Practice of Epidemiology

Differential Geographical Risk of Initial *Pseudomonas aeruginosa* Acquisition in Young US Children With Cystic Fibrosis

Kevin J. Psoter*, Margaret Rosenfeld, Anneclaire J. De Roos, Jonathan D. Mayer, and Jon Wakefield

* Correspondence to Dr. Kevin J. Psoter, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205 (e-mail: kpsoter1@jhu.edu).

Initially submitted August 2, 2013; accepted for publication March 11, 2014.

***Pseudomonas aeruginosa*** is the sentinel respiratory pathogen in cystic fibrosis patients. We conducted a retrospective study to examine whether state of residence affected risk of *P. aeruginosa* acquisition among US children under 6 years of age with cystic fibrosis by using data from the Cystic Fibrosis Foundation National Patient Registry, 2003–2009. The outcome was time to first isolation of *P. aeruginosa* from a respiratory culture. We used a Bayesian hierarchical Weibull regression model with interval-censored outcomes. Spatial random effects, included at the state level and modeled using an intrinsic conditional autoregressive prior, allowed estimation of the residual spatial correlation. The regression portion of the model was adjusted for demographic and disease characteristics potentially affecting *P. aeruginosa* acquisition. A total of 3,608 children met the inclusion criteria and were followed for an average of 2.1 (standard deviation, 1.6) years. *P. aeruginosa* was cultured in 1,714 (48%) subjects. There was a moderately elevated spatial residual relative risk. An estimated 95% credible interval for the residual hazard ratio under 1 of the fitted models was 0.64–1.57; the strongest positive association was observed in the Southern states. The fact that risk for *P. aeruginosa* acquisition displayed spatial dependence suggests that regional factors, such as climate, may play an important role in *P. aeruginosa* acquisition.

Bayesian hierarchical spatial models; cystic fibrosis; interval censoring; intrinsic conditional autoregression; *Pseudomonas aeruginosa*; Weibull regression

Abbreviations: CF, cystic fibrosis; GLMM, generalized linear mixed model; IG, inverse gamma; INLA, integrated nested Laplace approximation.

Cystic fibrosis (CF) is the most common life-shortening autosomal recessive disease in Caucasians and affects those of all races and ethnicities. CF is characterized by chronic lower airway bacterial infection and progressive lung function decline, which are associated with high morbidity and premature death. *Pseudomonas aeruginosa*, a ubiquitous gram-negative bacterium, is the most important respiratory pathogen in CF patients. Prevalence of *P. aeruginosa* in respiratory cultures increases with age from approximately 30% at ages 0–5 years to 80% at ages 18 years and older (1). Initial clinical *P. aeruginosa* isolates are present at low density (2, 3) and generally exhibit phenotypical characteristics of environmental isolates in that they are nonmucoid (4) and highly susceptible to antibiotics (4). Initially, *P. aeruginosa* infection is typically intermittent and asymptomatic; however, over time, *P. aeruginosa* adapts to the unique milieu of the CF lower airway. Infection eventually becomes chronic, characterized by biofilm formation, with mucoid (5, 6) and antibiotic-resistant (7–10) strains. Earlier *P. aeruginosa* acquisition is associated with worse clinical outcomes (7); consequently, early identification and aggressive eradication strategies are recommended to delay or prevent chronic infection (11–13).

The fact that the genotypes of initial *P. aeruginosa* isolates are similar to those of environmental isolates (4, 14, 15), coupled with the ubiquitous nature of *P. aeruginosa* in the environment, suggests that the majority of initial infections in CF patients are environmentally acquired. Although patient-level
risk factors for initial P. aeruginosa infection have been evaluated (8, 16–21), few studies have investigated potential environmental risk factors (21, 22). Recently, warmer annual ambient temperature was shown to be associated with increased P. aeruginosa prevalence and earlier age at P. aeruginosa acquisition (22), suggesting that risk of P. aeruginosa acquisition could potentially be geographically dependent. Further, initial P. aeruginosa acquisition has been shown to vary by season, with differential seasonal patterns observed for climate zones (23).

