Published in final edited form as:

Stat Med. 2019 May 20; 38(11): 1991–2001. doi:10.1002/sim.8081.

Estimating Seasonal Onsets and Peaks of Bronchiolitis with Spatially and Temporally Uncertain Data

Sierra Pugh^a, Matthew J. Heaton^a, Brian Hartman^a, Candace Berrett^a, Chantel Sloan^b,
Amber M. Evans^c, Tebeb Gebretsadik^d, Pingsheng Wu^{d,e}, Tina V. Hartert^e, and Rees L. Lee^f

^aDepartment of Statistics, Brigham Young University, Provo, UT, USA

^bDepartment of Health Science, Brigham Young University, Provo, UT, USA

^cHealth ResearchTx LLC, Treviso, PA, USA

^dDepartment of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

^eDepartment of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

^fNaval Medical Research Unit Dayton, Wright Patterson AFB, OH, USA

Abstract

RSV Bronchiolitis (an acute lower respiratory tract viral infection in infants) is the most common cause of infant hospitalizations in the United States. The only preventive intervention currently available is monthly injections of immunoprophylaxis. However, this treatment is expensive and needs to be administered simultaneously with seasonal bronchiolitis cycles in order to be effective. To increase our understanding of bronchiolitis timing, this research focuses on identifying seasonal bronchiolitis cycles (start times, peaks, and declinations) throughout the continental United States using data on infant bronchiolitis cases from the US Military Health System Data Repository. Because this data involved highly personal information, the bronchiolitis dates in the dataset were “jittered” in the sense that the recorded dates were randomized within a time window of the true date. Hence, we develop a statistical change point model that estimates spatially varying seasonal bronchiolitis cycles while accounting for the purposefully introduced jittering in the data. Additionally, by including temperature and humidity data as regressors, we identify a relationship between bronchiolitis seasonality and climate. We found that in general, bronchiolitis seasons begin earlier and are longer in the southeastern states compared to the western states with peak times lasting approximately one month nationwide.

Keywords

change point model; epidemic; Bayesian; jittering

1. Introduction

Respiratory syncytial virus (RSV) infects virtually all children in the United States before they reach three years old [1]. Most children infected with RSV will develop upper respiratory tract infection, and experience mild cold-like symptoms. However, approximately 2–3% of all infants in the United States require hospitalization for RSV bronchiolitis (a lower respiratory tract infection) each year [2]. Because the presence of RSV can only be confirmed via laboratory testing and because up to 80% of infant bronchiolitis cases are induced by RSV infection, bronchiolitis is often used as a surrogate to understand RSV spatio-temporal dynamics. In similar manner, this paper will focus on cases of infant bronchiolitis as a surrogate measure of RSV.

Infants who experience bronchiolitis remain at risk for long-term outcomes such as asthma later in life [3]. Severe outcomes are most common among infants who were born at less than 36 weeks gestation or have congenital abnormalities [4, 5, 6]. These high-risk infants typically receive immunoprophylaxis via a series of injections during the winter virus season to prevent RSV infection and subsequent bronchiolitis. There is currently no FDA approved RSV vaccine, though a number are in phase 3 clinical trials (see [7] and clinicaltrials.gov).

In temperate northern climates, RSV primarily circulates between the months of November and April. However, when a seasonal epidemic begins depends on location [8, 9]. Hospitalization season begins earlier and persists longer than the beginning and end of bronchiolitis season for the general population [10], making understanding of seasonal timing imperative to reducing bronchiolitis rates among high-risk groups. To this end, this research considers medical claims data for all infants under the age of one between July 2012 and July 2013 from the Department of Defense (DoD) Military Health System Data Repository (MDR). The MDR serves all active and retired military personnel and their dependents. While the MDR includes the entire United States (US) and outside the contiguous United States (OCONUS), our analysis focuses on the continental US, Alaska and Hawaii. The use of the MDR cohort includes infants living in multiple locations and environments across the United States who are of different socioeconomic backgrounds over a ten year time span. Therefore underlying factors associated with a specific location or year can be compared and described, and will not drive the results in a singular biased direction. While the MDR cohort does have some limitations, including less representation from certain highly rural areas of the United States and incomplete data on racial background, it remains statistically powerful and representative of the broader population of infants in the United States.

Originally, the MDR contained the date and location of birth for each child as well as the date a child was treated for bronchiolitis, if applicable. However, because personal health information (PHI) from the MDR is restricted, the analytic database included “jittered” dates of birth, dates of bronchiolitis treatment and locations for the children. For example, the recorded birthday in the database available to us for analysis has been randomized within a given length of time from the true birthday (the recorded dates of treatment and location of the child were also randomized in the same manner). In the statistical literature and recent calls from the National Institutes of Health (NIH), this lack of knowledge about a patient’s

geographic position and date of infection is called “spatio-temporal uncertainty” and we will adopt this phrase throughout this paper (see e.g., [11, 12]).

