University of Groningen

Effect of climate on incidence of respiratory syncytial virus infections in a refugee camp in Kenya

Nyoka, Raymond; Omony, Jimmy; Mwalili, Samuel M.; Achia, Thomas N. O.; Gichangi, Anthony; Mwambi, Henry

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Effect of climate on incidence of respiratory syncytial virus infections in a refugee camp in Kenya: A non-Gaussian time-series analysis

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

1 School of Mathematics, Statistics and Computer Science, University of KwaZulu- Natal, Scottsville, South Africa, **2** Molecular Genetics Department, University of Groningen, Groningen, Netherlands, **3** Statistics Department, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, **4** Jhpiego - an affiliate of John Hopkins University, Westlands, Nairobi, Kenya

* nyomus@gmail.com

Abstract

Respiratory syncytial virus (RSV) is one of the major causes of acute lower respiratory tract infections (ALRTI) in children. Children younger than 1 year are the most susceptible to RSV infection. RSV infections occur seasonally in temperate climate regions. Based on RSV surveillance and climatic data, we developed statistical models that were assessed and compared to predict the relationship between weather and RSV incidence among refugee children younger than 5 years in Dadaab refugee camp in Kenya. Most time-series analyses rely on the assumption of Gaussian-distributed data. However, surveillance data often do not have a Gaussian distribution. We used a generalized linear model (GLM) with a sinusoidal component over time to account for seasonal variation and extended it to a generalized additive model (GAM) with smoothing cubic splines. Climatic factors were included as covariates in the models before and after timescale decompositions, and the results were compared. Models with decomposed covariates fit RSV incidence data better than those without. The Poisson GAM with decomposed covariates of climatic factors fit the data well and had a higher explanatory and predictive power than GLM. The best model predicted the relationship between atmospheric conditions and RSV infection incidence among children younger than 5 years. This knowledge helps public health officials to prepare for, and respond more effectively to increasing RSV incidence in low-resource regions or communities.

Introduction

Respiratory syncytial virus (RSV) is one of the major causes of acute lower respiratory tract infections (ALTRI) in infants and young children [1][2]. RSV infections occur seasonally in temperate climate regions [3]. RSV adversely impacts the health of adults and immunocompromised patients, and is associated with significant mortality and morbidity, particularly in young children and vulnerable infants [4]. Children younger than 1 year are most susceptible

to RSV infection; often 60–70% of children in this age group have been infected at least once, and re-infection can occur throughout their lifetime [4][5][6].

RSV is shed in saliva and nasopharyngeal secretions [7]. Infected hosts shed higher quantities of viral particles upon exposure to higher-ambient temperatures [8]. Low humidity during winter enhances RSV viability, and enables its survival for up to 12 hours on nonporous surfaces [9]. In dry air conditions, large droplets evaporate and remain air-borne for longer periods of time. Some studies have shown that airborne transmission appears to be sensitive to ambient humidity and temperature in temperate regions [8][10]. RSV outbreaks show some seasonality that suggests a connection with atmospheric and environmental conditions [11] [12]. Most RSV infections in temperate locations occur between November and April [13]. RSV infection has been associated with winter in these regions because people spend more time indoors, potentially in crowded conditions [14]. Such climatic regions are different from those of Kenya, which is located on the equator and experiences bimodal seasonal rainfall due to the interaction of the Northern and Southern Hemisphere monsoon systems [15]. However, variations in climatic factors, such as humidity, temperature, wind speed, rainfall etc., can have a significant impact on disease dynamics. Therefore, it is essential that the RSV incidence be evaluated for equatorial climatic regions to aid accurate predictions of RSV outbreaks. [16] [17].

The wide range of statistical methods used to explore the link between RSV outbreaks and climate makes it difficult to elucidate a definitive relationship. Pearson correlation analysis was previously used to explain the associations of RSV-positive cases with meteorological variables [11]. The univariate analysis of variance (ANOVA), multiple regression analysis, and Spearman's rank correlation were used to assess the association between RSV incidence and meteorological parameters [18]. A better understanding of the relationship between climate and RSV helps in making reliable predictions of its incidence.

Worldwide, as of 2005, 99% of deaths from RSV were reported by the World Health Organization (WHO) to occur in developing countries [19]. It is, therefore, crucial to establish good RSV surveillance systems in developing countries to help understand the dynamics of the disease. In 2006, the U.S. Centers for Disease Control and Prevention (CDC) and the Kenya Medical Research Institute (KEMRI) established a respiratory illness surveillance system to detect disease outbreaks in Kenyan refugee camps [20]. We used RSV incidence data from this system to explore the best model that predicts the relationship between RSV incidence and climatic factors along spatio-temporal scales to determine whether a seasonal pattern of RSV infection exists. A generalized linear model (GLM) with a sinusoidal component over time was used to account for seasonal variation and compared with a generalized additive model (GAM) with smoothing cubic splines. Climatic factors were included as covariates in the models before and after timescale decompositions.

Methods

Data

Surveillance for viral respiratory illnesses, including adenovirus, human metapneumovirus, influenza virus, parainfluenza viruses 1, 2, and 3, and RSV was implemented in Dadaab refugee camp in north eastern Kenya in 2007. Paediatric and adult patients who presented at a camp medical unit, and met the case definition for influenza-like illness (ILI) or severe acute respiratory infection (SARI), were enrolled into the laboratory-enhanced respiratory surveillance system and tested for all of the above diseases after an informed consent form was completed by adults, older minors, and guardians of all minors <15 years [20]. The number of laboratory-confirmed cases was recorded on a daily basis from September 2007 to August 2011. The

monthly counts of all RSV cases among children younger than 5 years were included in the present analysis; the main outcome of interest being monthly RSV incidence rate in this age group. RSV incidence rate per 1,000 children younger than 5 years was calculated by dividing monthly RSV counts by the monthly population of children younger than age 5 years in the camp. Local weather and climatic data, including: the mean temperature and mean dew point for the day (both in ˚F); mean sea level pressure for the day in millibars; mean visibility for the day in miles; mean wind speed for the day in knots; minimum and maximum temperature (˚F) reported during the day; and the total precipitation (in inches) reported during the day 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 consisted of measurements recorded at successive, equally spaced time points (covariates used in the present study are provided in the supplementary materials, S3 Table). Data and R codes used in the analysis are available at https://figshare.com/s/feb61d236cad0abcf5b6 DOI 10.6084/m9.figshare.5010767.