If macroenvironmental factors are contributing to the risk of P. aeruginosa acquisition in patients with CF, then differential acquisition should be expected within a large geographical area in which variation in these factors is present. In this case, for patients residing in close proximity to each other, we would expect some degree of residual spatial correlation in the risk of P. aeruginosa acquisition after adjustment for patient-level characteristics known to be associated with P. aeruginosa acquisition. To date, no study has investigated spatial patterns of initial P. aeruginosa acquisition in any CF patient population. The purpose of this investigation was to evaluate the residual spatial correlation of initial P. aeruginosa acquisition in young children with CF after adjustment for known risk factors for P. aeruginosa acquisition.

METHODS

Study population and design

We conducted a retrospective study to describe the geographical distribution of and evaluate potential residual spatial dependence for initial P. aeruginosa acquisition in young children with CF using US Cystic Fibrosis National Patient Registry (herein, “registry”) data from 2003 to 2009. The registry is a national database containing a wide range of demographic and encounter-based clinical data from approximately 80% of all CF patients in the United States. The study population consisted of all CF patients born after December 31, 2002, with a recorded household zip code in the lower contiguous 48 states. Thus, participants were 6 years of age or younger at study completion. Patients were excluded if they did not have a respiratory culture recorded prior to 2 years of age. To evaluate incident P. aeruginosa cases, we also excluded patients in whom P. aeruginosa was isolated from the first recorded culture. Using zip code data from the registry, we identified state of residence using ArcGIS, version 10.1, software (Esri, Inc., Redlands, California). Individuals were assigned to the states in which they resided in either the year in which the first P. aeruginosa–positive cultures were obtained or the last clinical visits were recorded for those children who remained P. aeruginosa free during observation.

The primary outcome of interest was time to initial P. aeruginosa acquisition, defined as the first P. aeruginosa–positive culture recorded in the registry. Respiratory cultures are generally obtained quarterly (i.e., 4 times per year) (11) in accordance with Cystic Fibrosis Foundation clinical care guidelines. In these young children, the source of respiratory samples is generally an oropharyngeal swab. This study was approved by the institutional review board of the University of Washington (Seattle, Washington) and the Cystic Fibrosis Foundation Registry Committee (Bethesda, Maryland).

Statistical analysis

Descriptive statistics were produced and compared between children who acquired P. aeruginosa during follow-up and those who remained P. aeruginosa free. A χ² test was used to test the differences in proportions by P. aeruginosa acquisition status for categorical variables, and Student’s t tests with unequal variances were used for continuous variables. To describe the geographical distribution of the study population, we mapped the total number of children in each state and the national distribution of participants by P. aeruginosa acquisition status. To further describe P. aeruginosa acquisition, we also plotted the proportion of patients acquiring P. aeruginosa and the median time to the first positive P. aeruginosa culture at the state level.

The exact date of P. aeruginosa acquisition was unknown for each individual; rather, acquisition was known only to occur within an interval time period between the date of the first P. aeruginosa–positive culture and the date of the last prior P. aeruginosa–negative culture. Thus, we used Weibull regression with interval-censored outcomes to evaluate time to P. aeruginosa acquisition. Study participants entered the risk set upon their first recorded encounters in the registry and were right censored if they remained P. aeruginosa free at the last encounter recorded prior to January 1, 2010. The Weibull (2-parameter) density function can be expressed as

\[ f(t; \alpha, \lambda) = \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha} \]

where \( t \) is time to event, \( \alpha \) and \( \lambda \) are the shape and scale parameters, respectively. Considering interval-censored outcomes, an individual’s outcome, \( T_i \), can be expressed as a triple \((T_{i,lo}, T_{i,hi}, \delta_i)\), whereby \( T_{i,lo} \) represents the time for the interval’s left-hand endpoint, which, for all children in the present study, represents the last recorded negative P. aeruginosa culture; for those remaining P. aeruginosa free, \( T_{i,lo} \) represents the time of right censoring. Similarly, \( T_{i,hi} \) represents the time corresponding to the right-hand endpoint for the acquisition interval; for individuals, acquiring P. aeruginosa represents the date of the first positive culture, and for individuals remaining P. aeruginosa free, this represents an unobserved time. \( \delta_i \) is a censoring indicator and is defined by P. aeruginosa acquisition status.