Given the importance of timely administration of treatments, in this research we seek to use the MDR to (i) estimate the onset and peaks of bronchiolitis seasons so we can inform when to administer treatment and (ii) determine the impact of seasonal meteorology on peak bronchiolitis rates. In this regard, the primary focus of this research is statistical inference for parameters related to the timing of bronchiolitis and subsequent impact of seasonal meteorology as opposed to prediction. To accomplish these goals, we develop a novel temporal change point regression model where the change points are estimated under a Bayesian framework. Change point models are used in a variety of settings such as genetics, ecology, and economics to model inflection points in time as well as in space [13, 14, 15, 16, 17]. Because the onset and peak times of bronchiolitis seasons vary by location, as a novel statistical contribution we develop a change point model where the change points and associated parameters of the regression model vary over space which, unlike previous studies, allows us to make direct inference on the spatial variability of temporal parameters. We further propose to smooth across spatial regions using basis function expansions to allow information to be borrowed and facilitate parameter estimation.

An important aspect of accomplishing goals (i) and (ii) is to account for the spatial and temporal jittering in the data (i.e. spatio-temporal uncertainty). For example, [18] and [11] suggest excluding spatio-temporal uncertainty in the modeling framework can lead to “false positive” conclusions due to understating uncertainty. Methods by [19, 20, 21] and [12] model spatial uncertainty but do not consider temporal uncertainty as is needed here. Additionally, these methods were applied to smaller spatial regions and would be computationally challenging to extend to larger spatial regions such as the contiguous US under consideration in this research. In order to account for the spatio-temporal uncertainty in the MDR, we assume the parameters of our change point model are regionally (but not globally) constant within a partition of the spatio-temporal domain. We subsequently show this assumption leads to a computationally simple approach to account for the spatio-temporal uncertainty within a Markov chain Monte Carlo (MCMC) framework.

In summary, this paper provides a spatio-temporal change point model that accounts for the spatio-temporal uncertainty (“jittering”) in the MDR data. This study also provides insight on seasonal timing of bronchiolitis throughout the continental US and the impact of seasonal meteorology on the peaks. The paper outline is as follows. Section 2 outlines our spatio-temporal change point model. Section 3 applies our model to the data from the MDR, and Section 4 provides discussion and suggestions for future research.

2. A Spatio-Temporal Change Point Model

For purposes of this research, let $\mathcal{R} \subset \mathbb{R}^2$ denote the spatial domain of the contiguous US which we partition into R disjoint subregions $\mathcal{R}_1, \dots, \mathcal{R}_R$. Likewise, let $Y = \{1, \dots, D\}$ denote the set of $D = 365$ days of the year which we also partition into T disjoint sets of time periods $\mathcal{T}_1, \dots, \mathcal{T}_T$. In this section, we describe a hierarchical model that relates the

information in the MDR to spatio-temporal probabilities of bronchiolitis. In this way, we use this model to estimate start times, peaks, and end times of bronchiolitis seasons.

2.1. Distribution for Observed Data

Let b_i denote the day of birth and $s_i \in \mathcal{R}$ be the birth location for child i where $i = 1, \dots, N$.

As a distribution for $\{b_i, s_i\}$, we first assume that the birthdate is independent of the birth location such that $[b_i, s_i] = [b_i][s_i]$ where the notation $[\cdot]$ is used to specify a generic probability distribution. Census information suggests that birthdays are approximately uniformly distributed across dates in the calendar year so, for simplicity, we assume that $[b_i] \propto 1$. Populations, on the other hand, are not distributed uniformly across the contiguous US. Hence, we assume that birth location follows some density function which we denote by $p(s)$.

Based on the previous work of (author?) [1] who found that RSV infects virtually all children, we define child i to be “susceptible” to bronchiolitis in time period t if child i was born within a year of $\mathcal{T}_t$ for all children. For children who were treated for bronchiolitis, let $d_i \in \{1, \dots, D\}$ denote the treatment day for child i . We assume a constant probability of contracting bronchiolitis within time period and region such that

$$\mathbb{P}r(d_i | s_i \in \mathcal{R}_r, b_i) = \sum_{t=1}^T \frac{\rho_{rt}}{|\mathcal{T}_t|} \mathbb{1}_{\{d_i \in \mathcal{T}_t\}} \mathbb{1}_{\{b_i \in \mathcal{T}_t^*\}} \quad (1)$$

where $\mathbb{1}_{\mathcal{A}}$ is an indicator for the set $\mathcal{A}$ and we use $\mathcal{T}_t^*$ to denote all time periods within a year prior to $\mathcal{T}_t$ (in this way, according to (1), $\mathbb{P}r(d \in \mathcal{T}_t | s_i \in \mathcal{R}_r, b_i) = 0$ if child i is not susceptible to bronchiolitis in the t^{th} time period). Intuitively, given $s_i \in \mathcal{R}_r$ and the child is susceptible, (1) assumes that the probability of contracting bronchiolitis for any day $d \in \mathcal{T}_t$ is equal yielding $\mathbb{P}r(d_i \in \mathcal{T}_t | s_i \in \mathcal{R}_r, b_i \in \mathcal{T}_t^*) = \rho_{rt}$ such that ρ_{rt} is the probability of contracting bronchiolitis in $\mathcal{R}_r \times \mathcal{T}_t$. The $\{\rho_{rt}\}$ are, then, the primary parameters of interest. The estimation of ρ_{rt} can help determine the start times, peaks, and end times of the bronchiolitis season. Likewise, regressing ρ_{rt} onto meteorological covariates can help determine the effect of these covariates on seasonal bronchiolitis rates.

