The Journal of Infectious Diseases
MAJOR ARTICLE

Association Between Sporadic Legionellosis and River Systems in Connecticut

Kelsie Cassell,[1] Paul Gacek,[4] Joshua L. Warren,[2] Peter A. Raymond,[3] Matthew Cartter,[4] and Daniel M. Weinberger[1]

Departments of [1]Epidemiology of Microbial Diseases and [2]Biostatistics, Yale School of Public Health, and [3]Yale School of Forestry and Environmental Studies, New Haven, Connecticut, and [4]Connecticut Department of Public Health, Hartford

(See the editorial commentary by de Fisman, on pages 188–97.)

Background. There has been a dramatic increase in the incidence of sporadic legionnaires' disease in Connecticut since 1999, but the exact reasons for this are unknown. Therefore, there is a growing need to understand the drivers of legionnaires' disease in the community. In this study, we explored the relationship between the natural environment and the spatial and temporal distribution of legionellosis cases in Connecticut.

Methods. We used spatial models and time series methods to evaluate factors associated with the increase and clustering of legionellosis in Connecticut. Stream flow, proximity to rivers, and residence in regional watersheds were explored as novel predictors of disease, while controlling for testing intensity and correlates of urbanization.

Results. In Connecticut, legionellosis incidence exhibited a strong pattern of spatial clustering. Proximity to several rivers and residence in the corresponding watersheds were associated with increased incidence of the disease. Elevated rainfall and stream flow rate were associated with increases in incidence 2 weeks later.

Conclusions. We identified a novel relationship between the natural aquatic environment and the spatial distribution of sporadic cases of legionellosis. These results suggest that natural environmental reservoirs may have a greater influence on the spatial distribution of sporadic legionellosis cases than previously thought.

Keywords. Legionnaires' Disease; *Legionella*; hydrology; spatial analysis; time series; watershed; rivers; drinking water.

Since the first confirmed outbreak in 1976, diagnoses of legionnaires' disease, or legionellosis, have increased steadily in the United States, often garnering intense media attention during outbreaks [1]. Outbreaks tend to be associated with contaminated cooling towers and potable water sources [2, 3], but the majority of legionellosis cases are not known to be associated with a common exposure. These "sporadic" cases constitute between 3% and 7% of the total burden of community-acquired pneumonia [4, 5] and are particularly important causes of such pneumonia among immunocompromised individuals [6, 7].

The complex ecology and infection process of *Legionella* complicates the assessment of environmental risk factors and the identification of exposure sources [8]. To infect humans and cause disease, the bacteria must be aerosolized from an aquatic source and then inhaled. *Legionella* are pervasive in the environment, growing in rivers, lakes, ponds, and estuaries and along ocean coasts. The bacteria grow at temperatures between 25°C and 42°C and can form biofilms in both natural and built aquatic environments [9–12]. Meteorological factors, such as humidity and air temperature, have been linked to increases in sporadic legionellosis and seasonality of the disease [13–15]. In addition, the flooding and accelerated stream flow associated with large precipitation events can potentially disrupt both the structure and composition of river biofilms, dispersing the bacteria and negatively affecting water quality [16, 17].

Despite the ubiquity of *Legionella* in the environment, regional heterogeneity has been noted in case detection across the United States [9, 18–20] with New England, the Mid-Atlantic and East North Central areas reporting higher incidence [21]. During the past decade, Connecticut has experienced an unexplained increase of community-acquired (sporadic) legionellosis [22]. Given this unexplained increase in cases, it is necessary to identify potential drivers of sporadic cases. In the current study, we used spatial and time series regression models to test specific hypotheses about the relationship between the distribution of cases of legionellosis in Connecticut and proximity to rivers, stream flow, meteorological conditions, and sources of drinking water.

METHODS

Case Identification

In Connecticut, legionellosis is both physician and laboratory reportable to the Connecticut Department of Public Health (DPH) and to the patient's local health department. Laboratory confirmation of the infection is made through culture, urine

Received 25 May 2017; editorial decision 19 August 2017; accepted 29 September 2017; published online December 2, 2017.

Correspondence: K. Cassell, MPH, Department of Epidemiology of Microbial Diseases, Yale School of Public Health, 60 College St, New Haven, CT 06511 (kelsie.cassell@yale.edu).

The Journal of Infectious Diseases® **2018;217:179–87**

DOI: 10.1093/infdis/jix531

antigen testing (detects *Legionella pneumophila* serogroup 1 only), or paired serology with a ≥4-fold titer rise [23, 24]. Confirmed "legionellosis" refers to both legionnaires' disease and the milder form of the disease, Pontiac fever. To identify possible common sources of exposure, DPH staff conducted follow-up on all confirmed legionellosis cases by contacting the healthcare provider of record and conducting patient interviews.

Our analyses focused on patients hospitalized during the weeks between 1 January 1999 and 31 December 2015. If cases were not linked to an outbreak during follow-up investigation by public health officials, they were classified as "sporadic" and included in our analyses. All patients were included in the analysis, unless they had an overnight stay at a location with a confirmed outbreak during their exposure period. Along with week of hospitalization, age category, ZIP code of residence, and the drinking water supplier and sources were also available. Incidence rates were calculated as cases per 100 000 persons (using 2000 census data) during the 17-year study period. The study was approved by the human investigation committees at Yale University and the Connecticut DPH. Certain data used in this publication were obtained from the Connecticut DPH, and the authors assume full responsibility for analyses and interpretation of these data.