All statistical models were adjusted for a priori factors potentially associated with P. aeruginosa acquisition, including sex, race (white vs. nonwhite), ethnicity (Hispanic vs. non-Hispanic), insurance status (any private insurance vs. no private insurance), CF transmembrane conductance regulator functional class (minimal, residual, or unclassified) (18, 24, 25), age at diagnosis (in months), diagnosis by newborn screening (yes/no), and culture frequency (defined as the number of cultures per number of days to P. aeruginosa acquisition or censoring). CF transmembrane conductance regulator functional class was defined as follows: minimal function (“severe”) represents both mutations in Class I, II, or III (includes the most common CF transmembrane conductance regulator mutation, ΔF508, a Class II mutation); and residual function (“mild”) represents I or both mutations in Class IV or V.
Spatial modeling

Previous applications of spatial regression models to time-to-event analyses with interval-censored outcomes are limited (26, 27). If an outcome is spatially dependent, then ignoring this can usually lead to an underestimation of standard errors, resulting in overly narrow interval estimates and subsequent, incorrect inference (28). This (potential) spatial dependence, therefore, violates the independence assumption for observations required for generalized linear models. The more flexible class of generalized linear mixed models (GLMMs) can accommodate such dependence. Whereas frequentist methods are common in many GLMM applications, Bayesian methodology, using Markov chain Monte Carlo methods, has become the predominant choice for spatial applications. Recently, integrated nested Laplace approximation (INLA) (29) has emerged as a computationally efficient alternative to Markov chain Monte Carlo methods for performing Bayesian analysis when distributions, including the Weibull (26), can be expressed as a latent Gaussian model.

To evaluate the potential spatial dependence of time to initial *P. aeruginosa* acquisition, we used INLA for the previously described Weibull regression model with interval-censored outcomes. Our approach was to evaluate *P. aeruginosa* acquisition using the state as the areal unit of analysis. A finer level of analysis was not possible because of the sparseness of data at, for example, the county level. The so-called convolution random effects model (30) includes 2 random effects terms, an unstructured, independent random effect ($V_j$) assigned for each state $j$, and a spatially structured random effect ($U_j$). We model the latter as an intrinsic conditional autoregression for each state. The intrinsic conditional autoregression model borrows information from “neighboring” areal units (defined in our study as states sharing a common boundary). Thus, $V_j$ are independent, whereas $U_j$ depend on the random effects $U$ of neighboring areas. The spatial Weibull hazard model can be expressed as $h(t; y; \alpha) = \alpha t^{a-1} \exp(\eta_y)$, where the linear predictor is of the form $\eta_y = Z_{ij}^T \beta + V_j + U_{ij}$, and $\eta_y$ denotes the ith individual residing in state $j$. Additionally, $Z_{ij}^T$ is the design matrix; $V_j \sim N(0, \sigma^2_{Vj})$, $U_j$ is a Gaussian Markov random field where $U_j/k U_{k}, k \in \delta_j \sim N(U_{ij}, \sigma^2_u/m_j)$, with $U_{ij}$ being the mean of the spatial random effects of the neighbors, and $m_j$ being the number of neighbors of state $j$. The variances $\sigma^2_V$ and $\sigma^2_U$ represent unknown hyperparameters, with priors taken from inverse gamma (IG) distributions ($a$, $b$). Spatial and independent variances are on different scales and are not directly comparable, $\sigma^2_U$ is a conditional variance based on $U_k, k \in \delta_j$, and $\sigma^2_V$ is a marginal variance. The proportion of variance explained by the structured spatial component was estimated by the empirical posterior marginal variance.

For the independent and spatially structured variances, we took IG(1, 0.026) priors. These priors are such that a 95% interval for the residual hazard ratio is 0.5–2, and they are based on the residual hazard ratios following a log Student $t$ distribution with 2 degrees of freedom (31). The model was completed by assigning improper flat priors for $\beta$.

The formulation of the spatial model allows us to simultaneously 1) estimate the residual spatial dependence, and 2) investigate the impact of spatial correlation on regression coefficients for the predictors of time to *P. aeruginosa* acquisition. Results for regression coefficients are presented as hazard ratios and the associated 95% confidence intervals for models with no random effects and as 95% credible intervals for Bayesian random effects models. The random effect terms can be interpreted as the effect of state of residence on time to *P. aeruginosa* acquisition for each subject. At the state level, quartiles for the nonspatial and spatial random residual relative risks were plotted on a map for each state.