Statistical modeling

A Poisson distribution model was used in this analysis, as the outcome of interest (incident RSV cases) was non-Gaussian count data. Some authors have used Gaussian vector autoregressive models on multivariate counts that are serially correlated. Brandt and others used vector autoregressive methods that were based on Gaussian error process [21]. However, such an assumption is not applicable to event count data because it produces biased estimates [22]. So, as many of those methods apply for count series that approximate normality, they may not hold to dynamic events like the ones applied here. In the first model, seasonal effects on RSV incidence were analysed by using a generalized linear model (GLM) with a sinusoidal component to account for seasonal variation. The second model extended the GLM model to a generalized additive model (GAM) by applying smoothing cubic splines. The GAM is an extension of the GLM and is adaptable to non-normally distributed variables [23]. GLM uses linear predictors specified as the expected value of a response variable (Y_j), which is expressed as $\eta = \Sigma_j \beta_j(X_j)$. Here, β_j is a coefficient parameter and X_j represents the j-th explanatory variable. The GAMs extend these by replacing them with $\eta = \Sigma_j f_j(X_j)$, where $f_j(X_j)$ are unspecified nonparametric functions estimated by including smoothing splines [24]. GAMs allow for adjustments of the nonparametric, nonlinear, confounding effects of seasonality, trends, and weather variables, which have been previously used in modeling time-series data [25]. In the present analysis, climatic time-series covariates were included in the GLM and GAM models and implemented in R language v3.1.0 [26]. Both models were optimized for predictive accuracy and precision.

Data were decomposed into three components, namely: trend, seasonal, and random components, in order to independently evaluate the existence and strength of associations between RSV incidence and covariates on each time scale. Data decomposition was accomplished using Loess smoothing, a regression method that assigns a weighted polynomial to each component [25]. We introduced a GLM for time-series data, with a sinusoidal component over time to account for seasonal variations. The GLM was extended to include a smoothing function using the GAM approach to the Poisson distribution [27] In each model, a data-driven smoothing function of time was fitted, and compared with those fitted, using sine and cosine functions in the Fourier basis.

The observed number of RSV counts, Y_t at a given month $t = 1, \cdots, n$ from the population at risk is assumed to follow a Poisson random variable: $Y_t \sim \text{Poisson}(\mu_t)$. We let n_t be the population of children age 5 years and younger at risk at time t in the camp. The expected value of

Y_t is $E(Y_t) = \mu_t = n_t\Theta_t$ where the dependence of covariates on Θ_t is modeled by $\Theta_t = e^{x_t^T\beta}$. Therefore, a Poisson GLM of the form $E(Y_t) = \mu_t = n_t e^{x_t^T\beta}$ is used. More explicitly, to model the incidence, we use:

$$\log \mu_t = \beta_0 + \alpha y_{t-1}$$
$$+ \sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{s=1}^{r}\sum_{l=0}^{q}\beta_{ksl}x_{(t-l)ks} + \eta_1 cos\left[\frac{2\pi t}{T}\right] + \eta_2 sin\left[\frac{2\pi t}{T}\right] + \log n_t \quad (1)$$

Where μ_t is the infection rate for the month, t. β_0 is the intercept, α is the coefficient of the lagged RSV counts by one month, which is represented by y_{t-1}, $x_{(t-l)ks}$, is the decomposed measured covariate, β_{ksl} their corresponding coefficients with $k = 1, \cdots, m$ covariates and $s = 1, \cdots, r$ corresponding to r-th decomposition of the k-th covariate, $l = 0, 1, \cdots, q$ distributed lags where q is the maximum lag and $t = 1, \cdots, n$ are the time points. The terms η_1 and η_2 are coefficients of the sine and cosine function, respectively. Here, T is the number of time periods described by one cosine function over the interval $[0, 2\pi]$.

Using a cosine function, we specified two periods: one that defines the measure of RSV infection (month) and the other that is described by one cosine cycle. After fitting all covariates in the GLM model, the most parsimonious model was identified. The maximum lag for each covariate was obtained by comparing different lagged models using Akaike information criterion (AIC). The maximum lag for each covariate was used to run "crossbasis" in the "dlnm" package for time-series models [28][29]. The same covariates were used to fit the GAM model.

The corresponding GAM for the Poisson model is:

$$\log \mu_t = \beta_0 + \alpha y_{t-1}$$
$$+ \sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{s=1}^{r}\sum_{l=0}^{q}\beta_{ksl}\Psi_k(x_{(t-l)ks}, \lambda_{tks}) + \Psi_{k+1}(t, \lambda_{t(k+1)}) + \log n_t \quad (2)$$

Where λ_{tks} is the smoothing parameter or the degrees of freedom for covariates, λ_{k+1} is a smoothing parameter for time and Ψ. is the smoothing function. Larger values of λ. are indicative of a less-smooth function.

For Models (1) and (2), the additive time-scale decomposition of the k-th covariate into the seasonal (S), trend (T), and random (R) components is:

$$\beta_{ksl}x_{(t-l)ks} = \beta_{kSl}x_{(t-l)kS} + \beta_{kTl}x_{(t-l)kT} + \xi_{(t-l)kR} \quad (3)$$

for every k in $\{1, \cdots, m\}$. In the above case, s takes on three levels S, T, and R. This decomposition helps in assessing for the significance of the seasonal and trend components of the covariates in explaining the RSV incidence. The combination of the seasonal and trend components makes up the patterns in the covariates.

The trend cycles represent long-term changes in the levels or values of the covariate, while the periodic changes are the fluctuations of constant length. The GLM (1) has the Logit link function. The residual deviance for these models takes on the form $D = -2\log(L_{test}/L_{sat})$, where L_{test} and L_{sat} are the maximized likelihoods under the test and saturated models, respectively. The model selection and fitting was done using the "glmulti" package [30] and "gam" [31] in "mgcv" package [24] in the R language v3.1.0.

