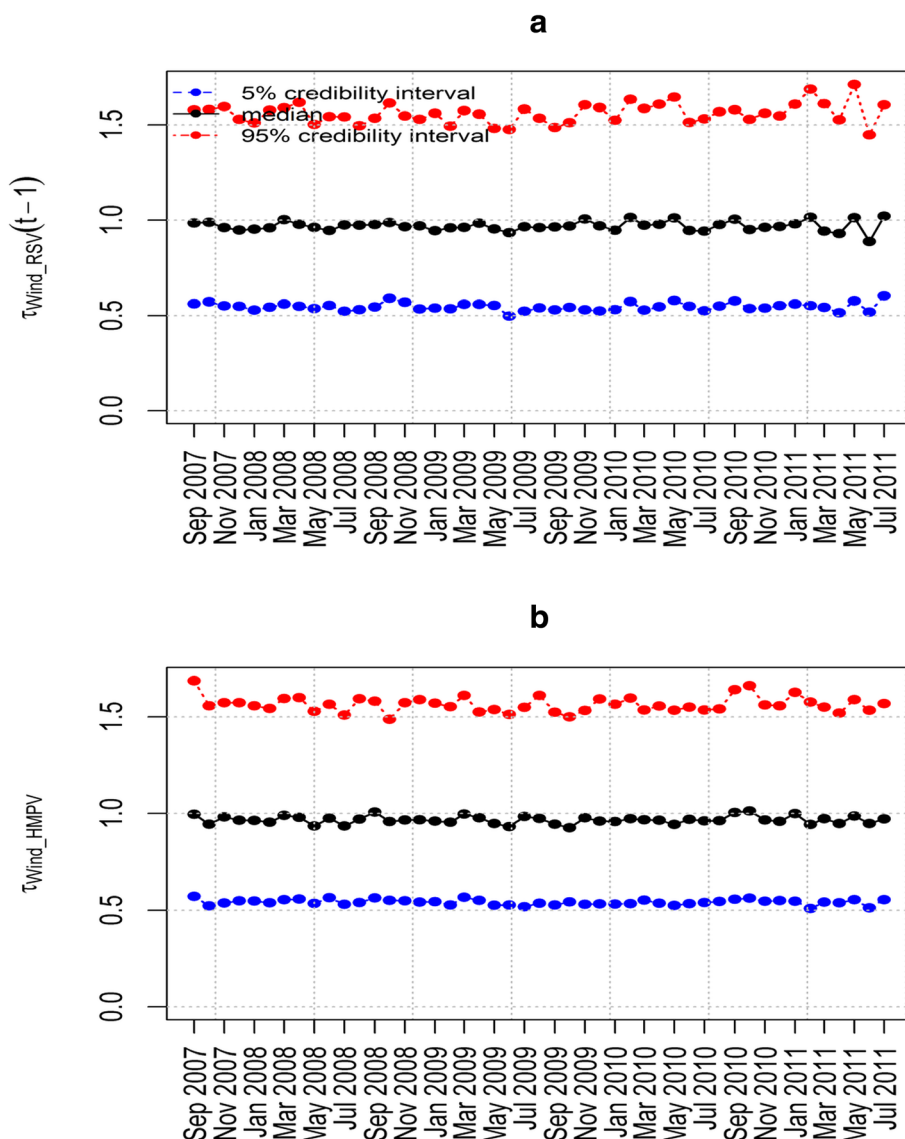
Nyoka *et al. BMC Public Health*   (2019) 19:807

Page 10 of 14

Fig. 6 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (**a**) τ_{Wind_RSV} and (**b**) τ_{Wind_HMPV} for model 6(iii)

Discussions

The RSV data shows bi-annual peaks of different severity during the rainy seasons in the Dadaab refugee camp (Kenya) [26, 27]. Wilkesmann et al. [28] showed that both HMPV and RSV cause similar symptoms and clinical severity with similar seasonality. A similar finding was reached by Kim et al. [29] who investigated the clinical and epidemiological assessment of HMPV and RSV in Seoul, Korea, 2003–2008. In their paper, Cuevas et al. [6] observed that HMPV incidence had increased with increases in RSV incidence. Another study in Yemen, children younger than 2 years identified co-infections of RSV and HMPV, and also showed that there were seasonal variations of RSV and HMPV with a peak of RSV

in December and January and a peak of HMPV in February and March [30].