We evaluated the sensitivity to the prior specifications by using different IG distributions for the random effect variances. Five priors were chosen for the variances of the random effects terms. These were chosen to represent the 95% posterior probability as outlined by Wakefield (32) and included IG(0.5, 0.0164), IG(0.5, 0.006), and IG(0.5, 0.0014), exponential distributions with 1 degree of freedom, and 95% ranges for the residual hazard ratios of (0.1, 10), (0.25, 4), and (0.5, 2), respectively. Additionally, the IG priors IG(1.0, 0.2864) and IG(1.0, 0.104) were used, which correspond to alternative Student $t$ distributions with 2 degrees of freedom and 95% ranges of (0.1, 10) and (0.25, 4), respectively. The initial Weibull model fits were verified by visual inspection of the log(−log($\hat{S}(t)$)) versus log($t$) plots. All $P$ values are 2-sided, and a $P$ value of less than 0.05 was considered statistically significant. All analyses were performed using the R, version 2.15.2, statistical environment (33).

Table 1. Distribution of Demographic and Disease Characteristics Among Young Children With Cystic Fibrosis Who Acquired or Remained *Pseudomonas aeruginosa* Free, United States, 2003–2009

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Acquired <em>P. aeruginosa</em> During Follow-Up, % ($n=1,714$)</th>
<th>Remained <em>P. aeruginosa</em> Negative, % ($n=1,909$)</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>48</td>
<td>52</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>White</td>
<td>92</td>
<td>92</td>
<td>0.61</td>
</tr>
<tr>
<td>Hispanic</td>
<td>11</td>
<td>10</td>
<td>0.61</td>
</tr>
<tr>
<td>Any private insurance</td>
<td>53</td>
<td>54</td>
<td>0.26</td>
</tr>
<tr>
<td>AF508 mutation category</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Homozygous</td>
<td>53</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Heterozygous</td>
<td>36</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>CFTR mutation class</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Minimal</td>
<td>73</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>6</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>22</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Age at CF diagnosis, months</td>
<td></td>
<td></td>
<td>2.3a</td>
</tr>
<tr>
<td>Diagnosis by newborn screening</td>
<td>35</td>
<td>49</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Abbreviations: CF, cystic fibrosis; CFTR, cystic fibrosis transmembrane conductance regulator.

* Value expressed as mean.
RESULTS

A total of 3,608 children were included in the final analysis, of which 48% (n = 1,714) acquired *P. aeruginosa* during a mean observational period of 2.1 (standard deviation, 1.6) years. The mean number of recorded respiratory cultures per participant was 8.3 (standard deviation, 6.1). The median time to *P. aeruginosa* acquisition was 470 days (75th percentile–25th percentile, 195–634 days). Demographic characteristics of the study population by *P. aeruginosa* acquisition status are presented in Table 1. Patients remaining *P. aeruginosa* free were more likely to be male and to have been diagnosed by newborn screening, whereas those acquiring *P. aeruginosa* were more likely to receive public health interventions.
insurance, be ΔF508 homozygous, and have CF transmembrane conductance regulator mutations with minimal function.


The nationwide distribution of the study population is presented by state in Figure 1A and by *P. aeruginosa* acquisition status in Figure 1B. The proportion of patients acquiring
**Table 2.** Results of Spatial and Nonspatial Weibull Regression Models Evaluating Time to *Pseudomonas aeruginosa* Acquisition in Young Children With Cystic Fibrosis, United States, 2003–2009