The assumption that the probability of contracting bronchiolitis is uniform within each $\mathcal{R}_r \times \mathcal{T}_t$ facilitates statistical learning of ρ_{rt} . That is, this assumption allows us to borrow information from all bronchiolitis cases in $\mathcal{R}_r \times \mathcal{T}_t$ to estimate ρ_{rt} . To see this, let Y_{rt} denote the number of children in $\mathcal{R}_r \times \mathcal{T}_t$ who contract bronchiolitis and N_{rt} be the number of children who are susceptible in $\mathcal{R}_r \times \mathcal{T}_t$. Because contracting bronchiolitis is a binary outcome and each susceptible child in $\mathcal{R}_r \times \mathcal{T}_t$ has an equal probability of contracting bronchiolitis, it follows that $Y_{rt} | N_{rt} \sim \text{Bin}(N_{rt}, p_{rt})$ and standard estimation of methods for p_{rt} (e.g., logistic regression) can be used. Our model for ρ_{rt} is given in Section 2.2.

The validity of the uniform probability within each $\mathcal{R}_r \times \mathcal{T}_t$ assumption in (1), however, is contingent upon the size of $\mathcal{R}_r$ and the number of days in $\mathcal{T}_t$. That is, the partitions $\mathcal{R}_1, \dots, \mathcal{R}_R$ and $\mathcal{T}_1, \dots, \mathcal{T}_T$ need to be sufficiently fine for this assumption to hold (or approximately hold). For example, if $\mathcal{T}_t$ corresponds to a 3 month time period, then this uniform assumption is likely to be violated because ρ_{rt} will almost certainly change within $\mathcal{T}_t$. Due to the jittering in the MDR data, we do not directly observe b_i , s_i and d_i . Rather, for each child in the MDR, we observe a jittered version of these dates. Let b_i^* , s_i^* and d_i^* denote the jittered birthday, birth location, and treatment date. Given that dates and locations were jittered uniformly, we assume

$$b_i^* | b_i \sim \mathcal{U}(\{b_i - K, \dots, b_i + K\}) \quad (2)$$

and

$$d_i^* | d_i \sim \mathcal{U}(\{d_i - K, \dots, d_i + K\}) \quad (3)$$

where $\mathcal{U}$ denotes the uniform distribution. Similarly, we assume that $s_i^* | s_i \sim \mathcal{U}(\mathcal{C}_\ell(s_i))$ where $\mathcal{C}_\ell(s)$ is a square centered at s with side-length ℓ (hence, the jittering length is $\ell/2$). We treat the jittering lengths K and $\ell/2$ as known but do not share these values to maintain confidentiality.

2.2. Model Parameterization

To estimate start, peak, and end times for bronchiolitis seasons, we let ρ_{rt} follow a spatio-temporal change point model (i.e. a linear spline) where

$$\log\left(\frac{\rho_{rt}}{1 - \rho_{rt}}\right) = \quad (4)$$

$$\begin{cases} \alpha_r & \text{if } t \leq \Delta_{r0} \\ \alpha_r + \frac{\psi_r}{\Delta_{r1}}(t - \Delta_{r0}) & \text{if } \Delta_{r0} < t \leq \Delta_{r0} + \Delta_{r1} \\ \alpha_r + \psi_r & \text{if } (\Delta_{r0} + \Delta_{r1}) < t \leq (\Delta_{r0} + \Delta_{r1} + \Delta_{r2}) \\ \alpha_r + \psi_r - \frac{\psi_r}{\Delta_{r3}}\left(t - \sum_{j=0}^2 \Delta_{rj}\right) & \text{if } (\Delta_{r0} + \Delta_{r1} + \Delta_{r2}) < t \leq (\Delta_{r0} + \Delta_{r1} + \Delta_{r2} + \Delta_{r3}) \\ \alpha_r & \text{if } t > (\Delta_{r0} + \Delta_{r1} + \Delta_{r2} + \Delta_{r3}). \end{cases}$$

Each parameter in the above change point model can be interpreted in an intuitive manner. Namely, for Δ_{r0} time periods bronchiolitis is “offseason” such that the probability of bronchiolitis in region $\mathcal{R}_r$ is low and is given by $\rho_{rt} = \text{logit}^{-1}(\alpha_r)$. For the next Δ_{r1} time periods (i.e. $t \in (\Delta_{r0}, \Delta_{r0} + \Delta_{r1}]$), $\text{logit}(\rho_{rt})$ is increasing linearly until at time $\Delta_{r0} + \Delta_{r1}$ it hits a peak of $\alpha_r + \psi_r$ such that $\psi_r > 0$ denotes the increase in probability from the offseason rate of α_r . After staying at the peak for Δ_{r2} time periods, $\text{logit}(\rho_{rt})$ then decreases linearly for Δ_{r3} time periods until it returns to the offseason rate at time $\Delta_{r0} + \Delta_{r1} + \Delta_{r2} + \Delta_{r3}$. An illustration of this curve is shown in Figure 1.