From our previous work using the same dataset, we noted a similar conclusion that the use of climatic factors explained the seasonality of RSV [27]. This implies that having considered the different rate for each virus at every time point, the models with the best fit to data were those with climatic factors. In our study, we have shown that the incidence of RSV influenced that of HMPV from the best model fit. It is therefore crucial to establish good RSV surveillance systems in developing countries to help understand the dynamics of the disease. This will aid in knowing when to put up an intervention to control for RSV and HMPV outbreaks. Some
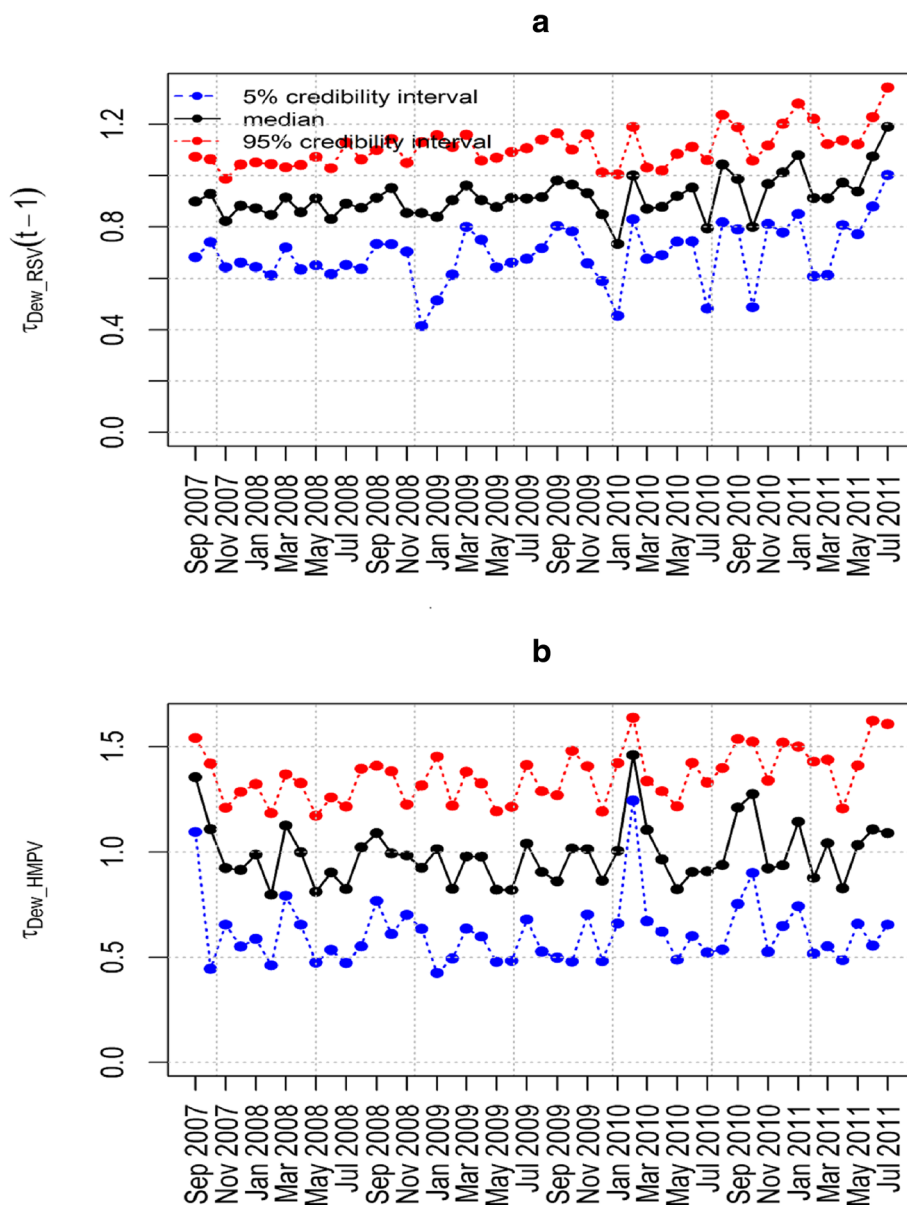
Fig. 7 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (**a**) τ_{Dew_RSV} and (**b**) τ_{Dew_HMPV} for model 6(iii)