Ethical considerations

Ethical approval for the surveillance activities was obtained from the KEMRI Ethical Review Committee (SSC Protocol Number 1161). Institutional review was waived by 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.
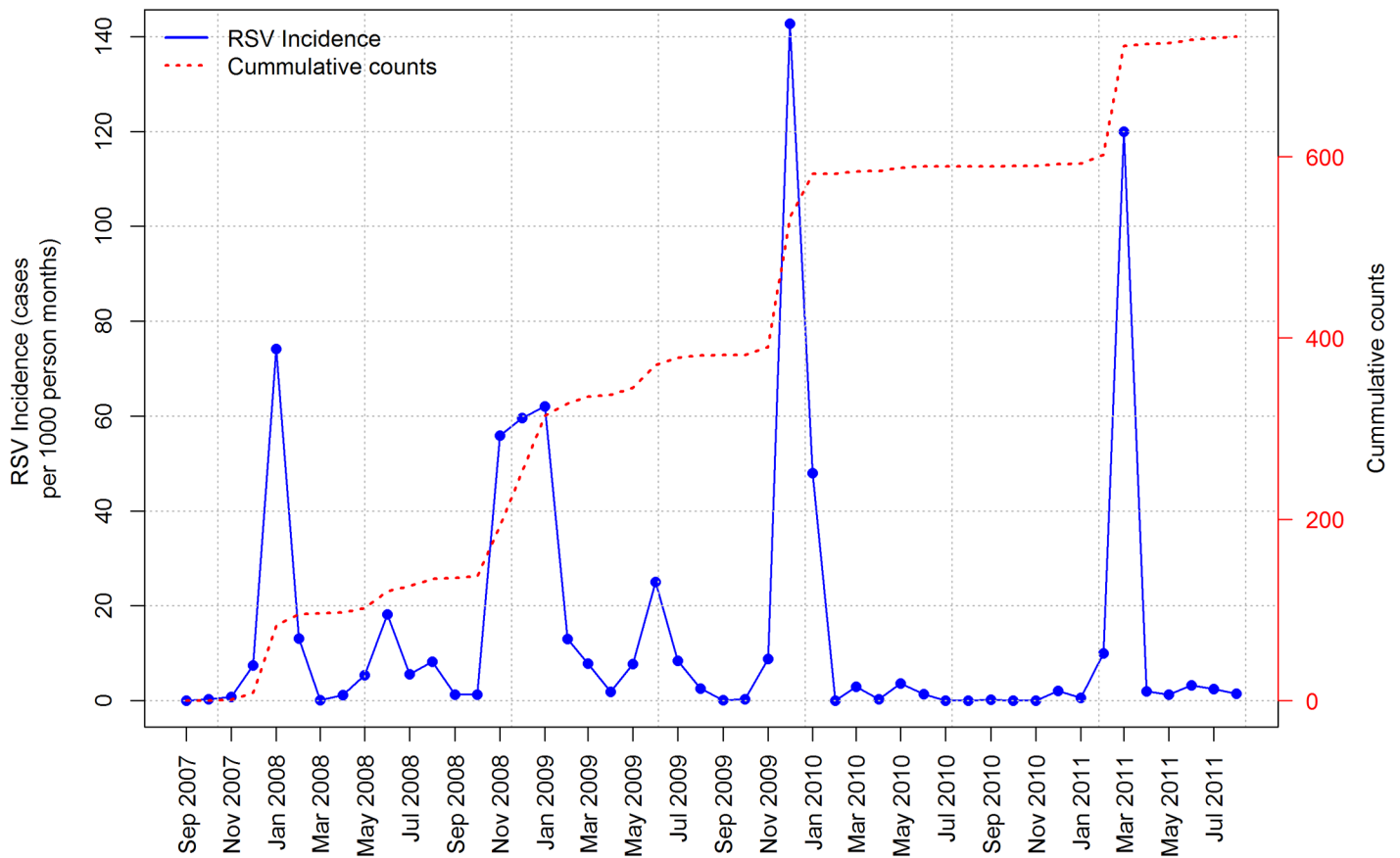
Fig 1. Plot of RSV incidence in Dadaab. Fluctuations in the data are roughly constant over time, indicating that the RSV time series could likely be described using an additive model.

Results and discussions

Data exploration

A peak in RSV incidence occurred every 11–12 months, particularly from October to January (Fig 1). Other than these peaks, there was relatively low RSV incidence (≤ 20 cases per 1000 person months).