In (4), the parameters determine the time frame of a bronchiolitis season (e.g. the start of the bronchiolitis season in region r would be Δ_{r0} while bronchiolitis rates in region r peak at time $\Delta_{r0} + \Delta_{r1}$). Because a bronchiolitis season cannot extend beyond a year, we need to incorporate a constraint that $\Delta_{r0} + \Delta_{r1} + \Delta_{r2} + \Delta_{r3}$ must be less than one year. To incorporate this constraint, we use a stick-breaking construction such that $\Delta_{r0} = T\omega_{r0}$ and

$$\Delta_{rJ} = T\omega_{rJ} \prod_{j=0}^{J-1} (1 - \omega_{rj}) \quad (5)$$

for $J \in \{1, 2, 3\}$ where $\omega_{rj} \in (0, 1)$ can be interpreted as a proportion of time periods corresponding to Δ_{rj} . Intuitively, ω_{r0} is the proportion of all time periods T that bronchiolitis rates in $\mathcal{R}_r$ spent at baseline. Of the remaining time periods after baseline ($T(1 - \omega_{r0})$), $100 \times \omega_{r1}\%$ of these correspond to the time in which bronchiolitis rates were increasing. Interpretations for ω_{r2} and ω_{r3} can be extended similarly.

Let $\alpha = (\alpha_1, \dots, \alpha_R)'$ denote the vector of baseline (“offseason”) rates and let $\omega_j = (\omega_{1j}, \dots, \omega_{Rj})'$ for $j \in \{0, \dots, 3\}$. Because we partitioned the spatial domain $\mathcal{R}$ into the disjoint regions $\mathcal{R}_1, \dots, \mathcal{R}_R$, each of α and $\omega_0, \dots, \omega_3$ are defined areally (i.e. describe spatial regions) and exhibit spatial correlation. Particularly important for this research, we wish to exploit this correlation by borrowing information across neighboring regions to facilitate estimation in regions with few observations. To capture this spatial structure, we use the Moran basis functions described in [22] and set

$$\alpha = \theta_{\alpha 0} \mathbf{1}_R + \mathbf{M} \theta_{\alpha} \quad (6)$$

where $\theta_{\alpha} = (\theta_{\alpha 1}, \dots, \theta_{\alpha P})'$ is a vector of unknown coefficients corresponding to the Moran bases $\mathbf{M}$ and $\mathbf{1}_R$ is a column vector of R ones. For each ω_j , we need to enforce that $\omega_{rj} \in (0, 1)$ for all r and j . Hence, we define

$$\omega_j = \Phi(\theta_{j0} \mathbf{1}_R + \mathbf{M} \theta_j) \quad (7)$$

where $\Phi(v)$ is the Gaussian cumulative distribution function applied to each element in the vector v .

The vector $\psi = (\psi_1, \dots, \psi_R)' > \mathbf{0}$ represents the increase in bronchiolitis rates experienced at peak season. As stated in the Introduction, one of the goals of this analysis is to understand why certain areas of the US experience higher bronchiolitis rates than others. Hence, we model,

$$\log(\psi) = \mathbf{X} \beta + \mathbf{M}_{\psi} \theta_{\psi} \quad (8)$$

where $\mathbf{X}$ is a $R \times (Q + 1)$ matrix of covariates (where the first column of $\mathbf{X}$ is unity) with corresponding coefficients $\beta = (\beta_0, \dots, \beta_Q)'$ and $\mathbf{M}_{\psi}$ is the matrix of spatial Moran bases (without a column of 1's) constructed to be orthogonal to $\mathbf{X}$ with corresponding coefficients θ_{ψ} . The vector β then identifies factors which either contribute to (e.g. $\beta_q > 0$) or detract from (e.g. $\beta_q < 0$) higher bronchiolitis rates.

2.3. Parameter Estimation

For this research, we opt to use a Bayesian approach to parameter estimation. In this case, we argue that the Bayesian approach is natural because parameter estimation via Markov chain Monte Carlo (MCMC) sampling will account for the uncertainty associated with $b_i^{\star}, s_i^{\star}$ and $d_i^{\star}$ via repeated sampling of the true birthdates (b_i), birth locations (s_i), and treatment dates (d_i) given the jittered dates. Admittedly, a multiple imputation algorithm could also be devised to account for this uncertainty, but we leave this possibility as future work and adopt the Bayes approach here.

The unknown parameters from the model in Section 2.2 are the basis coefficients $\theta_a, \theta_0, \theta_1, \theta_2, \theta_3, \theta_\psi$ and the covariate coefficients β . Recall that the baseline probability, $\text{logit}^{-1}(\alpha_r)$, should be low so as to reflect an offseason rate. Previous work from, e.g., [9] estimate that these offseason probabilities should be approximately spatially constant with a mean rate of 0.001 (i.e. $\alpha_r \approx -6.6$ for all r) and no offseason probability should exceed 0.005 (i.e. $\alpha_r < -5.8$ for all r). Hence, we assume the informative prior $\theta_{a0} \sim \mathcal{N}(-6.6, 0.01^2)$ and $\theta_\alpha \sim \mathcal{N}(0, (0.27^2)I)$ to reflect these prior beliefs. In contrast, we wish the data to inform the seasonal timing and peaks of the bronchiolitis season. Hence, we assume vague $\mathcal{N}(0, 2^2)$ priors for the remaining θ and β parameters. We verified these priors are quite vague by simulating draws from the prior and back transforming to the original variables.