of the interventions include washing hands with soap and avoiding overcrowding. A similar observation was made by Lazar et al. who noted that HMPV did not contribute to the severity of RSV [31]. This is corroborated in findings from a similar investigation of the influence of RSV on HMPV by Greensill et al. [10] in which 70% of children infected with RSV were co-infected with HMPV. Elsewhere, Cuevas et al. [6] observed that HMPV incidence increased with increasing number of RSV cases suggesting the presence of a strong association between the dynamics of the two epidemics.

Some of the limitations of this study were that the available time series data for the viruses was only for a four-year time-frame which is short for time series analysis and that the climatic factors were from the neighboring weather station which is about 100 km away from the Dadaab camp. Nevertheless, the weather measurements are a good representation of the actual weather around Dadaab. There was no establishment of whether patients were co-infected during virus testing. We used the DIC which is an approximation of a penalized loss function based on the deviance to evaluate the models. The application is valid only when the number of
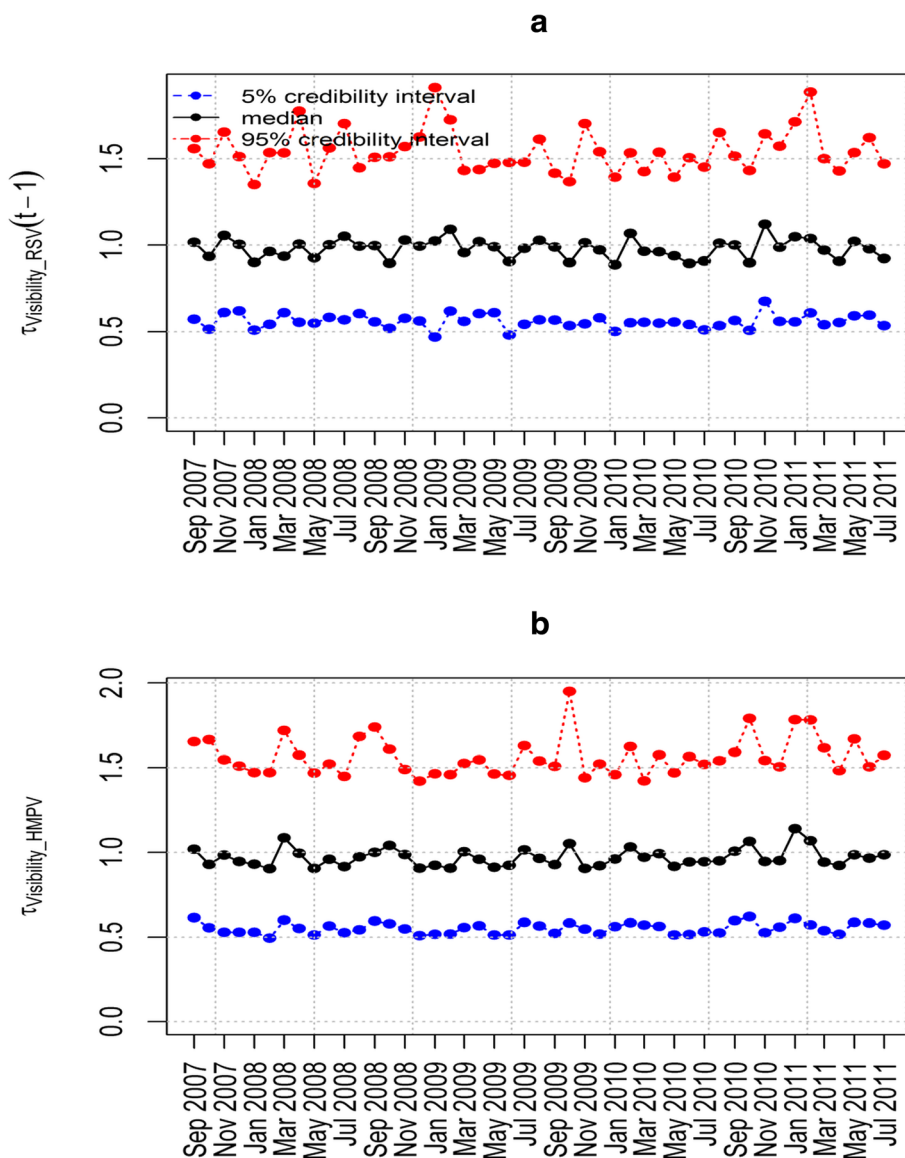
Fig. 8 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (**a**) $\tau_{Visibility\_RSV}$ and (**b**) $\tau_{Visibility\_HMPV}$ for model 6(iii)