The decomposed data, seasonal pattern, trend line, and random component of the RSV, wind, rainfall, and temperature time series are shown in S1–S4 Figs. The seasonal pattern of RSV incidence regularly repeated itself, with two distinct peaks annually (S1 Fig). The data show that overall, lower wind speeds and higher temperatures were associated with higher RSV incidence. The magnitude of the seasonal components of the decomposed covariates did not vary annually (S1–S6 Figs). This justifies the use of additive, rather than multiplicative decomposition. There was a positive correlation between temperature and RSV incidence (Fig 2B). There was a significant moderate correlation between RSV incidence and wind speed ($\rho = -1.603$, $p = 0.003$)(Fig 2A); an insignificant weak correlation between RSV incidence and temperature($\rho = 0.809$, $p = 0.289$) (Fig 2B); an insignificant weak correlation between RSV incidence and dew point ($\rho = -0.763$, $p = 0.201$) (Fig 2A); and for temperature and wind speed (Fig 2D);, the parabolic curve was fitted using:$x_3 = Ⴤ_0 + Ⴤ_1(x_1 - Ⴤ_2)^2$ where $Ⴤ_{0,1,2}$ are constants, and the regression fit was significant ($p < 0.001$). Here, x_1 and x_3 represent wind speed and temperature, respectively.
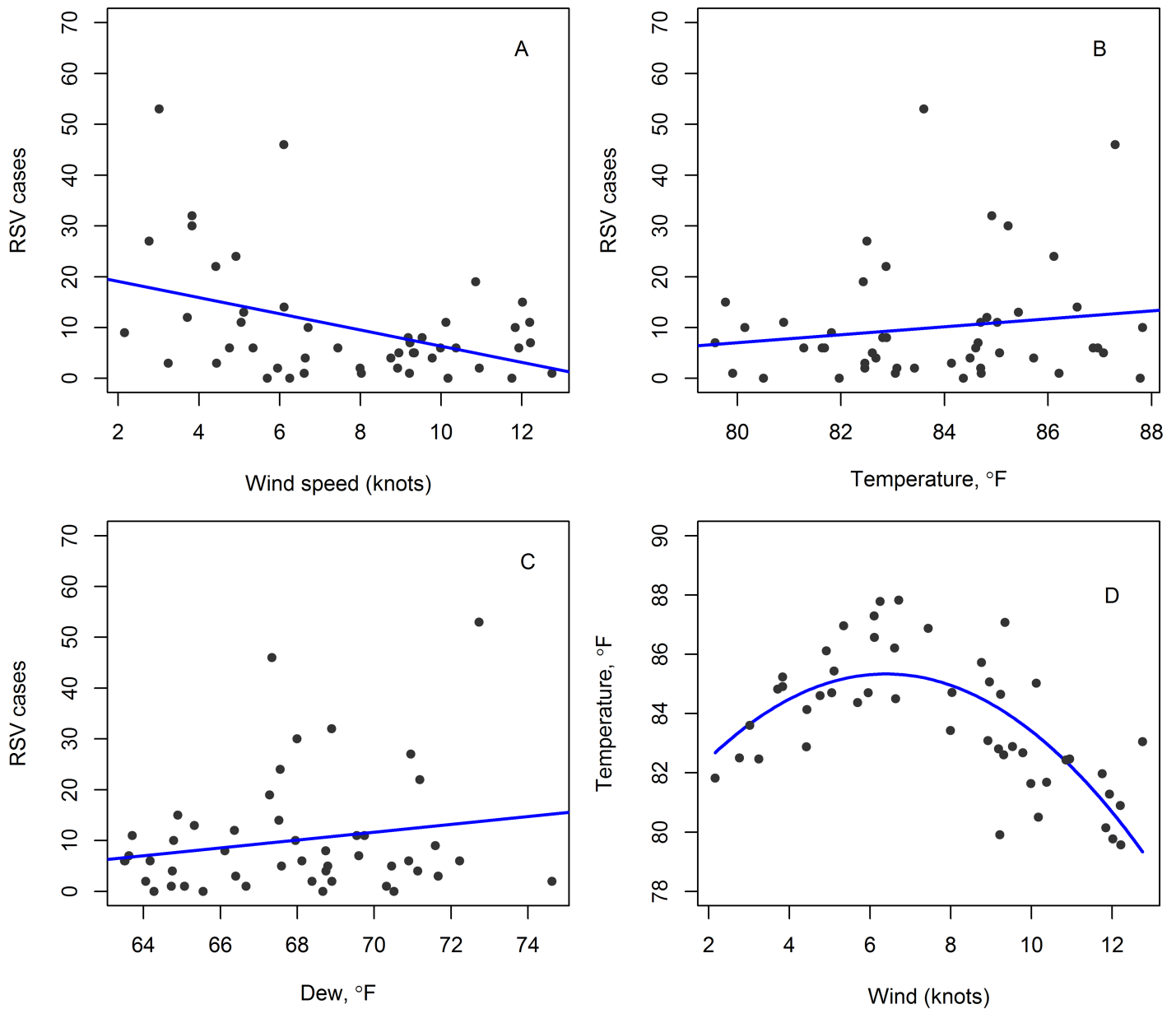
Fig 2. Correlation-regression analysis. A: Correlation between RSV incidence and wind speed; B: Correlation between RSV incidence and temperature; C: Correlation between RSV incidence and dew point; and D: Correlation between temperature and wind speed. In these plots, the regression lines of best fit are indicated by bold blue lines.