Hydrological Data

The hydrological variables included data from both the natural environment (major rivers and regional watersheds) and built environment (public drinking water supplier and sources) in Connecticut. ZIP codes were assigned to one of Connecticut's 42 regional watersheds by mapping which watershed the majority of a ZIP code's area fell (Supplementary Figure S1). The 31 rivers used in the spatial analysis included both large and medium sized rivers. Distance from ZIP code *i* to river *j* was calculated as the shortest euclidean distance from the centroid of ZIP code *i* to river *j*.

Weekly mean of stream flow for rivers (measured in cubic feet per second) were obtained from daily data from the United States Geological Survey. Stream flow rates were measured at 38 stations along all 34 streams for which data were consistently collected from 1 January 1999 to 31 December 2015. For the time series analysis, data for individual streams were aggregated at the weekly level and then standardized (by subtracting the mean and dividing by the standard deviation). The standardized values for each river were then averaged across all streams to provide an mean weekly stream flow value for the entire state. As part of a secondary analysis, we examined aquatic life use support assessment of the rivers as a proxy for a river's ecological health (see Supplementary Text for more information).

Meteorological/Climatic Data

Weekly mean of precipitation (in millimeters), relative humidity (as percentages) and temperature (in degrees Celsius) were collected by the National Oceanic and Atmospheric Administration and obtained from the National Climatic Data Center. Temperature and precipitation were recorded daily at 18 stations across the state, and relative humidity was recorded hourly at 4 stations.

Demographic and Testing Data

ZIP code population and area were obtained from the US Census Bureau's 2000 Summary File 1 for Connecticut and ZIP code density was calculated as population/area [25]. The median age of housing for each ZIP code in 2000 was also collected from the US Census Bureau [26]. *Legionella* urine antigen testing rates were available by hospital for 2009 only. The proportion of hospitalized pneumonia discharges tested via urine antigen testing was included in all spatial models. Testing proportions were matched to ZIP code through the hospital service area–ZIP code crosswalk provided by the Dartmouth Atlas of Health Care (see Supplementary Text). Hospital service areas were defined as the primary hospital where the greatest proportion of Medicare residents in each ZIP code would be hospitalized (Supplementary Figure S2).

Statistical Methods

Spatial Analysis

Three random effects models were used to estimate the incidence of legionellosis by ZIP code, where the observed number of cases of legionellosis in ZIP code *i*, Y_i, is assumed to be distributed as a Poisson random variable:

$$Y_i \mid \lambda_i \sim \text{Poisson}(\lambda_i), i = 1, \ldots, n,$$

$$\begin{aligned}\ln(\lambda_i) = \ln(\text{offset}_i) + \beta_0 + \beta_1(\ln(\text{density}_i))\\ + \beta_2(\text{housing age category}_{1vs5,i})\\ + \beta_3(\text{housing age category}_{2vs5,i})\\ + \beta_4(\text{housing age category}_{3vs5,i})\\ + \beta_5(\text{housing age category}_{4vs5,i})\\ + \beta_6(\ln(\text{proportion\_tested}_i)) + u_{0i}\end{aligned}$$

This model included the log of density in the ZIP code as a measure of urbanization, a categorical median age of housing as a measure of aging infrastructure/plumbing, and the proportion of pneumonia cases tested for *Legionella* in 2009 by ZIP code [27].

In the first model, the random intercept (u_{0i}) was $u_{0i} \sim N(0, \sigma^2)$, and n is the total number of ZIP codes in the analysis (n = 266). In this model, the ZIP codes are treated independently. The second model accounts for spatial correlation in counts from different ZIP codes such that the random intercepts, u_{0i}, are modeled using a multivariate normal distribution with spatial covariance structure given as

$\text{Cov}\left[u_{0i}, u_{0j}\right] = \sigma^2 \exp\left(-\frac{d_{ij}}{\alpha}\right)$, where $d_{ij} = \| c_i - c_j \|$, c_i are the coordinates of the centroid of ZIP code *i,* and $\|\cdot\|$ represents the euclidean distance function. The predicted incidence rate from this model was calculated as $\lambda_i / offset_i$. This model assumes that ZIP codes separated by shorter distances may be more similar where the range of spatial correlation is controlled by the unknown parameter α. The third random effects model was constructed to estimate incidence rate ratios for regional watersheds. Instead of using a random intercept for ZIP code, an *iid* random intercept for regional watershed *i* was used (similar to model 1).

To evaluate the relationship between cases of legionellosis in ZIP code *i* and distance to specific rivers *j*, a fourth model was developed. Model 4 was a Poisson regression *without* random effects, because this outperformed models with random effects. The log of density, median age of housing, and proportion undergoing urine antigen testing for each ZIP code were included in the model. Variables for proximity to rivers were added to the model in forward fashion and retained in the model if its Akaike information criterion (AIC) value was lowered. Proximity to river variables were dichotomized such that if river *j* was within *d* distance of the centroid of a ZIP code *I,* then proximity equals 1; otherwise, proximity equals 0. Distance *d* was explored at lengths of 5, 10, and 15 km from ZIP code centroids in univariate analysis for each river and the distance for which a river had the lowest AIC value was the proximity used in the final multivariate model.