We perform posterior inference by retaining 250,000 draws from the joint posterior distribution of all model parameters after an initial 100,000 iterations of “burn-in” in an MCMC algorithm. In order to account for the uncertainty associated with our observed jittered data, at each iteration of the MCMC algorithm we sample the region and temporal indices of b_i, s_i and d_i (note that the region and temporal indices are sufficient statistics for ρ_{rt}). For example, the complete conditional probability of the temporal index for the treatment date is given by

$$\begin{aligned} \mathbb{P}\text{r}(d_i \in \mathcal{T}_t | d_i^*, -) &\propto \mathbb{P}\text{r}(d_i^*, d_i \in \mathcal{T}_t, -) \\ &= \sum_{u \in \mathcal{T}_t} \mathbb{P}\text{r}(d_i^* | u, -) \mathbb{P}\text{r}(u | -) \\ &\propto \sum_{u \in \mathcal{T}_t} \mathbb{1}_{\{d_i^* - K \leq u \leq d_i^* + K\}} \rho_{rt} \\ &= |\{d_i^* - K, \dots, d_i^* + K\} \cap \mathcal{T}_t| \rho_{rt} \end{aligned} \quad (9)$$

where “ $-$ ” is shorthand notation to denote all other parameters. Because d_i^* and K are both known, the set intersection $|\{d_i^* - K, \dots, d_i^* + K\} \cap \mathcal{T}_t|$ can be calculated just once *a priori* (outside the MCMC loop) for all $t = 1, \dots, T$ leading to a simple update for the temporal index of the actual treatment date. A similar derivation can be used for the temporal index of the true birthday b_i .

When drawing the region index of the birth location s_i at each iteration of the MCMC algorithm, similar calculations to (9) above show the complete conditional probability of the regional index to be

$$\mathbb{P}\text{r}(s_i \in \mathcal{R}_r | s_i^*, -) \propto \rho_{rt} \int_{\mathcal{R}_r \cap \mathcal{C}(s_i^*)} p(u) du \quad (10)$$

where, recall, that $\mathcal{C}_\ell(s_i^\star)$ is the square of side-length ℓ centered at $s_i^\star$ and $p(s)$ is the spatial density of birth locations over $\mathbb{R}$ (see Section 2.1). Because $p(s)$ is unknown, the integral in (10) is also unknown. [23] treat $p(s)$ as an additional unknown parameter estimated from the data but doing so here would require calculating (10) at each iteration of our MCMC algorithm which vastly increases the computation required for model inference. Alternatively, $p(s)$ could be approximated by some known population distribution (e.g. from the US census) so that calculation of (10) could be done outside the MCMC algorithm. However, the data we consider here are military health records so that the spatial distribution of these individuals (i.e. locations of high and low population density) would not match the spatial distribution from the general US census but rather follow large military base locations. Hence, here we carefully select the partition $\mathcal{R}_1, \dots, \mathcal{R}_R$ so that $\mathcal{R}_r \cap \mathcal{C}_\ell(s_i^\star) = \emptyset$ for all but a single $r \in \{1, \dots, R\}$. Using this carefully chosen partition, given the jittered location $s_i^\star$, we know with certainty the regional index of the true birth location $\{r: s_i \in \mathcal{R}_r\}$ (but, notably, not the true birth location itself). In other words, our regions are defined such that $\Pr(s_i \in \mathcal{R}_r | s_i^\star, -) \equiv 1$ for a certain region r . This, then, allows us to avoid sampling the regional index at each iteration and greatly reduces computation time.

The remaining parameters associated with ρ_H were drawn using adaptive Metropolis methods [24]. Convergence of our MCMC algorithm was monitored via trace plots and Monte Carlo standard error (MCSE) diagnostics [25].

3. Application to Military Data Repository

To apply the methodology described in Section 2 to the data on bronchiolitis from the MDR, we partitioned $\mathcal{R}$ (the continental US) into $R=127$ spatial regions. Figure 2 gives the centroids of each region and $\mathcal{R}_k$ is the set of points nearest the centroid. This choice of regionalization was chosen based on a trade-off between computational tractability and the ability to still capture regional scale variability in bronchiolitis rates. Namely, these regions are sufficiently fine to describe the spatial variability of bronchiolitis seasons yet sufficiently large so that $\mathcal{R}_r \cap \mathcal{C}_\ell(s_i^\star) = \emptyset$ for all but a single $r \in \{1, \dots, R\}$ to facilitate computation. The temporal partitions $\mathcal{T}_1, \dots, \mathcal{T}_T$ are taken to be half month intervals, resulting in a temporal resolution fine enough to accurately describe temporal variations during a bronchiolitis season while large enough to allow for borrowing of information from all bronchiolitis cases in $\mathcal{R}_r \times \mathcal{T}_t$ and reduce the computational burden. By choosing half month intervals rather than a smaller interval, $\Pr(d_i \in \mathcal{T}_t | d_i^\star, -) = 0$ for more $\mathcal{T}_t$, generally. This means less $\mathcal{T}_t$ need to be considered when accounting for the temporal uncertainty both because T is smaller and because $\Pr(d_i \in \mathcal{T}_t | d_i^\star, -) = 0$ for more $\mathcal{T}_t$.

To determine the impact of seasonal meteorology on the peaks, we define X as consisting of a column of ones, average deviations in temperatures during τ_1 (the time period where bronchiolitis rates are increasing), average deviations in humidity during τ_1 , and the interaction between the temperature and humidity deviations during τ_1 . Because the

columns of $\mathbf{X}$ are defined in terms of β_1 , the temperature and humidity values included in the columns of $\mathbf{X}$ could be updated at each iteration of the MCMC algorithm. At early stages of this research, we attempted to update $\mathbf{X}$ at each iteration but doing so greatly increased the computational complexity of the algorithm (partially because $\mathbf{X}$ is used to define the $\mathbf{M}$ matrices below) while negatively impacting the mixing of the chain. For this reason, we *a priori* estimated β_1 using maximum likelihood estimation and fixed $\mathbf{X}$ accordingly. We also chose to model the temperature and humidity residuals after removing the spatially varying mean, rather than the actual values to avoid confounding the effects of seasonal meteorology with spatial location (e.g. southern states have a warmer climate than northern states such that deviations in temperature and humidity from the local climate better account for local adaptations to weather).