Model assessment and comparison

The trend component of the wind decomposition model decreased slightly immediately after 2008, then increased steadily to a peak in early 2009, followed by a decrease to a minimum value in late 2010 (S2 Fig). These finding indicate that the wind variable has a seasonal impact on RSV incidence. A similar seasonal effect was observed in the RSV and temperature decomposition (S1–S4 Figs). To determine the best predictive model, we compared the performance of the four models described in the methods section. The best GLMs and GAMs from the Poisson were compared using the AIC and residual deviances (S1 Table). In the models with
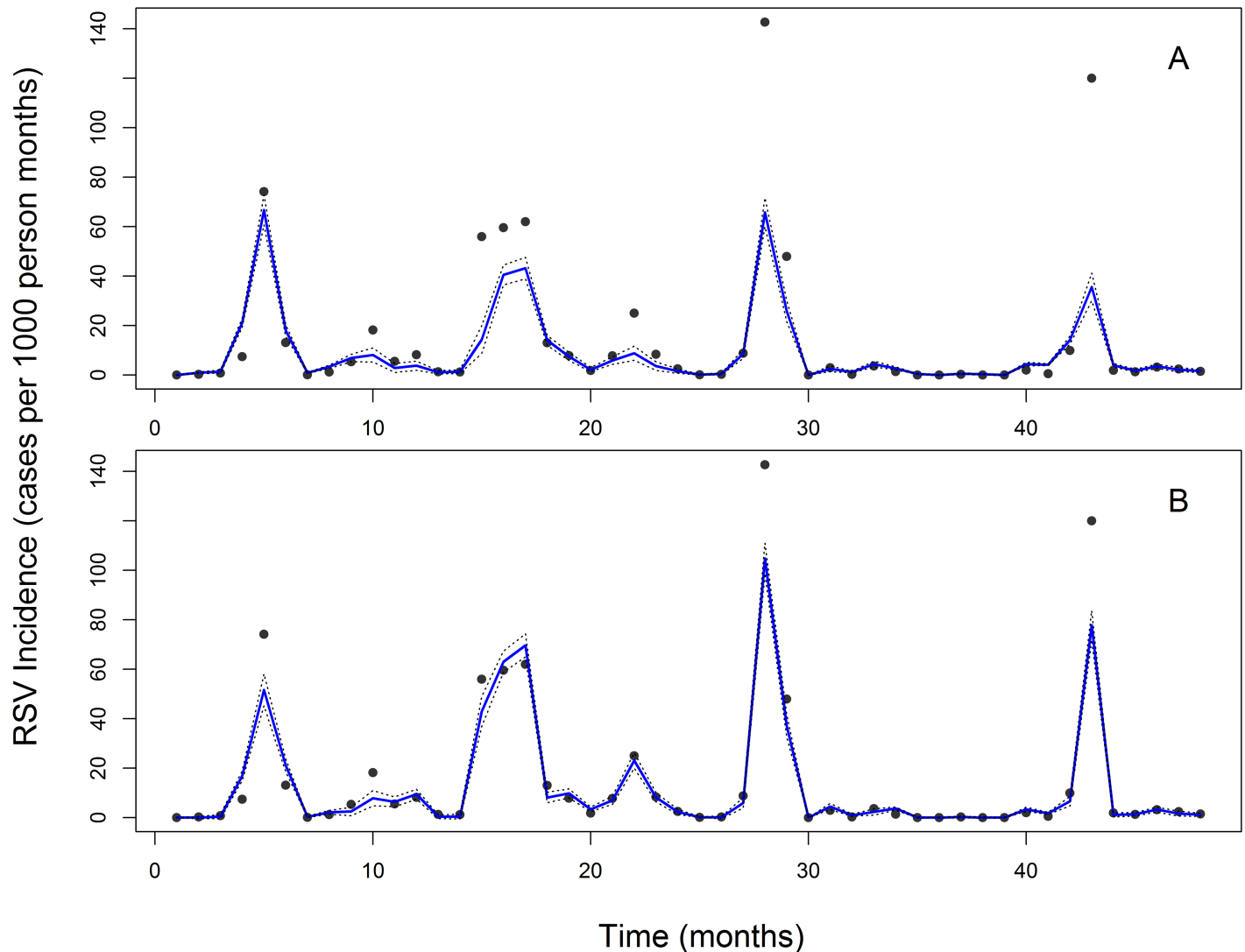
Fig 3. Best model fit to the RSV incidence data (bold lines) with decomposed covariates. A: Poisson, GLM. B: Poisson, GAM. The standard error bars to the model fit are indicated by the dotted lines (95% confidence bounds). The base year in all these plots was September 2009.

decomposed covariates for both GLM and GAM, the current cases of RSV did not depend on the previous observations. The AIC was used to judge the best model from the set of models that had a good fit. The best models all had covariates with $p < 0.05$. This was the case for models with and without decomposed covariates. Of all the models that were evaluated, the Poisson GAM with decomposed covariates had the best fit to the data (AIC = 317.17 and a Deviance explained = 65.3%, S1 Table). Fig 3 shows the best model fit to the RSV incidence data with decomposed covariates comparing the Poisson GLM and the Poisson GAM, where the Poisson GAM fits the data well. The best model in its reduced form is the Poisson GAM[a] (S4 Table). S2 Table contains the corresponding ANOVA results for the Poisson GAM[a]. From this table, the wind with both the trend and seasonal effects (seasonal effect of rainfall, trend mean dew point, and the trend effect of visibility) significantly explained RSV incidence. We note that time in months did not significantly explain RSV incidence, further demonstrating the importance of using climactic factors to explain the seasonality of RSV.

The direction of effects demonstrated nonlinear relationships with RSV incidence, except in the case of seasonal wind speed, which had a linear relationship (Fig 4). High wind speed within the same month had a significant negative effect on the RSV incidence. The trend component of the wind speed in the 2 months preceding incident RSV cases had a nonlinear relationship with RSV incidence. As the wind speed increased, incidence fluctuated from low to high, returning to low incidence when the speeds were highest. An increase in the seasonal component of rainfall in the four months preceding RSV cases was associated with an increase in RSV incidence. When rainfall was at its lowest, RSV incidence increased then returned to baseline when rainfall reached its maximum. The trend effect of the mean dew point 1 month preceding incident cases was associated with an increase in RSV incidence until dew point reached its maximum. The increase in visibility trend component 2 months preceding incident RSV cases demonstrated a constant effect on RSV incidence, which peaked when the visibility was 19.5 miles and troughed when the visibility was at its highest.

Implication of results and comparison to related studies

Our data showed seasonal variations for RSV incidence (S1 Fig). The Poisson GAM with decomposed covariates out-performed the GLM variant, thereby relaxing its linearity. Generally, the role of climatic factors in determining disease dynamics is rather complex to decipher [32]. In the literature, there is strong evidence that the relationship between climatic factors and RSV incidence varies widely between geographical regions [18]. Previous studies have shown that climatic factors might be associated with RSV, although it remains unclear what these factors are or exactly how they impact RSV incidence. We performed a correlation analysis for each covariate with RSV by fitting regression lines to test the level of significance between the climatic variables (Fig 2A–2C). A recent study by Agoti et al [33] on RSV strains using the same RSV surveillance data showed that there were six epidemic peaks within the 3 year study period: two peaks each year; the first and the last peaks were composed of group B strains and the other four peaks were composed of group A strains. Agoti's study, in conjunction with our findings, show that onset of RSV infections in Kenya can be reliably predicted. Our findings, in comparison with other studies, also suggest that the relationship between RSV incidence and climatic factors varies widely; for instance, from 2004 to 2012 in tropical and sub-tropical zones such as Hong Kong, China, Singapore, Kuala Lumpur, Malaysia, Medellin and Colombia outbreaks occurred primarily during the hot and rainy seasons [14].