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Univariate Model</th>
<th>Multivariate Nonspatial Model</th>
<th>Multivariate Model With Area-Level Effect Only</th>
<th>Multivariate Spatial Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% Confidence Interval</td>
<td>HR</td>
<td>95% Confidence Interval</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00</td>
<td>Referent</td>
<td>1.00</td>
<td>Referent</td>
</tr>
<tr>
<td>Female</td>
<td>1.09</td>
<td>0.97, 1.22</td>
<td>1.09</td>
<td>0.98, 1.21</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td>Referent</td>
<td>1.00</td>
<td>Referent</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>0.98</td>
<td>0.80, 1.17</td>
<td>1.00</td>
<td>0.76, 1.23</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>1.00</td>
<td>Referent</td>
<td>1.00</td>
<td>Referent</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.12</td>
<td>0.92, 1.31</td>
<td>1.15</td>
<td>0.95, 1.35</td>
</tr>
<tr>
<td>Insurance status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No private insurance</td>
<td>1.00</td>
<td>Referent</td>
<td>1.00</td>
<td>Referent</td>
</tr>
<tr>
<td>Any private insurance</td>
<td>0.85</td>
<td>0.73, 0.97</td>
<td>0.86</td>
<td>0.74, 0.98</td>
</tr>
<tr>
<td>CFTR mutation class</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal</td>
<td>1.00</td>
<td>Referent</td>
<td>1.00</td>
<td>Referent</td>
</tr>
<tr>
<td>Residual</td>
<td>0.48</td>
<td>0.22, 0.73</td>
<td>0.49</td>
<td>0.23, 0.74</td>
</tr>
<tr>
<td>Other</td>
<td>0.78</td>
<td>0.63, 0.93</td>
<td>0.75</td>
<td>0.59, 0.91</td>
</tr>
<tr>
<td>Age at diagnosis (years)</td>
<td>0.66</td>
<td>0.18, 2.15</td>
<td>0.88</td>
<td>0.24, 3.10</td>
</tr>
<tr>
<td>Diagnosis by newborn screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00</td>
<td>Referent</td>
<td>1.00</td>
<td>Referent</td>
</tr>
<tr>
<td>Yes</td>
<td>0.95</td>
<td>0.94, 0.96</td>
<td>0.98</td>
<td>0.97, 1.00</td>
</tr>
<tr>
<td>Standard deviation of random effect(s)**</td>
<td></td>
<td></td>
<td>0.23</td>
<td>0.16, 0.32</td>
</tr>
<tr>
<td>Independent (σ_v)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial (σ_u)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** CFTR, cystic fibrosis transmembrane receptor; HR, hazard ratio; IG, inverse gamma.

* Credible interval (2.5% and 97.5%) based on IG(1.0, 0.026) prior(s).

* Reflects the median of the standard deviation.

*P. aeruginosa* over the study period varied by state and ranged from 21% to 71% (Figure 1C). The distribution by state of median age at first *P. aeruginosa*–positive culture for children acquiring *P. aeruginosa* is presented in Figure 1D by quartile of age (in days). Generally, delayed acquisition was observed in the Western states compared with the rest of the country.

Results from unadjusted and adjusted Weibull survival models for time to *P. aeruginosa* acquisition, including multivariate regression models with and without spatial random effects, are presented in Table 2. When comparing the adjusted Weibull regression models, it can be seen that including independent random effects for states did not materially change the observed associations for other potential risk factors. For example, the hazard ratios for the effect of sex were comparable for all models, indicating minimal confounding by location for these variables. Although newborn screening for CF was governed by state-by-state policy during this time period, associations with time to *P. aeruginosa* acquisition varied little between models. The interpretation of covariate coefficients in the models without random effects takes the usual form (i.e., the change in the hazard ratio associated with a unit change in the covariate of interest, with all other covariates held constant), whereas for the random effects models, we add to this “for individuals in the same state (or in states with the same random effects).”

In the multivariate regression models including only an independent random effect term for each state, the posterior estimated standard deviation of the residual hazard ratio was 1.26 (95% credible interval: 1.17, 1.38), indicating there is excess variability. An estimate of the 95% credible interval for the residual hazard ratio is, therefore, $\exp(\pm 1.96 \times 0.23) = (0.64, 1.57)$. Here, the “residual” refers to excess risk associated with the independent random effects after accounting for other covariates. Similarly, the inclusion of independent and spatial random effects indicated that there was a moderate, statistically significant independent random component associated with
time to *P. aeruginosa* acquisition (hazard ratio = 1.20, 95% credible interval: 1.12, 1.20). For this model, inclusion of the structured spatial random effect had minimal effect on the residual risk associated with the independent random effect term; however, inclusion of the spatial random effect resulted in an estimate of the excess spatial variability that was similar in magnitude to the independent component. The estimated proportion of the total residual variability of the log hazard ratio explained by the spatial component was approximately 45%, based on the empirical variance of $U_j$.