Finally, the Moran basis $\mathbf{M}$ was composed of the 61 columns associated with each positive eigenvalue according to the methods described in [22]. This choice resulted in a 48% reduction in the spatial dimension, which further facilitated analysis. The matrix $\mathbf{M}_\psi$ was also constructed using the methods in [22] but added the further constraint that $\mathbf{M}_\psi$ is orthogonal to the columns of $\mathbf{X}$.

Figure 3 displays various aspects of the spatial variability in the 2012–2013 bronchiolitis season (birth dates between 2011 and 2013 were used to determine the number of susceptible children) according to the posterior mean of the 250,000 post-burn draws from the MCMC algorithm. (Uncertainty estimates for each of these maps are also available from the posterior draws but are not shown here for brevity.) Namely, Figure 3 (a) shows the posterior mean for the time when that region first reached peak rates ($\beta_0 + \beta_1$) which ranges from November 2012 to March 2013. In general, the bronchiolitis season seemed to peak earlier in southeastern US with the season peaking around December or January. In comparison, seasons on the West coast tended to peak around February or March 2013. This agrees with the findings of [9] who also saw an earlier onset of the bronchiolitis in the southeast.

Figure 3 (b) displays the posterior mean for β_2 , the length of time bronchiolitis rates were at the peak. In contrast to Panel (a), the time at peak has less spatial variability. That is, most regions experienced peak bronchiolitis rates for approximately one month. We note this as an interesting conclusion in that peak rate times last are approximately equal across the contiguous US with a few notable exceptions in south Texas, South Dakota, Iowa, and Pennsylvania. However, other than south Texas, these anomaly areas had very little data such that one or two cases dramatically increased the proportion of all susceptible children who contracted bronchiolitis. Additionally, the few data points available in these regions resulted in higher uncertainty in the time at peak.

Consider, next, Figure 3 (c) that shows the total time that rates were higher than baseline ($\beta_1 + \beta_2 + \beta_3$). Similar to Panel (a), the East coast and West coast seem to have different patterns in the length of time above baseline. Namely, the bronchiolitis season was longer in the Eastern US and shorter in the Western US. This difference in the length of the season along the southern coast can partially be explained by Figure 4 which shows bronchiolitis probabilities as a function of time for several regions well dispersed across the continental

US labeled by major cities contained within each region. Notably, for Orlando in Figure 4 (a), the underlying data (thin black line) shows there is not a well-pronounced start and end to the bronchiolitis season but rather slow increases followed by slow decreases that essentially includes most of the calendar year. By contrast, the bronchiolitis season in Salt Lake City, Minneapolis and Salem (Figure 4 (b), (c) and (d)) show clear peaks to the bronchiolitis season. This result is consistent with previous studies that found tropical regions tend to have less defined seasons with a generally lower proportion of children being infected [26, 27].

Finally, Figure 3 (d) shows the posterior mean of ρ during the peak season (i.e. the maximum probability of bronchiolitis). The probability of a case of bronchiolitis appears to be lower along the coasts and along the Canadian border. In most areas, during the peak season about 1 or 2 percent of infants are expected to contract bronchiolitis but some areas (mainly in Texas) experience rates as high as 5 or 6 percent. To partially explain this spatial variability among peak bronchiolitis rates, Table 1 shows the posterior median and 95% credible interval for $\exp\{\beta\}$. We found temperature had a statistically significant affect on peak bronchiolitis rates while humidity did not. That is, because the coefficient for temperature was positive, we conclude that warmer winters (relative to the local average temperature) are associated with more extreme peaks in bronchiolitis rates. More specifically, we can say ψ is expected to be 1.0092 times larger as the average deviation in temperature during τ_1 increases by one.

Figure 4 further shows the spatial variability in bronchiolitis seasonality for the 2012–2013 season where we display the entire seasonal curve for a few chosen regions. Notably, the fitted curves closely resemble the data suggesting adequate model fit. Using these curves, we can characterize the cumulative impact of the bronchiolitis season to the area by calculating the area under each seasonal curve (see Table 2). Orlando has the largest cumulative impact and is statistically different from the cumulative impact in Salem and Minneapolis. We can also see that despite Salt Lake City and Minneapolis showing very differently shaped seasons, the cumulative impact is approximately the same for both cities. Salt Lake City experienced a large probability of bronchiolitis at the peak, but these peak times were not drawn out. However, Minneapolis experienced the longest peak time (lasting almost two months) but, notably, with a smaller peak rate. Both had approximately the same area under its seasonal curve, but they distributed the mass very differently.

We analyzed data from Alaska and Hawaii as its own region which we assume as independent of the other regions due to the lack of spatial contiguity. The posterior means of these epidemic curves are shown as Panels (e) and (f) in Figure 4. Alaska seems to follow a pattern relatively similar to those we saw in the contiguous US, but the peak rate is lower and lasts longer than most. Hawaii has a low peak rate as well, and maintains the peak considerably longer than seen for any of the regions in the contiguous US. Interestingly, Hawaii does not seem to follow the same pattern as the regions with more temperate climates (e.g. Florida) where the total time the rates are above baseline is shorter and the peak is maintained for much longer.