The ability to predict increases in RSV incidence, based on prevailing meteorological conditions, could potentially inform the application of public health interventions and provisions of healthcare in Kenya, and perhaps, in other regions with a similar climate and equatorial location. Currently, there is no RSV vaccine available; however, in developed countries, infants at risk of severe outcomes can be administered monthly doses of the anti-RSV antibody, palivizumab, during outbreaks of RSV [3][8]. Because predicting the incidence of RSV could optimize the cost-effectiveness of immunoprophylaxis; our model might be useful to apply in a cost-benefit analysis of this approach in Kenya. In most temperate climate regions, RSV occurs as an annual epidemic. For instance, Noyola and Mandeville found that temperature was the predominant atmospheric condition explaining the annual spread and variability of RSV incidence in San Luis Potosi, Mexico [34]. Using correlation and regression analysis, Noyola and colleagues observed that the weekly number of RSV incidence between October 2002 and May 2006 was correlated to ambient temperature, barometric pressure, relative humidity, vapor tension, dew point, precipitation, and hours. Our findings corroborate what they observed for the same climatic factors. The modeling has aided identification of factors influencing RSV incidence and provided indicators for devising measures to prevent the spread of the disease.
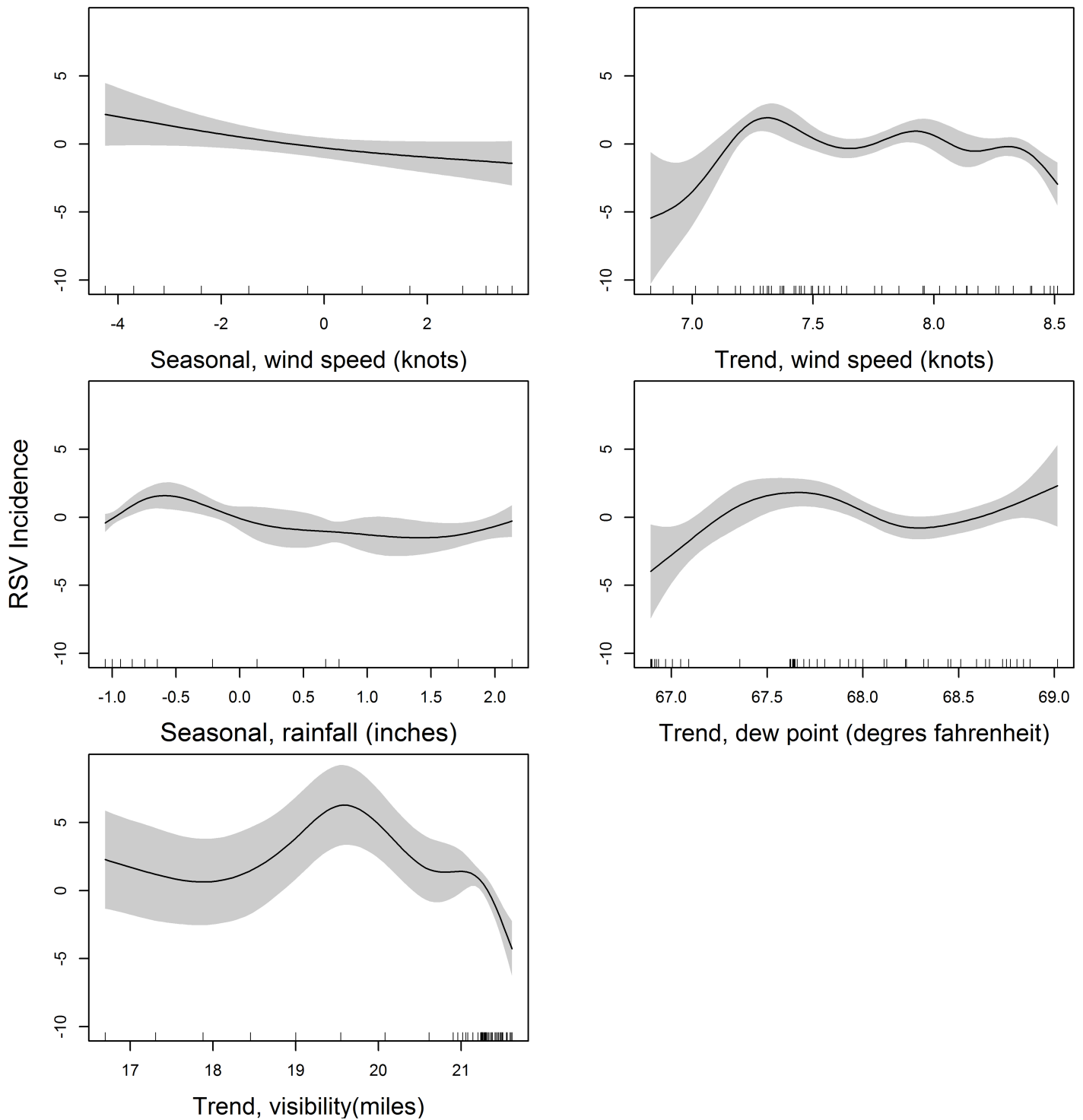
Fig 4. Best model fit (Poisson, GAM) to the RSV incidence data with the signifficant decomposed covariates. Seasonal, wind speed; Trend, wind speed; Seasonal, rainfall; Trend, rainfall; and Trend, visibility. The standard error bars to the model fit are indicated by the gray shade (95% confidence bounds). RSV incidence units as cases per 1,000 person months.

Our analysis showed that other climatic factors affecting RSV seasonality can improve the performance of a predictive model. Khor et al [18] demonstrated that, in Malaysia, ambient temperature was inversely associated with RSV activity, even though the highest number of cases may not always coincide with the lowest temperature. A negative correlation between the mean minimum temperatures and RSV incidence was recently reported in Italy [11]. RSV transmission that occurs during cold weather is facilitated by its stability in secretions, since inhalation of cold air slows down the mucociliary escalator. This reduces phagocytic activity of leukocytes, increasing the host's vulnerability to infection. There is evidence of RSV epidemics occurring in tropical areas with high temperatures during rainy seasons, a phenomenon that our data are exhibiting [35][36]. However, the exact mechanisms of how climatic factors affect RSV incidence requires further investigations, especially across geographically diverse regions. The relationship between the dynamics in wind speed and direction, and how these dynamics influence the climate of geographical regions like Dadaab, remains unclear. Understanding such complex relationships between the co-factors explaining the spread of RSV is essential to predict its incidence.