parameters is much smaller than the number of independent observations [32]. The classical model selection was used that assumes that there is at least a best model for deducing inferences from the data. The criterion used to select the best model did not allow for the computation of weights of each fitted model to quantify for uncertainty, that is the model averaging techniques were not used [33].

Conclusion

We provided a comprehensive comparison of RSV and HMPV in a refugee camp setting by using a bivariate non-Gaussian model to jointly model the epidemics. By comparing various model structures, we identified a

model that satisfactorily fits the epidemic data, thereby explaining most of the observed variation therein. The models and estimated parameters also provided clues into the dynamics and stability of the two epidemics. Our results demonstrated the influence of RSV on HMPV while adjusting for climatic factors. The climatic factors played a significant role in explaining the influence of RSV incidence on HMPV incidence. These models are important to the public health implication since controlling the incidence of RSV would consequently reduce the incidence of HMPV.

Nyoka *et al. BMC Public Health*   (2019) 19:807

Page 13 of 14

criterion; HMPV: Human metapneumovirus; KEMRI: Kenya Medical Research Institute; LRTIs: Lower respiratory tract infections; RSV: Syncytial virus; RVIs: Respiratory virus infections

Authors' contributions

RN, TNOA and JO contributed to the conception and design of the study, acquisition, analysis and interpretation of data, and have been involved in drafting the manuscript and revising it critically for important intellectual content. RN JO SMM TNOA AG and HM contributed in study implementation and methodology. SMM TNOA AG and HM contributed to the study supervision and validation of the project including revising the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Availability of data and materials

The data files and supplementary materials used for this study can be found at https://figshare.com/s/e8a735c22f554d8372e3 DOI: https://doi.org/10.6084/m9.figshare.5340724 .

Ethics approval and consent to participate

Ethical approval for the surveillance activities was obtained from the Kenya Medical Research Institute (KEMRI) Ethical Review Committee (SSC Protocol Number 1161). Institutional review was waived by US Centers for Disease Control and Prevention (CDC) because the study was considered to be a non-research public health activity. Informed written consent was obtained from all participants and from the guardians of minors.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, South Africa. [2]Molecular Genetics Department, University of Groningen, 9747 AG Groningen, The Netherlands. [3]Statistics Department, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000-00200, Nairobi, Kenya. [4]Jhpiego - an affiliate of John Hopkins University, P.O. Box 66119, Westlands, Nairobi 00800, Kenya. [5]Nairobi, Kenya.