4. Conclusions

RSV-associated bronchiolitis in children is correlated with various respiratory health complications later in life [3]. Because prevention of RSV with immunoprophylaxis for high-risk infants is dependent upon time-sensitive, in this research we developed a temporal change point model that was used to (i) estimate the onset and peaks of RSV and (ii) determine the impact of seasonal meteorology on the peaks. The ability to predict the onset of the bronchiolitis season in specific regions using available data would allow regional recommendations for the timing of the start of immunoprophylaxis injections in any given season and/or the implementation of any newly discovered primary prevention, such as an RSV vaccine.

The primary challenge of estimating the parameters associated with our change point model was that the information in the MDR was jittered to preserve patient confidentiality (i.e. included spatio-temporal uncertainty). Hence, our proposed modeling strategy included the jittered locations and dates of infection in a hierarchical framework. The associated uncertainty in these jittered locations and dates could then be accounted for via iterative sampling in a MCMC framework. In summary, we proposed a modeling strategy that hierarchically deals with the jittered locations and dates of bronchiolitis cases to reveal epidemic curves and assess climate associations with RSV-associated bronchiolitis. While our model was specific to the MDR data, the model framework could be applied to any dynamic time series data or information that is tainted purposefully with noise.

This work parameterized the jittering distributions based on discussions about how the original data was jittered. However, we note that our use of uniform distributions may not apply to all applications. For example, alternative jittering distributions such as the Gaussian distribution may be more appropriate for other applications [see ?]. In general, parameterizing this distribution is a difficult task as other geocoding errors (e.g. typos) are challenging to parameterize without strong prior information.

This methodology could also be used in some cases where the information is not purposefully tainted but still has uncertainty (e.g. geocoding). In these cases we could estimate the distribution of the actual data given the recorded data. For example, if we know a case occurred near a certain date, we could explore the accuracy of the recorded is and select an appropriate distribution as the temporal and spatial jittering distribution (such as a normal distribution with the variance based on how accurate we believe the information to be).

The developed methodology was contingent on a partition of the spatial and temporal domain. The regionalization used in this research shown in Figure 2 was chosen as a tradeoff between the ability to capture important features in spatial variability of the bronchiolitis season, the degree of statistical learning of the change point model, and computational tractability. While our regionalization was useful here to discover spatial variations in the seasonal pattern of bronchiolitis over a large spatial domain, this partition is likely too large to be able to estimate other important drivers of the bronchiolitis seasons such as pollution levels which operate at finer spatial scales [23]. In this regard, further research is warranted

on the choice of regionalization so that these important factors could be included in a regression model for the peak bronchiolitis rates. Or, potentially a spatial partition could be learned from the data via some form of a Dirichlet process mixture model (see, e.g., [28]), but this avenue is left open for future research.

The results of this analysis are only applicable to the 2012–2013 bronchiolitis season. Epidemiologically, it would be interesting to contrast the spatial variation seen here with other years. Namely, open questions include (i) what other seasonal factors are associated with bronchiolitis, (ii) is the estimated positive effect of temperature consistent from year to year, and (iii) how much variability in peak bronchiolitis exists from year to year?

Acknowledgments

The authors wish to thank the Navy and Marine Corps Public Health Center for its support during the conduct of this study and Christy Fox for helping to compile the MDR database and providing clarifications on the jittering process.

Disclaimers: Dr. Rees Lee a military service member. This work was prepared as part of his official duties. Title 17 U.S.C. §105 provides that 'Copyright protection under this title is not available for any work of the United States Government.' Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, the National Institutes of Health nor the U.S. Government.

The study protocol was approved by the Naval Medical Research Unit Dayton Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects.

Contract/grant sponsor: Research reported in this publication was supported by the National Institute of Environmental Health Science of the National Institutes of Health under award number R03ES025295.

References

- [1]. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA, Auinger P, Griffin MR, Poehling KA, Erdman D, et al. The burden of respiratory syncytial virus infection in young children. *New England Journal of Medicine* 2009; 360(6):588–598. [PubMed: 19196675]
- [2]. Meissner HC. Viral bronchiolitis in children. *New England Journal of Medicine* 2016; 374(1):62–72. [PubMed: 26735994]
- [3]. James KM, Gebretsadik T, Escobar GJ, Wu P, Carroll KN, Li SX, Walsh EM, Mitchel EF, Sloan C, Hartert TV. Risk of childhood asthma following infant bronchiolitis during RSV season. *The Journal of allergy and clinical immunology* 2013; 132(1):227. [PubMed: 23419541]
- [4]. Helfrich AM, Nylund CM, Eberly MD, Eide MB, Stagliano DR. Healthy late-preterm infants born 33–36+ 6 weeks gestational age have higher risk for respiratory syncytial virus hospitalization. *Early human development* 2015; 91(9):541–546. [PubMed: 26186560]
- [5]. Stagliano DR, Nylund CM, Eide MB, Eberly MD. Children with down syndrome are high-risk for severe respiratory syncytial virus disease. *The Journal of pediatrics* 2015; 166(3):703–709. [PubMed: 25596098]
- [6]. Granbom E, Fernlund E, Sunnegårdh J, Lundell B, Naumburg E. Respiratory tract infection and risk of hospitalization in children with congenital heart defects during season and off-season: A swedish national study. *Pediatric cardiology* 2016; 37(6):1098–1105. [PubMed: 27090653]
- [7]. Giersing BK, Karron RA, Vekemans J, Kaslow DC, Moorthy VS. Meeting report: WHO consultation on respiratory syncytial virus (RSV) vaccine development, geneva, 25–26 april 2016. *Vaccine* 2017;.
- [8]. Noveroske DB, Warren JL, Pitzer VE, Weinberger DM. Local variations in the timing of RSV epidemics. *BMC Infectious Diseases* 2016; 16(1):674. [PubMed: 27835988]