Time Series Analysis

Potential covariates were assessed in a Poisson regression model that controlled for secular trends and seasonality, such that

$$Y_t \mid \lambda_t \sim \text{Poisson}(\lambda_t), t = 1, \ldots, n$$

$$\ln(\lambda_t) = \beta_0 + \beta_1(\text{week}_t) + \beta_2(\text{week}_t^{\,2}) + \beta_3\left[\sin\left(2\pi \times \frac{\text{week}_t}{52.25}\right)\right] + \beta_4\left[\cos\left(2\pi \times \frac{\text{week}_t}{52.25}\right)\right] + \beta_5(\text{Covar}_t)$$

where λ_t is the expected number of cases in a week of a given year and $Covar_t$ is one of the climatic/stream flow variables assessed. Here n (881) is the total number of weeks in the 17 years studied, excluding the first 3 and last 3 (lead and lag) weeks. Variables tested included weekly mean of precipitation, maximum temperature, humidity and stream flow with lags and leads of up to 3 weeks tested for each. All climatic variables were assessed in univariate analysis and only variables that were not collinear in their original (nonlagged) form (assessed in a Spearman rank test with $P > .05$) were eligible for multivariate analysis. Forward selection of variables was used, and covariates were retained in the multivariate model if they had a P value <.05 and lowered the AIC of the full model. (See Supplementary Text for the R packages and SAS procedures used.)

RESULTS

Epidemiological Characteristics of Legionellosis in Connecticut

Between 1 January 1999 and 31 December 2015, a total of 736 confirmed cases of community-acquired legionellosis cases were reported to the Connecticut DPH. The highest incidence was seen in the ≥65-year age group, with 65.29 cases per 100 000 population, whereas no cases were reported among residents aged ≤17 years during the study period. There were 29 cases for which ZIP code was unknown, and these cases were thus excluded from the spatial analysis. Excluded from the time series analysis were 27 cases for which week of hospitalization was unknown. Incidence rate per year shows an increasing

Table 1. Demographics of Sporadic Legionellosis Cases in Connecticut, 1999–2015

Population	Cases, N (%)[a]	Age-Specific Rate or IR
Total	736	…
Overnight exposure		
Yes	115 (15.63)	…
No	332 (45.11)	…
Data missing	289 (39.27)	…
Age category, 7		Age-specific rate (per 100 000)[b]
0–17	0	…
18–39	60 (8.15)	6.02
40–64	362 (49.18)	33.53
≥65	307 (41.71)	65.29
Data missing	7 (0.95)	…
Year		IR (per 100 000)
1999	18	0.53
2000	18	0.53
2001	14	0.41
2002	19	0.56
2003	28	0.82
2004	24	0.70
2005	35	1.03
2006	58	1.70
2007	45	1.32
2008	51	1.50
2009	55	1.62
2010	55	1.62
2011	82	2.41
2012	55	1.62
2013	64	1.89
2014	58	1.70
2015	57	1.67

[a]Percentages may not sum to 100% owing to rounding [25].

[b]Age-specific rates were calculated using the US Census Bureau's 2000 population assessment of Connecticut.

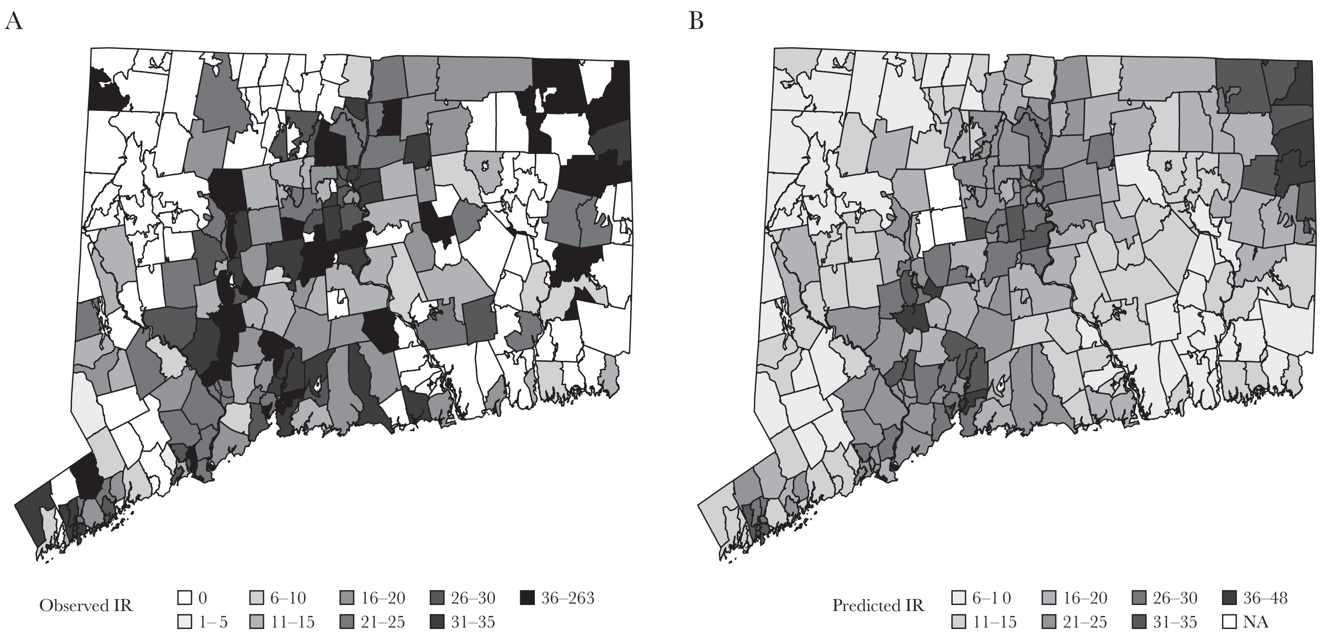

Figure 1. Observed and predicted incidence rates (IRs) (per 100 000) of sporadic legionellosis cases by ZIP code, 1999–2015. *A,* Observed IR of sporadic legionellosis cases by ZIP code, years 1999–2015. *B,* Spatially smoothed predicted IR from model 2, which adjusts for density, median housing age, and proportion tested for *Legionella* in each ZIP code. The spatial smoothing effect is due to the covariance structure, which assumes that ZIP codes near to each other are more similar. NA, not available.

