

# The relationship between meteorological variables and sporadic cases of Legionnaires' disease in residents of England and Wales

K. D. HALSBY<sup>1</sup>\*, C. A. JOSEPH<sup>2</sup>, J. V. LEE<sup>2</sup> AND P. WILKINSON<sup>3</sup>

<sup>1</sup> Gastrointestinal, Emerging and Zoonotic Infections Department, Public Health England, London, UK

<sup>2</sup> Independent Consultant (formerly of the Health Protection Agency)

<sup>3</sup> Department of Social and Environmental Health Research, London School of Hygiene & Tropical Medicine, London, UK

Received 11 June 2013; Final revision 16 October 2013; Accepted 5 December 2013;  
first published online 9 January 2014

## SUMMARY

We studied the timing of occurrence of 1676 sporadic, community-acquired cases of Legionnaires' disease in England and Wales between 1993 and 2008, in relation to temperature, relative humidity, rainfall, windspeed and ultraviolet light using a fixed-stratum case-crossover approach. The analysis was conducted using conditional logistic regression, with consideration of appropriate lag periods. There was evidence of an association between the risk of Legionnaires' disease and temperature with an apparently long time lag of 1–9 weeks [odds of disease at 95th vs. 75th centiles: 3·91, 95% confidence interval (CI) 2·06–7·40], and with rainfall at short time lags (of 2–10 days) (odds of disease at 75th vs. 50th centiles: 1·78, 95% CI 1·50–2·13). There was some evidence that the risk of disease in relation to high temperatures was greater at high relative humidities. A higher risk of Legionnaires' disease may be indicated by preceding periods of warmer wetter weather.

**Key word:** Legionnaires' disease.

## INTRODUCTION

Legionnaires' disease (LD) is an atypical pneumonic infection, with over 5000 cases diagnosed across Europe each year [1]. *Legionella* bacteria are aquatic and aerobic, and are ubiquitous in natural and artificial water environments worldwide [2]. They are frequently found in domestic and public water systems, and in evaporative cooling water systems used for air conditioning or industrial cooling [3]. The organism infects humans through aerosol transmission (or rarely through aspiration).

Laboratory studies have demonstrated that temperature and relative humidity (RH) play important roles in the growth and replication of legionellae [4–7], and that ultraviolet (UV) light can inhibit growth of the bacteria [8]. Epidemiological observational studies suggest that there is an association between case numbers and both temperature and RH [9], and also suggest that there may be an association with rainfall [10, 11]. Sunlight duration may also play a role, but its effects are less clear [11]. A combination of warm temperature and high RH have been recorded multiple times in association with outbreaks [12, 13], while gentle winds [14, 15] and heavy rains have also been implicated [16, 17].

In order for legionellae in the environment to cause disease, the conditions in the water source must be suitable for bacterial growth and replication, aerosols

\* Author for correspondence: Dr K. D. Halsby, Gastrointestinal, Emerging and Zoonotic Infections Department, Public Health England, 61 Colindale Avenue, London, NW9 5EQ, UK.  
(Email: kate.halsby@phe.gov.uk)

containing the organism must be formed by an apparatus capable of generating an aerosol, and the organism must survive in the atmosphere long enough to disseminate widely in order to encounter a susceptible host. Meteorological factors may influence any of the stages of this process and thereby influence the probability of infection.

This paper uses a case-crossover analysis to examine the relationship between sporadic cases of community-acquired LD in England and Wales and five meteorological variables: temperature, RH, rainfall, wind-speed, and UV light.

## METHODS

### The case dataset

Information on all cases of LD diagnosed in residents of England and Wales is collected by Public Health England (PHE) [formerly the Health Protection Agency (HPA), Colindale]. Cases are diagnosed by local laboratories and reported to their health protection team which has responsibility for completing a national surveillance questionnaire for each case containing demographic, clinical and microbiological information, as well as information on the case's exposure history. This is then sent to the Communicable Infectious Disease Surveillance Centre, PHE and entered into the national dataset. All community-acquired cases (individuals who did not report overnight travel or hospital stay during their incubation period) that occurred between 1993 and 2008, and which had not been associated with an outbreak, were selected for inclusion in this study.

Between 1993 and 2008, 2173 community-acquired cases were reported for English or Welsh residents. Of these, 470 were known to be associated with an outbreak and were removed from the dataset. To aid the analyses, those cases that could not be allocated to a region of residence ( $n=13$ ), and those cases without a date of onset ( $n=14$ ), were also removed from the study. This left 1676 cases for analysis.

### The weather dataset

Weather data were obtained from Met Office Land Surface Observation Stations Data ('MIDAS' – Met Office Integrated Data Archive System) held by the British Atmospheric Data Centre [18]. Data for each variable were downloaded from all available weather stations which had readings for at least 75% of days

for the period 1993–2008 for that variable. A single regional series was derived for each parameter, weighted by population.