- [9]. Sloan C, Heaton M, Kang S, Berrett C, Wu P, Gebretsadik T, Sicignano N, Evans A, Lee R, Hartert T. The impact of temperature and relative humidity on spatiotemporal patterns of infant bronchiolitis epidemics in the contiguous united states. *Health & Place* 2017; 45:46–54. [PubMed: 28285184]
- [10]. Glick AF, Kjelleren S, Hofstetter AM, Subramony A. RSV hospitalizations in comparison with regional RSV activity and inpatient palivizumab administration, 2010–2013. *Hospital Pediatrics* 2017; :hpeds-2016.
- [11]. Cressie N, Kornak J. Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statistical science* 2003; :436–456.
- [12]. Fanshawe T, Diggle P. Spatial prediction in the presence of positional error. *Environmetrics* 2011; 22(2):109–122.
- [13]. Minin VN, Dorman KS, Fang F, Suchard MA. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* 2005; 21(13):3034–3042. [PubMed: 15914546]
- [14]. Beckage B, Joseph L, Belisle P, Wolfson DB, Platt WJ. Bayesian change-point analyses in ecology. *New Phytologist* 2007; 174(2):456–467. [PubMed: 17388908]
- [15]. Bai J Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 1997; 79(4):551–563.
- [16]. Raftery AE. Change point and change curve modeling in stochastic processes and spatial statistics. *Journal of Applied Statistical Science* 1994; 1(4):403–423.
- [17]. Otto P, Schmid W. Detection of spatial change points in the mean and covariances of multivariate simultaneous autoregressive models. *Biometrical Journal* 2016; 58(5):1113–1137. [PubMed: 27374408]
- [18]. Gabrosek J, Cressie N. The effect on attribute prediction of location uncertainty in spatial data. *Geographical Analysis* 2002; 34(3):262–285.
- [19]. Barber JJ, Gelfand AE, Silander JA. Modelling map positional error to infer true feature location. *Canadian Journal of Statistics* 2006; 34(4):659–676.
- [20]. Zimmerman DL. Estimating the intensity of a spatial point process from locations coarsened by incomplete geocoding. *Biometrics* 2008; 64(1):262–270. [PubMed: 17680833]
- [21]. Chakraborty A, Gelfand AE, Wilson AM, Latimer AM, Silander JA Jr. Modeling large scale species abundance with latent spatial processes. *The Annals of Applied Statistics* 2010; :1403–1429.
- [22]. Hughes J, Haran M. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2013; 75(1):139–159.
- [23]. Heaton MJ, Berrett C, Pugh S, Evans A, Sloan C. Modeling bronchiolitis incidence proportions in the presence of spatio-temporal uncertainty. *Journal of the American Statistical Association* ; Submitted.
- [24]. Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli* 2001; :223–242.
- [25]. Jones GL, Haran M, Caffo BS, Neath R. Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association* 2006; 101(476):1537–1547.
- [26]. Shek LPC, Lee BW. Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric respiratory reviews* 2003; 4(2):105–111. [PubMed: 12758047]
- [27]. Chan P, Chew F, Tan T, Chua K, Hooi P. Seasonal variation in respiratory syncytial virus chest infection in the tropics. *Pediatric pulmonology* 2002; 34(1):47–51. [PubMed: 12112797]
- [28]. Müller P, Quintana FA, Jara A, Hanson T. Bayesian nonparametric data analysis. Springer, 2015.

Figure 1. Example epidemic curve for a single region on the logit scale. For brevity, the region subscript, r , of each parameter is dropped.

Figure 2.
Centroids for the R spatial regions.

Figure 3.

Posterior means of (a) $t_0 + t_1$ (time to peak), (b) t_2 (time at peak) in months, (c) $t_1 + t_2 + t_3$ (total time rates were higher than baseline) in months, and (d) ρ at the peak values.

Figure 4.

Epidemic curve for the region containing (a) Orlando, FL, (b) Salt Lake City, UT, (c) Minneapolis, MN, (d) Salem, OR (e) Alaska and (f) July 2013.

Table 1.

Posterior medians and 95% credible intervals for $\exp\{\beta\}$. These numbers can be interpreted as the multiplicative increase in ψ given a unit increase in the covariate.

	2.5%	50%	97.5%
Temperature	1.0030	1.0092	1.0155
Humidity	0.9901	0.9965	1.0026
Interaction	0.9986	0.9992	0.9998

Table 2.

Posterior medians and 95% credible intervals for the total area under the epidemic curve for select regions containing the named cities.

	2.5%	50%	97.5%
Orlando	0.18	0.20	0.22
SLC	0.13	0.16	0.19
Minneapolis	0.11	0.13	0.16
Salem	0.10	0.13	0.16