trend over time, peaking at 2.41 in 2011 and stabilizing at about 1.70 cases per 100 000 thereafter (Table 1).

Figure 1 shows the incidence rates by ZIP code, which ranged from 0 (105 ZIP codes never reported a case during the study period) to 263, with notable clustering in urban areas (Figure 1A). The spatial smoothing model confirmed this distinctive spatial pattern of incidence, with high incidence in the northeast and east central regions relative to the statewide adjusted mean (Figure 1B and Supplementary Figure S3) and low incidence in the southeastern part of the state. Incidence was elevated around urban centers (Bridgeport, New Haven, and Hartford) despite controlling for density, median housing age, and testing rates. Three ZIP codes were excluded because they did not have a matching hospital service area to relate to testing practices.

Elevated Incidence Around Key Rivers

Two regional watersheds (Naugatuck and Quinebaug) were associated with an incidence of disease higher than the statewide adjusted mean (Figure 2A and Supplementary Table S1). Six rivers were associated with statistically significant differences in legionellosis incidence in the surrounding ZIP codes in a multivariate model (Table 2). In multivariate models, ZIP codes within 10 km of the Quinebaug River and the Hockanum Brook were associated with 4.37-fold (95% confidence interval, 2.96–6.44) and 1.71-fold (1.21–2.42) increases in legionellosis (Figure 2B). Close proximity to the Naugatuck and Farmington rivers was also associated with statistically significant increases in legionellosis risk. The Saugatuck and Shetucket rivers were associated with decreased rates of legionellosis. This final multivariate model had the lowest AIC score of all the spatial models explored in this study (Supplementary Table S2). Secondary analysis showed no relationship between the health of a river (aquatic life use support category) and a river's inclusion in the model (Supplementary Table S3).

Of the 707 case patients matched to ZIP codes, 566 were matched to a public water supplier (PWS), and the remaining 137 were assumed to obtain their drinking water from a private well (see Supplementary Text for more information). For 41 of the 161 ZIP codes that experienced ≥1 case, multiple drinking water suppliers provided water to affected ZIP code, either through combinations of multiple PWSs or combinations of PWS and private well water. Secondary analysis did reveal a higher proportion of case patients residing within watersheds with an elevated incidence (Naugatuck and Quinebaug) used private well water (31.40%), compared with case patients living in watersheds without a significant association (17.01%; $P < .001$) (Table 3).

Association Between Incidence of Legionellosis and Meteorological Factors and Stream Flow

Finally, we evaluated the association between the weekly incidence of legionellosis and different lags and leads of climatic variables and stream flow. For precipitation, humidity and stream flow, a lag of 2 weeks was most predictive of incidence (Supplementary Figure S4). Precipitation was associated with the greatest risk of sporadic disease, with every 5-mm average increase in rainfall over a week being associated with a 48% increased risk of legionellosis 2 weeks later (Table 4 and