Figures 2A and 2B depict the mapping of the residual relative risk of the spatial and independent random effect terms.
As is evident, the spatial random effect terms display a high level of spatial structure and smoothness, with the highest spatial dependence observed in the Southern states and the lowest levels observed in the Western and Northeastern states. As expected, the independent random effects terms do not display any clear geographical pattern.

The interpretation of results did not materially change when different priors were applied (Table 3). The variance attributable to the independent random effect (states) was decreased when the spatial random effects were included in the models, as expected. The variances of the spatial random effects terms were slightly higher when an IG prior based on the posterior residual hazard ratios following an exponential distribution was placed on the random effects compared with when an IG based on a log Student t distribution was used.

**DISCUSSION**

The identification of factors associated with initial *P. aeruginosa* acquisition is of great importance to the clinical management of CF patients. We have demonstrated that time to initial *P. aeruginosa* acquisition in young children with CF in the United States exhibits moderate residual spatial correlation, with the strongest positive associations observed in the southern United States. These results suggest that macroenvironmental factors, such as temperature, humidity, and air pollution may be contributing to the risk of *P. aeruginosa* acquisition, particularly in the Southern states.

Strengths of this investigation included a large, national study population and individual-level covariate data to adjust for previously reported factors associated with incident *P. aeruginosa* acquisition. Further, in contrast to previous investigations of *P. aeruginosa* acquisition, our study accounted for the interval-censored nature of the outcome. We are unaware of any other studies that have used spatial methodology for investigating *P. aeruginosa* acquisition; therefore, comparison with other studies is limited.

Effect estimates of patient-level covariates in the nonspatial Weibull regression models were similar to those obtained in a recent investigation by Rosenfeld et al. (21) reporting on a corresponding registry study population and subjects enrolled in a large US observational study (34, 35). Importantly, in the present investigation, the estimated effect sizes for the covariates were not significantly altered after inclusion of the random spatial and nonspatial components, indicating that confounding by location was most likely not present, presumably because the covariates did not have a strong spatial pattern. In the present investigation, the mean age at initial *P. aeruginosa* acquisition was 1.2 years, an earlier age than previously reported (19–22), likely because of our eligibility criteria, which included having a first clinical encounter and initial negative *P. aeruginosa* culture prior to 2 years of age, as well as limiting our cohort to children 6 years of age or younger.

The complex interplay between genetic and environmental factors in explaining the heterogeneity of CF lung disease is an area of active investigation. In the CF Twins and Siblings Study, Collaco et al. (36) demonstrated that approximately half of the lung function variation in this cohort was explained by environmental/stochastic factors as opposed to genetic factors. In this same cohort, genetic factors contributed little to initial *P. aeruginosa* acquisition (37). Thus, the identification of specific environmental factors contributing to this variation is paramount to understanding the natural history of the disease.

Spatial epidemiology provides a powerful set of statistical tools for understanding the etiology of infectious diseases. The roles of environmental factors in infectious disease outbreaks, spread, and acquisition have been well described (38). For example, climatic conditions are important factors in the distribution of vectors and may also increase host susceptibility to infection. In the context of *P. aeruginosa* acquisition in CF patients, initial acquisition is generally considered to be from the environment rather than transmission between patients, presenting a somewhat unusual infectious disease paradigm. Identification of geographical variability in *P. aeruginosa* acquisition in this study suggests that there may be factors such as climate associated with environmental *P. aeruginosa* proliferation and/or affecting individual susceptibility or exposure to *P. aeruginosa*. It may be important
in future work to consider applications to predict *P. aeruginosa* acquisition in both space and time.

Our analysis used a Bayesian GLMM. The GLMMs are a rich family of models characterized by their inclusion of normal random effect term(s), the distinguishing feature from the more traditional generalized linear models. The Bayesian framework is flexible, and in our applications included spatially structured random effects. A limitation of implementing Bayesian GLMMs, often performed with WinBUGS software (39), is the computing time required. INLA provides a relatively fast computation method to accommodate such models and accommodates a wide variety of other models including generalized additive models. We found that interval-censored data, a common occurrence in other epidemiologic settings, can be easily implemented using the INLA package in R using a Weibull distribution. Additionally, INLA allows implementation of Cox-type GLMM survival models; however, this requires specifying the baseline hazard to be piecewise constant (i.e., modeling the baseline hazard with knots). A sample of the R code used for conducting the present analysis is provided in the accompanying Appendix; in all models, computation took less than 3 seconds to complete. R codes to implement the methods presented here, along with a simulated data set, are available from J.W.’s website (40).