A foreseeable limitation of our models is that with log- or logit-links; the mean value zero corresponds to an infinite range on a linear predictor scale. For count data with a relatively large number of zeros clustered closely within the covariate space, GAMs might suffer from identifiability problems, especially the Poisson family. For the over-dispersion parameter, the assumption of equal mean and variance inherent in the Poisson GAM might be violated; hence, it has to be replaced by variances that exceed the mean. Our data show a cyclic and seasonal behavior for RSV incidence among children (Fig 1). The Poisson GAM from this analysis demonstrated that climatic factors, including wind speed, rainfall, dew point and visibility, significantly affected RSV incidence. The use of atmospheric condition data help public health officials predict increases in RSV infection incidence among children and help them prepare and respond more swiftly to increasing RSV incidence in low-resource regions or communities. While specific vaccines, antiviral medications, and immunoglobulins are not available to control RSV in these settings, agencies responsible for managing healthcare in crisis-affected populations can increase preparedness for RSV outbreaks by establishing additional patient-isolation areas and bed space, ensuring that all healthcare workers are provided with adequate personal protective equipment (e.g., facial masks and gloves) and appropriate amounts of hand sanitizers and adequate hand-washing facilities for healthcare workers are readily available.

Health education is important; crisis-affected populations should be made aware of the symptoms and signs of RSV, how it spreads, and how to protect themselves and their loved ones. Health education should focus on how to cover coughs, keep appropriate social distancing (e.g., not being too close to others, not shaking hands), and the importance of washing hands with soap. In particular, our model indicates that when the wind speed in knots change from high to low, these interventions should be enhanced to prevent spread of RSV infections in Kenya. In the future, these models could be validated with new RSV surveillance data to see how well they perform to predict increases in RSV incidence particularly for geographical regions with similar climatic attributes to Dadaab.

Supporting information

S1 Fig. Decomposition of RSV time-series data. The variation in the remainder component is approximately the same as the variation in the data. The variation in the seasonal and trend components are about 3–4x smaller than that observed in the data. The long-term trend components appear to be generally increasing. The random (remainder), the bottom plot, show

the residual variation in the data after the long-term trend and seasonality are removed.
(TIFF)

S2 Fig. Decomposition of wind time-series data. The variation in the trend is much smaller than that in the data. The variations in the seasonal and remainder components are marginally smaller than the variation in the data (gray bars on the right).
(TIFF)

S3 Fig. Decomposition of rainfall time-series data. The variations in the seasonal and remainder components do not deviate much from that in the data. The variation in the trend component is roughly 4x less than the variation in the data.
(TIFF)

S4 Fig. Decomposition of temperature time-series data. The trend component has a much smaller variation than that in the data. The seasonal and remainder components show marginally smaller variation than that observed in the data. The long-term trend components appear to be generally increasing.
(TIFF)

S5 Fig. Decomposition of dew time-series data. The trend exibits approximately 3x the overall variation in the Dew data (large gray bar relative to the gray bar on the right-hand of the data plot). The long-term trend components appear to be generally increasing.
(TIFF)

S6 Fig. Decomposition of visibility time-series data. The season accounts for a very small portion of the overall variation in the visibility value (large gray bar relative to the gray bar on the right-hand of the data plot). The long-term trend components seem to be generally decreasing.
(TIFF)

S1 Table. Model diagnostic and performance results.
(DOCX)

S2 Table. ANOVA model for the best performing model, the Poisson GAM with covariate decomposition.
(DOCX)

S3 Table. Covariates and their description. Non-decomposed and decomposed covariates into the seasonal (S), trend (T), and random (R) components.
(DOCX)

S4 Table. Selected Poisson candidate models.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: RN.

Data curation: RN JO.

Formal analysis: RN JO.

Funding acquisition: SMM AG HM.

Investigation: RN.

Methodology: RN JO SMM TNOA AG HM.

Project administration: SMM TNOA AG HM.

Resources: RN SMM AG.

Software: RN JO TNOA.

Supervision: SMM TNOA AG HM.

Validation: TNOA HM.

Visualization: RN JO.

Writing – original draft: RN JO.

Writing – review & editing: JO TNOA HM.

References

1. Iwane MK, Edwards KM, Szilagyi PG, Walker FJ, Griffin MR, Weinberg GA, et al. Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. Pediatrics [Internet]. 2004; 113(6):1758–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15173503

2. Stensballe LG, Devasundaram JK, Simoes EAF. Respiratory syncytial virus epidemics: the ups and downs of a seasonal virus. Pediatr Infect Dis J. 2003; 22(2):S21–32.

3. Leecaster M, Gesteland P, Greene T, Walton N, Gundlapalli A V, Rolfs RT, et al. Modeling the variations in pediatric respiratory syncytial virus seasonal epidemics. BMC Infect Dis [Internet]. 2011; 11 (1):105. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3094225&tool=pmcentrez&rendertype=abstract

4. Simões EAF, DeVincenzo JP, Boeckh M, Bont L, Crowe JE, Griffiths P, et al. Challenges and opportunities in developing respiratory syncytial virus therapeutics. J Infect Dis. 2015; 211:S1–20. https://doi.org/10.1093/infdis/jiu828 PMID: 25713060

5. Mangtani P, Hajat S, Kovats S, Wilkinson P, Armstrong B. The association of respiratory syncytial virus infection and influenza with emergency admissions for respiratory disease in London: an analysis of routine surveillance data. Clin Infect Dis. 2006; 42(5):640–6. https://doi.org/10.1086/499810 PMID: 16447109

6. Mwambi H, Ramroop S, White LJ, Okiro EA, Nokes DJ, Shkedy Z, et al. A frequentist approach to estimating the force of infection and the recovery rate for a respiratory disease among infants in coastal Kenya. Stat Methods Med Res [Internet]. 2011; 20(5):573. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19221168

7. Loeffelholz M, Chonmaitree T. Advances in diagnosis of respiratory virus infections. International Journal of Microbiology. 2010.