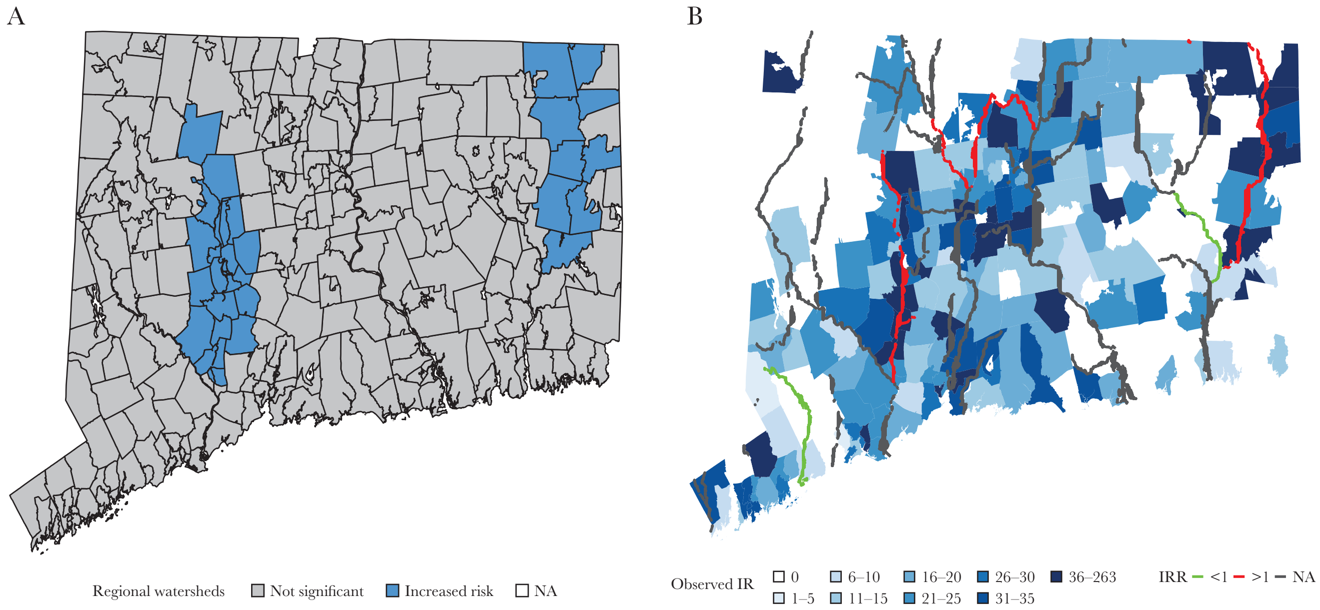

Figure 2. Risk associated with regional hydrology. *A,* Two watersheds (Naugatuck and Quinebaug) were associated with increased disease in model 3 (random *iid* intercept for regional watershed) at $P < .05$. NA, not available. *B,* ZIP codes shaded according to their observed incidence rates (IR). All 31 rivers are plotted on the map. Those rivers not included in the final multivariate model are shown in gray; those that increased predicted incidence in the surrounding ZIP codes, in red; and those that decreased predicted incidence in the surrounding ZIP codes in green. IRR, incidence rate ratio.

Figure 3). Table 4 shows that a 5% average increase in relative humidity and a 1–standard deviation average increase in stream flow over a period of a week were also predictors of disease 2 weeks later in univariate analysis. Increasing temperature was associated with a decrease in disease rates, most prominently at a 1-week lag. No combinations of variables in a multivariate analysis created a model fit equal to or better than that of precipitation alone.

DISCUSSION

This study analyzed a variety of environmental predictors of legionellosis, revealing a relationship between incidence and the proximity of ZIP codes to specific rivers and watersheds. While previous work by Ng et al [28] has shown that changes in watershed hydrology can predict cases over time, these results suggest that environmental reservoirs may have a greater influence on the spatial distribution of sporadic legionellosis cases than previously thought.

There are a number of plausible explanations for the association between elevated incidence of legionellosis and proximity to rivers. These possibilities can broadly be divided into (1) direct exposure to aerosolized river water and (2) contamination of drinking water supplies in an area. *Legionella* could be aerosolized directly from river water by several mechanisms. Some power plants make use of river water as part of the cooling process. Although there are regulations on the temperature of both the water entering and the water discharged from a power plant, these temperatures are still warm enough to promote the growth the *Legionella* within plant cooling towers, and at least one outbreak of pneumonia has been linked to a power plant cooling tower [29–32]. Likewise, sewage treatment plants can aerosolize bacteria during their treatment process, and a study

Table 2. Change in Rates of Legionellosis Associated With Proximity to Rivers

	Univariate IRR (95% CI) by Distance from River or Brook			
River/Brook	<5 km	<10 km	<15 km	Multivariate Model (Model 4)
Quinebaug River	2.31 (1.55–3.43)	2.63 (1.83–3.76)[a]	2.14 (1.52–3.02)	4.37 (2.96–6.44)
Hockanum Brook	2.18 (1.38–3.44)	1.87 (1.36–2.56)[a]	1.41 (1.13–1.75)	1.71 (1.21–2.42)
Farmington River	0.92 (.61–1.37)	1.15 (.93–1.43)	1.29 (1.07–1.56)[a]	1.41 (1.15–1.73)
Naugatuck River	1.42 (1.13–1.77)[a]	1.20 (.98–1.47)	1.22 (1.02–1.45)	1.32 (1.02–1.69)
Saugatuck River	0.56 (.36–.86)	0.50 (.36–.69)[a]	0.65 (.52–.80)	0.60 (.42–.84)
Shetucket River	0.28 (.09–.90)[a]	0.80 (.46–1.39)	0.67 (.41–1.10)	0.14 (.04–.45)

Abbreviations: CI, confidence interval; IRR, incidence rate ratio.

[a]Distance at which a river/brook had its most significant effect (lowest Akaike information criterion score) in each univariate model; this was also the distance cutoff used in the multivariate river model (model 4).

Table 3. Association Between Water Supplier and Watershed

Water Supplier	Cases, No. (Column %)[a] Watersheds Positively Associated (Naugatuck, Quinebaug)	Watersheds Not Associated	*P* Value (χ^2 Test)
Private well water	38 (31.40)	99 (17.01)	<.001
Municipal water	83 (68.60)	483 (82.99)	

[a]Percentages may not sum to 100% owing to rounding. The 4 cases assigned to multiple public water suppliers were excluded from analysis.

Table 4. Association Between Meteorological/Stream Flow Variables and Legionellosis

Variable	IRR (95% CI)[a]
5-mm increase in precipitation (2-wk lag)[b]	1.48 (1.38–1.59)
5% increase in relative humidity (2-wk lag)	1.24 (1.17–1.31)
1-SD increase in stream flow (2-wk lag)	1.26 (1.18–1.35)
5°C increase in maximum air temperature (1-wk lag)	0.81 (0.70–0.93)

Abbreviations: CI, confidence interval; IRR, incidence rate ratio; SD, standard deviation.

[a]IRRs corresponding to the most significant lag period for each variable. IRRs were calculated from the univariate Poisson regression for each meteorological variable at each lead/lag period.