References

1. Simoes E A. F, Cherian T, Chow J, Shahid-Salles S A., Laxminarayan R, John TJ. Acute Respiratory Infections in Children. Dis Control Priorities Dev Ctries [Internet]. 2006;1–24. Available from: http://www.ncbi.nlm.nih.gov/books/NBK11786/. Accessed 3 June 2019.
2. Checchi F, Gayer M, Grais RF, Mills EJ. Public Health in Crisis-Affected Populations: A Practical Guide for Decision-Makers [Internet]. Humanitarian Practice Network. 2007. Available from: https://fieldresearch.msf.org/handle/10144/52753. Accessed 3 June 2019.
3. Bellos A, Mulholland K, O'Brien KL, Qazi SA, Gayer M, Checchi F. The burden of acute respiratory infections in crisis-affected populations: a systematic review. Confl Health. 2010;4(1):3 Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-77954739853&partnerID=tZOtx3y1.
4. Tregoning JS, Schwarze J. Respiratory viral infections in infants: causes, clinical symptoms, virology, and immunology. Clin Microbiol Rev. 2010;23(1):74–98.
5. Pastula ST, Hackett J, Coalson J, Jiang X, Villafana T, Ambrose C, et al. Hospitalizations for respiratory syncytial virus (RSV) among adults in the United States, 1997 - 2012. Open Forum Infect Dis. 2017;48105:ofw270 Available from: https://academic.oup.com/ofid/article-lookup/doi/10.1093/ofid/ofw270.
6. Cuevas LE, Ben NAM, Dove W, Gurgel RQ, Greensill J, Hart CA. Human metapneumovirus and respiratory syncytial virus, Brazil. Emerg Infect Dis. 2009;9(12):1626–8.
7. Peiris JSM, Tang WH, Chan KH, Khong PL, Guan Y, Lau YL, et al. Children with respiratory disease associated with metapneumovirus in Hong Kong. Emerg Infect Dis. 2003;9(6):628–33.
8. Guerrero-Plata A, Casola A, Suarez G, Yu X, Spetch L, Peeples ME, et al. Differential response of dendritic cells to human metapneumovirus and respiratory syncytial virus. Am J Respir Cell Mol Biol. 2006;34(3):320–9.
9. Le Nouen C, Munir S, Losq S, Winter CC, McCarty T, Stephany DA, et al. Infection and maturation of monocyte-derived human dendritic cells by human respiratory syncytial virus, human metapneumovirus, and human parainfluenza virus type 3. Virology. 2009;385(1):169–82.
10. Greensill J, McNamara PS, Dove W, Flanagan B, Smyth RL, Hart CA. Human metapneumovirus in severe respiratory syncytial virus bronchiolitis. Emerg Infect Dis. 2003;9(3):372–5.
11. Konig B, Konig W, Arnold R, Werchau H, Ihorst G, Forster J. Prospective study of human Metapneumovirus infection in children less than 3 years of age. Society. 2004;42(10):4632–5.
12. Tang RS, Schickli JH, Macphail M, Fernandes F, Bicha L, Spaete J, et al. Effects of Human Metapneumovirus and Respiratory Syncytial Virus Antigen Insertion in Two 3J Proximal Genome Positions of Bovine / Human Parainfluenza Virus Type 3 on Virus Replication and Immunogenicity. J Virol. 2003;77(20):10819–28.
13. Akhras N, Weinberg JB, Newton D. Human metapneumovirus and respiratory syncytial virus: subtle differences but comparable severity. Infect Dis Rep. 2010;2(2):35–9.
14. Moe N, Krokstad S, Stenseng IH, Christensen A, Risnes KR, Arne S, et al. Comparing human Metapneumovirus and respiratory syncytial virus : viral co- detections , genotypes and risk factors for severe disease. PLoS One [Internet. 2017;12(1):1–19. https://doi.org/10.1371/journal.pone.0170200.
15. You H-L, Chang S-J, Yu H-R, Li C-C, Chen C-H, Liao W-T. Simultaneous detection of respiratory syncytial virus and human metapneumovirus by one-step multiplex real-time RT-PCR in patients with respiratory symptoms. BMC Pediatr [internet]. BMC Pediatr. 2017;17(1):89 Available from: http://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-017-0843-7.
16. Ditt V, Lüsebrink J, Tillmann RL, Schildgen V, Schildgen O. Respiratory infections by HMPV and RSV are clinically indistinguishable but induce different host response in aged individuals. PLoS One. 2011;6(1):1–9.
17. Held L, Höhle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. Stat Modelling. 2005;5(3):187–99.
18. O'Hara RB, Kotze DJ. Do not log-transform count data. Methods Ecol Evol. 2010;1(2):118–22. https://doi.org/10.1111/j.2041-210X.2010.00021.x.
19. Nishiura H. Early Detection of Nosocomial Outbreaks Caused by Rare Pathogens: A Case Study Employing Score Prediction Interval. Osong Public Heal Res Perspect. 2012;3(3):121–7. https://doi.org/10.1016/j.phrp.2012.07.010.
20. Buckeridge DL, Okhmatovskaia A, Tu S, O'Connor M, Nyulas C, Musen MA. Predicting outbreak detection in public health surveillance: quantitative analysis to enable evidence-based method selection. AMIA Annu Symp Proc. 2008;2008:76–80.
21. Ver Hoef JM, Boveng PL. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology. 2007;88(11):2766–72.
22. Paul M, Held L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Stat Med. 2011;30(10):1118–36.
23. Paul M, Meyer S. The function hhh4 in the R -package surveillance; 2014. p. 1–15.
24. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity anf fit. J R Stat Soc Ser B (Statistical Methodol). 2002;64(4):583–639.
25. Adrion C, Mansmann U. Bayesian model selection techniques as decision support for shaping a statistical analysis plan of a clinical trial: an example from a vertigo phase III study with longitudinal count data as primary endpoint. BMC Med Res Methodol. 2012;12:137 Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3554595&tool=pmcentrez&rendertype=abstract.