8. Piedimonte G, Perez MK. Respiratory syncytial virus infection and bronchiolitis. Pediatr Rev [Internet]. 2014; 35(12):519–30. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84913533040&partnerID=tZOtx3y1

9. Hall CB. Nosocomial respiratory syncytial virus infections: the "Cold War" has not ended. Clin Infect Dis [Internet]. 2000; 31(2):590–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10987726

10. Aintablian N, Walpita P, Sawyer MH. Detection of Bordetella pertussis and respiratory synctial virus in air samples from hospital rooms. Infect Control Hosp Epidemiol [Internet]. 1998; 19(12):918–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9872529

11. Vandini S, Corvaglia L, Alessandroni R, Aquilano G, Marsico C, Spinelli M, et al. Respiratory syncytial virus infection in infants and correlation with meteorological factors and air pollutants. Ital J Pediatr

[Internet]. 2013; 39(1):1. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3553040&tool=pmcentrez&rendertype=abstract

12. Vitor P, Simas M, Gustavo L, Gardinassi A, Nogueira FC, Measso C, et al. Virus Reviews and Research, Analysis of climatic factors impact on RSV infection distribution in children attending child-care at northwest region of Sao Paulo, Brazil. Virus Rev Res. 2012;23–8.

13. Ampofo K, Bender J, Sheng X, Korgenski K, Daly J, Pavia AT, et al. Seasonal invasive pneumococcal disease in children: role of preceding respiratory viral infection. Pediatrics [Internet]. 2008; 122(2):229–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18676537

14. Bloom-Feshbach K, Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller MA., et al. Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. PLoS One. 2013; 8(2):3–4.

15. Wolff C, Kristen-Jenny I, Schettler G, Plessen B, Meyer H, Dulski P, et al. Modern seasonality in Lake Challa (Kenya/Tanzania) and its sedimentary documentation in recent lake sediments. Limnol Oceanogr [Internet]. 2014; 59(5):1621–36. Available from: http://doi.wiley.com/10.4319/lo.2014.59.5.1621

16. Hall CB, Simões EAF, Anderson LJ. Clinical and epidemiologic features of respiratory syncytial virus. Curr Top Microbiol Immunol [Internet]. 2013; 372:39–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24362683

17. DeVincenzo JP, de Swart RL, Osterhaus ADME. 2007 International Congress on Respiratory Viruses, Colorado, USA, 20–22 July 2007. [Internet]. Pediatric Infectious Disease Journal. 2008. p. S1–145. Available from: http://www.pidj.com/

18. Khor C-S, Sam I-C, Hooi P-S, Quek K-F, Chan Y-F. Epidemiology and seasonality of respiratory viral infections in hospitalized children in Kuala Lumpur, Malaysia: a retrospective study of 27 years. BMC Pediatr [Internet]. 2012; 12(1):32. Available from: http://www.biomedcentral.com/1471-2431/12/32

19. Dnb N, Theodoratou E, Rudan I, Campbell H, Nokes DJ, Hnd N, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet. 2010; 375:1545–55. https://doi.org/10.1016/S0140-6736(10)60206-1 PMID: 20399493

20. Kim C, Nyoka R, Ahmed JA, Winchell JM, Mitchell SL, Kariuki Njenga M, et al. Epidemiology of respiratory infections caused by atypical bacteria in Two Kenyan refugee camps. J Immigr Minor Heal. 2012; 14(1):140–5.

21. Brandt Patrick T. and Williams John T.. Multiple time series models. No: 148. Sage; 2007.

22. Brandt Patrick T. and Williams John T.. A Linear Poisson Autoregressive Model: The Poisson AR(p) Model "Political Analysis, vol. 9, no. 2" [Internet]. 2001. 164–184 p. Available from: www.jstor.org/stable/25791638

23. Hastie TJ, Tibshirani R. Generalized additive models [Internet]. Statistical Science. 1990. p. 297–318. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Generalized+additive+models#0

24. Wood SN. Generalized additive models : an introduction with R. Texts Stat Sci [Internet]. 2006; xvii, 392. http://catdir.loc.gov/catdir/enhancements/fy0702/2006040209-d.html%5Cnhttp://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&doc_number=014829960&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA

25. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol. 2002; 156(3):193–203. PMID: 12142253

26. Development Core Team R. R: A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing Vienna Austria. 2011. {ISBN} 3-900051-07-0. http://www.r-project.org

27. Peng RD, Dominici F. Statistical Methods for Environmental Epidemiology with R [Internet]. Media. 2008. 1–151 p. http://link.springer.com/10.1007/978-0-387-78167-9%5Cnhttps://books.google.com/books?id=QbDBxSSXIjsC&pgis=1

28. Gasparrini A. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. J Stat Softw [Internet]. 2011; 43(8):1–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3191524&tool=pmcentrez&rendertype=abstract

29. Gasparrini A. Distributed lag linear and non-linear models for time series data. 14310721250 [Internet]. 2014;1–12. http://143.107.212.50/web/packages/dlnm/vignettes/dlnmTS.pdf

30. Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. J Am Stat Assoc [Internet]. 1988; 83(403):596–610. Available from: http://dx.doi.org/10.2307/2289282

31. Calcagno V, De Mazancourt C. glmulti : An R Package for Easy Automated Model Selection with (Generalized) Linear Models. J Stat Softw. 2010; 34(12):1–29.

32. Jónsson G. The environmental factor in migration dynamics–a review of African case studies. Int Migr Inst Work Pap [Internet]. 2010;1–34. http://www.imi.ox.ac.uk/pdfs/wp-21-gjonsson

33. 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 [Internet]. 2014; 14(1):178. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4021307&tool=pmcentrez&rendertype=abstract

34. Noyola DE, Mandeville PB. Effect of climatological factors on respiratory syncytial virus epidemics. Epidemiol Infect [Internet]. 2008; 136(10):1328–32. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2870732&tool=pmcentrez&rendertype=abstract

35. Haynes AK, Manangan AP, Iwane MK, Sturm-Ramirez K, Homaira N, Brooks WA, et al. Respiratory syncytial virus circulation in seven countries with global disease detection regional centers. J Infect Dis. 2013; 208(SUPPL. 3).

36. Donaldson GC. Climate change and the end of the respiratory syncytial virus season. Clin Infect Dis [Internet]. 2006; 42(5):677–9. Available from: http://discovery.ucl.ac.uk/61843/