[b]A 5-mm increase in precipitation with a 2-week lag had the lowest Akaike information criterion (AIC) score in the univariate analysis. See Supplementary Figure S4*B* for the complete list of corresponding AIC scores.

by the Environmental Protection Agency noted high antibody prevalence rates (23% seroconverted) among persons residing within 1.6 km of sewage treatment plants [33].

These explanations would require more knowledge regarding the exact location of residence and movement patterns of case patients in relation to these industries. Direct aerosolization might also occur after rain events owing to increased turbidity of the river water. Finally, a study by Sakamoto et al [34] put forth the hypothesis that car tires aerosolize *Legionella* after rain events, thereby increasing its presence in the environment during warm, wet weather.

Individuals could also be exposed to *Legionella* at their residence via contaminated aerosolized potable water. The spatial clustering present around certain river systems and within watersheds could then be related to localized contamination of surface water or ground water source that case patients are exposed to either through private wells or inadequately treated municipal water systems [35]. The subsequent increase in cases after rainfall events could results from disruption of sediments or biofilms and release of the bacteria, which might strain water treatment plants or contaminate ground water sources [35]. Studies of tap and groundwater samples have shown that, despite water treatment, *Legionella* can still be isolated from drinking water and groundwater and that its presence can fluctuate over time [36]. Data on drinking water disruptions, contamination of well water and aquifers, and ZIP code-level data on the proportion of residents with private well water would be necessary to build on these hypotheses.

Although we did control for population density, median age of housing, and testing rates in 2009, there could be other demographic factors that were not appropriately captured in our model and are associated with rivers, such as aging industrial infrastructure (eg, cooling towers) [37]. We did not have access to each year's proportion of pneumonia discharges tested for *Legionella*, which could prove an issue if testing rates drastically fluctuate by hospital over time. Using the proportion tested does not provide information on the severity of disease

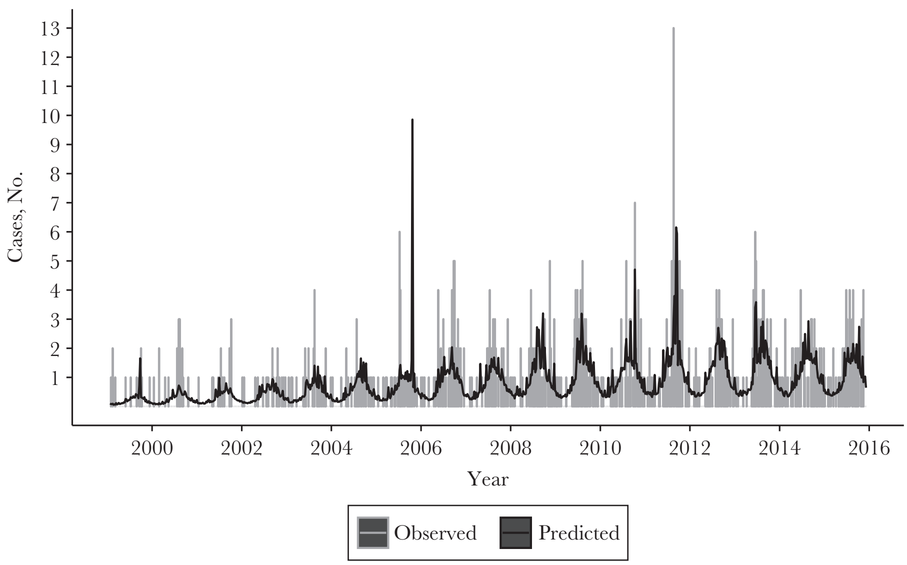

Figure 3. Precipitation 2 weeks earlier as a predictor of weekly case patient hospitalizations. Data represent observed (gray) and predicted (black) cases of legionellosis by week in Connecticut, determined using a Poisson regression model with covariates for week, week squared, yearly sine and cosine terms, and precipitation (with a 2-week lag).

being tested. It is possible that some hospitals test infrequently but test only high-risk patients and therefore have a higher case-capture rate.

We also explored the role of the drinking water suppliers and water quality of the rivers as possible explanations for the relationship with rivers. The aquatic life use support assessment did not explain a river's association with increased disease, which could be due to the inability of this metric to capture the intricacies of a river's capacity to harbor and spread pathogenic *Legionella* [38–40]. Because not all case patients in each ZIP code obtained their water from the same PWS, PWS was not analyzed as a predictor of ZIP code incidence rate in our models. Although previous studies have noted private water supply as a risk factor for legionellosis [41, 42], the higher proportion of private well water among case patients in high-risk watersheds seen in our study could be confounded by the lack of municipal water supply in more rural areas of the state. Neither the Environmental Protection Agency nor the state of Connecticut require or regulate the testing of residential private well water, which could allow for the accumulation of *Legionella* and biofilms if owners do not adequately test and properly maintain their own wells [43].

These analyses also confirmed previous findings of a temporal relationship between climatic conditions and the incidence of legionellosis, but the exact mechanisms behind this association remain to be delineated [13–15]. An increase in cases 2 weeks after increases in precipitation and humidity aligns with the estimated 2–10-day incubation period for legionellosis [44]. These results showed that increases in temperature had a protective effect on cases at all lags, but it is likely that any single week of temperature is not predictive; rather, consecutive weeks of warmer weather are predictive [45]. Although we were limited by the availability of stream temperature data, it is also likely that warming stream temperatures could be predictive of increases in cases during the summer months. It is probable that stream flow did not perform as well as precipitation in the model because rivers have a certain capacity to handle excess rainfall, which could explain the tendency of stream flow to underestimate cases (Supplementary Figure S5).