Nyoka *et al. BMC Public Health*   (2019) 19:807

Page 14 of 14

26. Agoti CN, Mayieka LM, Otieno JR, Ahmed JA, Fields BS, Waiboci LW, et al. Examining strain diversity and phylogeography in relation to an unusual epidemic pattern of respiratory syncytial virus (RSV) in a long-term refugee camp in Kenya. BMC Infect Dis. 2014;14(1):178 Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4021307&tool=pmcentrez&rendertype=abstract.

27. Nyoka R, Omony J, Mwalili SM, Achia TNO, Gichangi A, Mwambi H. Effect of climate on incidence of respiratory syncytial virus infections in a refugee camp in Kenya: a non-Gaussian time-series analysis. PLoS One. 2017;12(6):1–14.

28. Wilkesmann A, Schildgen O, Eis-Hübinger AM, Geikowski T, Glatzel T, Lentze MJ, et al. Human metapneumovirus infections cause similar symptoms and clinical severity as respiratory syncytial virus infections. Eur J Pediatr. 2006; 165(7):467–75.

29. Kim CK, Choi J, Callaway Z, Bin KH, Chung JY, Koh YY, et al. Clinical and epidemiological comparison of human metapneumovirus and respiratory syncytial virus in Seoul, Korea, 2003-2008. J Korean Med Sci. 2010;25(3):342–7.

30. Al-Sonboli N, Hart CA, Al-Aeryani A, Banajeh SM, Al-Aghbari N, Dove W, et al. Respiratory syncytial virus and human metapneumovirus in children with acute respiratory infections in Yemen. Pediatr Infect Dis J. 2005;24(8):734–6.

31. Lazar I, Weibel C, Dziura J, Ferguson D, Landry ML, Kahn JS. Metapneumovirus and severity of respiratory disease. Emerg Infect Dis. 2004;10(7):7–9 Available from: http://www.ncbi.nlm.nih.gov/pubmed/16607540.

32. Plummer M. Penalized loss functions for Bayesian model comparison. Biostatistics. 2008;9(3):523–39.

33. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research. 2004;33(2):261–304.

Publisher's Note