There are several limitations to our study. First, approximately 90% of respiratory cultures were from oropharyngeal swabs. The sensitivity of oropharyngeal cultures for detecting *P. aeruginosa* in the lower respiratory tract is low (44%); therefore, results reported herein may more accurately reflect upper airway and not lower airway colonization. Second, the spatial analysis was conducted at the state level because our sample size did not allow analysis at a finer geographical level. An assumption of the modeling approach was a common state-level effect, so within-state differences could have been masked. Third, individual zip code was taken as the zip code recorded in the year that a child acquired *P. aeruginosa* or the year in which the child was censored. During follow-up, approximately 12% (*n* = 421) of the patient population reported a change in zip code; however, only 2.2% (*n* = 81) reported a change in state of residence. Accordingly, because of the minimal number of individuals who reported a change in state residence, potential disease latency, in relation to spatial location, was not explored. Fourth, the Cystic Fibrosis Foundation Registry includes approximately 80% of the population with CF in the United States. In general, those patients who are not recorded in the registry tend to be older, because infants and young children are more likely to visit Cystic Fibrosis Foundation care centers. Additionally, there is no indication that there is systematic underreporting of CF cases at the state level. Therefore, we believe that the results of this investigation reflect the clinical population of CF patients who would be treated. Finally, spatial correlation of health outcomes is often related to underlying differences in processes of care (affecting the endpoint) or unmeasured confounding variables such as socioeconomic status. We did adjust for Cystic Fibrosis Foundation care center and culture frequency; however, residual confounding could still remain after adjustment for insurance status.

In conclusion, we demonstrated that there are significant geographical differences in time to *P. aeruginosa* acquisition in young children with CF. Results of this spatial analysis provide evidence to justify the conduct of future studies to identify potential climatic and environmental risk factors for *P. aeruginosa* acquisition.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Kevin J. Psoter); Department of Pediatrics, School of Medicine, University of Washington, Seattle, Washington (Margaret Rosenfeld); Division of Pulmonary Medicine, Seattle Children’s Hospital, Seattle, Washington (Margaret Rosenfeld); Department of Environmental and Occupational Health, School of Public Health, Drexel University, Philadelphia, Pennsylvania (Anneckaire J. De Roos); Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington (Jonathan D. Mayer); Department of Global Health, School of Public Health, University of Washington, Seattle, Washington (Jonathan D. Mayer); Department of Health Services, School of Public Health, University of Washington, Seattle, Washington (Jonathan D. Mayer); Department of Family Medicine, School of Medicine, University of Washington, Seattle, Washington (Jonathan D. Mayer); Division of Allergy and Infectious Diseases, Department of Medicine, School of Medicine, University of Washington, Seattle, Washington (Jonathan D. Mayer); Department of Biostatistics, School of Public Health, University of Washington, Seattle, Washington (Jon Wakefield); and Department of Statistics, University of Washington, Seattle, Washington (Jon Wakefield).

We thank Dr. Bruce Marshall and the Cystic Fibrosis Foundation for providing the Cystic Fibrosis Foundation Registry data.

Conflict of interest: Dr. Rosenfeld has served as a consultant for Vertex Pharmaceuticals Corp. and for the North American Scientific Advisory Group of the Epidemiologic Study of Cystic Fibrosis sponsored by Genentech, Inc. Dr. Rosenfeld has received institutional funding from the Cystic Fibrosis Foundation.

REFERENCES


APPENDIX

Sample R Code for Hierarchical Multivariable Spatial Weibull Regression with Interval-Censored Outcomes

```r
formula=inla.surv(LeftEndpointTime, Censor, RightEndpointTime)~Sex + PrivateInsurance + DiagNeonatal + Race +
as.factor(CFTR) + Ethnicity + AgeDx + CultureFreq + f(State, model="iid", param=c(1,0.026)) + f(State2, model="besag", graph="US_adjacency", param=c(1,0.026))
spatialmodel=inla(formula, family="weibull", data=a)
summary(spatialmodel)
```