The current study has important limitations, including availability of patient information and changing trends in diagnostic measures. Patients reporting out-of-state travel during the exposure period were not excluded because this information was not routinely collected until 2007. Exclusion of these potential travel-related cases did not alter the spatial distribution for cases occurring after 2007, and high-incidence watersheds were not more likely to have travel-related cases, indicating little correlation between potential travel-related exposure and case patient residence (Supplementary Table S4). Information on workplace location was not available, but if patients were exposed at the workplace this could alter the spatial pattern seen in this analysis.

Future analysis will need to focus on the association between precipitation events and the presence (and density) of *Legionella* in natural hydrological systems and determine whether such an association accounts for the spatial clustering around river systems seen in this study. Further research is also needed to determine mechanisms by which *Legionella* aerosols are disseminated from natural hydrological systems. Due to the relationship between legionellosis, precipitation, humidity, and the optimal growth of *Legionella* in warm waters, investigation into the potential influence of changes in climatic factors on sporadic cases of disease may also be necessary when making predictive models [46].

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Financial support. This work was supported by the Yale Climate Change and Health Initiative through a grant from the Overlook International Foundation and by the National Center for Emerging and Zoonotic Infectious Diseases at the Centers for Disease Control and Prevention through a gift from the Overlook Emerging Infections Program.

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Neil K, Berkelman R. Increasing incidence of legionellosis in the United States, 1990–2005: changing epidemiologic trends. Clin Infect Dis **2008**; 47:591–9.
2. Fraser DW, Tsai TR, Orenstein W, et al. Legionnaires' disease. N Engl J Med **1977**; 297:1189–97.
3. van Heijnsbergen E, Schalk JAC, Euser SM, Brandsema PS, den Boer JW, de Roda Husman AM. Confirmed and potential sources of *Legionella* reviewed. Environ Sci Technol **2015**; 49:4797–815.
4. Marston BJ, Plouffe JF, File TM, et al. Incidence of community-acquired pneumonia requiring hospitalization. Arch Intern Med **1997**; 157:1709–18.
5. von Baum H, Ewig S, Marre R, et al. Community-acquired *Legionella* pneumonia: new insights from the German Competence Network for Community Acquired Pneumonia. Clin Infect Dis **2008**; 46:1356–64.
6. Stout JE, Yu VL. Legionellosis. N Engl J Med **1997**; 337:682–7.

7. Yu VL, Ramirez J, Roig J, Sabria M. Legionnaires disease and the updated IDSA guidelines for community-acquired pneumonia. Clin Infect Dis **2004**; 39:1734–1737.
8. Fields BS, Benson RF, Besser RE. *Legionella* and legionnaires' disease: 25 years of investigation. Clin Microbiol Rev **2002**; 15:506–26.
9. Grimes DJ. Ecology of estuarine bacteria capable of causing human disease: a review. Estuaries **1991**; 14:345–60.
10. van Heijnsbergen E, de Roda Husman AM, Lodder WJ, et al. Viable *Legionella pneumophila* bacteria in natural soil and rainwater puddles. J Appl Microbiol **2014**; 117: 882–90.
11. Fields BS, Benson RF, Besser RE, et al. *Legionella* and legionnaires' disease: 25 years of investigation. Clin Microbiol Rev **2002**; 15:506–26.
12. Declerck P. Biofilms: the environmental playground of *Legionella pneumophila*. Environ Microbiol **2010**; 12:557–66.
13. Fisman DN, Lim S, Wellenius GA, et al. It's not the heat, it's the humidity: wet weather increases legionellosis risk in the greater Philadelphia metropolitan area. J Infect Dis **2005**; 192:2066–73.
14. Halsby KD, Joseph CA, Lee JV, Wilkinson P. The relationship between meteorological variables and sporadic cases of legionnaires' disease in residents of England and Wales. Epidemiol Infect **2014**; 142:2352–2359.
15. Karagiannis I, Brandsema P, Van Der Sande M. Warm, wet weather associated with increased legionnaires' disease incidence in The Netherlands. Epidemiol Infect **2009**; 137:181–187.
16. Tilburg CE, Jordan LM, Carlson AE, Zeeman SI, Yund PO. The effects of precipitation, river discharge, land use and coastal circulation on water quality in coastal Maine. R Soc Open Sci **2015**; 73:501–16.
17. Wellnitz T, Rader RB. Mechanisms influencing community composition and succession in mountain stream periphyton: interactions between scouring history, grazing, and irradiance. J North Am Benthol Soc **2003**; 22: 528–41.
18. Fliermans CB, Cherry WB, Orrison LH, Smith SJ, Tison DL, Pope DH. Ecological distribution of *Legionella pneumophila*. Appl Environ Microbiol **1981**; 41:9–16.
19. Den Boer JW, Coutinho RA, Yzerman EPF, van der Sande MAB. Use of surface water in drinking water production associated with municipal legionnaires' disease incidence. J Epidemiol Community Health **2008**; 62:e1.
20. Rudbeck M, Jepsen MR, Sonne IB, Uldum SA, Viskum S, Molbak K. Geographical variation of sporadic legionnaires' disease analysed in a grid model. Epidemiol Infect **2010**; 138:9–14.
21. Centers for Disease Control and Prevention. Legionellosis—United States, 2000–2009. MMWR. **2011**; 60:1083–6.
22. Gacek P, Soto K, Brockmeyer J, Krasnitski Mph J, Marquez C, Rabatsky-Ehr T. Legionellosis in Connecticut, 2003–2012. Connect Epidemiol **2013**; 33:13–6.
23. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance (ABCs) report: Emerging Infections Program Network: legionellosis, 2012. **2012** https://www.cdc.gov/abcs/reports-findings/survreports/leg12.pdf. Accessed 27 March 2017.
24. Sosa L, Hadler J, Cartter M, et al. Investigation of a possible outbreak of legionnaires' disease—Connecticut, 2006. Connect Epidemiol **2006**; 26:21–4.
25. US Census Bureau. Census 2000 Summary File 1: Geographic identifiers. **2000**. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF1_G001&prodType=table. Accessed 2 February 2017.
26. US Census Bureau. 2000 Summary file 3: housing characteristics. **2000.** https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF3_DP4&prodType=table. Accessed 24 July 2017.
27. Valciņa O, Pūle D, Lucenko I, et al. *Legionella pneumophila* seropositivity-associated factors in Latvian blood donors. Int J Environ Res Public Health **2015**; 13:58.
28. Ng V, Tang P, Jamieson F, et al. Going with the flow: legionellosis risk in Toronto, Canada is strongly associated with local watershed hydrology. Ecohealth **2008**; 5:482–90.
29. Madden N, Lewis A, Davis M. Thermal effluent from the power sector: an analysis of once-through cooling system impacts on surface water temperature. Environ Res Lett **2013**; 8:35006.
30. USGS Water Science School. Thermoelectric power water use. **2016.** https://water.usgs.gov/edu/wupt.html. Accessed 28 March 2017.
31. Walczak M, Kletkiewicz H, Burkowska A. Occurrence of *Legionella pneumophila* in lakes serving as a cooling system of a power plant. Environ Sci Process Impacts **2013**; 15:2273–2278.
32. Morton S, Bartlett CL, Bibby LF, Hutchinson DN, Dyer JV, Dennis PJ. Outbreak of legionnaires' disease from a cooling water system in a power station. Br J Ind Med **1986**; 43:630–5.
33. Northrop R, Becker C, Cordell R, et al. Health effects of sewage aerosols: additional serological surveys and search for *Legionella pneumophila* in sewage. Cincinnati, Ohio; **1981.** https://nepis.epa.gov/Exe/ZyNET.exe/20016SUX.TXT?ZyActionD=ZyDocument&Client=EPA&Index=1981+Thru+1985&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C81thru85%5CTxt%5C00000013%5C20016SUX.

txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=1&SeekPage=x&ZyPURL.
34. Sakamoto R, Ohno A, Nakahara T, et al. *Legionella pneumophila* in rainwater on roads. Emerg Infect Dis **2009**; 15:1295–7.
35. Riffard S, Douglass S, Brooks T, Springthorpe S, Filion LG, Sattar SA. Occurrence of *Legionella* in groundwater: an ecological study. Water Sci Technol **2001**; 43:99–102.
36. Donohue MJ, O'Connell K, Vesper SJ, et al. Widespread molecular detection of *Legionella pneumophila* serogroup 1 in cold water taps across the United States. Environ Sci Technol **2014**; 48:3145–52.
37. Bhopal RS, Fallon RJ, Buist EC, Black RJ, Urquhart JD. Proximity of the home to a cooling tower and risk of non-outbreak legionnaires' disease. BMJ **1991**; 302:378–83.
38. Taylor M, Ross K, Bentham R. *Legionella,* protozoa, and biofilms: interactions within complex microbial systems. Microb Ecol **2009**; 58:538–47.
39. Ponsatí L, Corcoll N, Petrović M, et al. Multiple-stressor effects on river biofilms under different hydrological conditions. Freshw Biol **2016**; 61:2102–15.
40. Marrão G, Veríssimo A, Bowker RG, Costa MS. Biofilms as major sources of *Legionella* spp. in hydrothermal areas and their dispersion into stream water. FEMS Microbiol Ecol **1993**; 12:25–33.
41. Straus WL, Plouffe JF, File TM, et al. Risk factors for domestic acquisition of legionnaires disease. Arch Intern Med **1996**; 156:1685.
42. Flannery B, Gelling LB, Vugia DJ, et al. Reducing *Legionella* colonization in water systems with monochloramine. Emerg Infect Dis **2006**; 12:588–96.
43. Connecticut Department of Public Health. Private well water systems in Connecticut: best management practice checklist. **2016**. http://www.ct.gov/dph/lib/dph/environmental_health/private_wells/pdf/091416_bmps_private_well_water_systems.pdf. Accessed 5 August 2017.
44. Farnham A, Alleyne L, Cimini D, Balter S. Legionnaires' disease incidence and risk factors, New York, New York, USA, 2002–2011. Emerg Infect Dis **2014** ; 20:1795–802.
45. Brandsema PS, Euser SM, Karagiannis I, den Boer JW, van der Hoek W. Summer increase of legionnaires' disease 2010 in the Netherlands associated with weather conditions and implications for source finding. Epidemiol Infect **2014**; 142:2360–71.
46. Greer A, Ng V, Fisman D. Climate change and infectious diseases in North America: the road ahead. CMAJ **2008**; 178:715–22.