Temperature data was extracted as maximum daily air temperatures in degrees Celsius ( $^{\circ}\text{C}$ ) to the nearest  $0.1^{\circ}\text{C}$ . RH data was extracted as hourly dewpoint temperature and hourly air temperature, both measured to the nearest  $0.1^{\circ}\text{C}$ . RH was calculated using the following formula:

$$\text{RH} = 100 \times \exp\left(2.453 \times \frac{10^6}{461.5} \times \left(\frac{1}{273.15 + \text{airtemp}} - \frac{1}{273.15 + \text{dewpt}}\right)\right).$$

Readings were available for 09:00 and 15:00 hours each day, and the mean of these two readings was used to comprise the daily dataset for the study.

Rainfall data was extracted as the daily precipitation amount in millimetres, and wind data was extracted as the daily mean wind speed in knots. Data was imputed for any missing daily readings for each weather variable (with the exception of rainfall data), using the AIRGENE method which was developed for use with air pollution data [19]. Missing rainfall data were calculated separately because of a strong positive skew observed in the data, using a simple alternative formula:

$$\hat{x}_{ij} = \bar{z}_i \times \frac{\bar{x}_{jk}}{\bar{z}_k},$$

where  $i$  = date,  $j$  = monitor,  $k$  = month,  $\bar{x}_{jk}$  = period average of monitor,  $\bar{z}_i$  = mean regional value for 1 day, and  $\bar{z}_k$  = mean regional value for 1 month.

Data on UV radiation for England and Wales was obtained from HPA Chilton. Data series were available for three sites (Camborne, Chilton, Leeds), and each region was allocated the data series for the closest geographical station. The UV data are based on 5-min averages of erythemally effective irradiance (EEI) (measured in  $\text{mW}/\text{m}^2$ ). These readings are averaged across the hour, multiplied by 3600 s, and summed across the day to give the total radiant exposure for each day.

### Analytical methods

The relationship between the risk of LD and each meteorological variable was examined using a fixed stratum case-crossover analysis. Time was divided into periods of 28 days ('strata'), and the stratum within which each case fell provided the control set for that

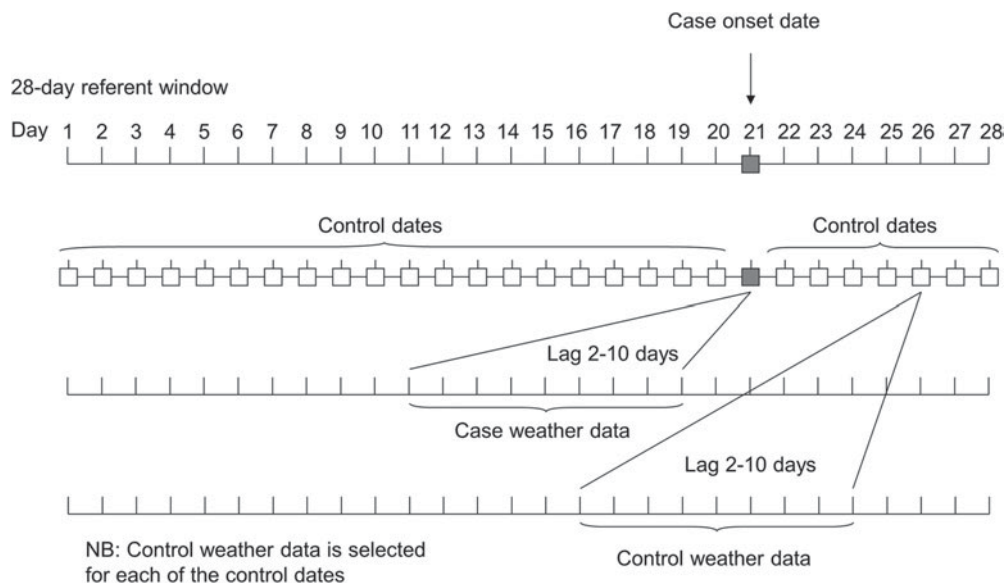


Fig. 1. Schematic representation of selection of controls and weather data used in the analysis.

case. The date of onset for each case was therefore allocated 27 control dates from the rest of the 28-day stratum, and together they formed a matched case-control set (see Fig. 1). These matched sets were analysed by conditional logistic regression, adjusting for 'day of the week' and for a 28-day linear term representing day within each lunar month.

Time lags between meteorological conditions and disease onset were selected to reflect a lag due to (i) the incubation period (2–10 days) of LD (representing the delay between the meteorological conditions at the time of dispersal and the subsequent onset of disease) and (ii) for temperature and rainfall, the potentially longer lagged effect of environmental conditions on the growth of the organism in the environment. Seven-day moving averages were used to construct sets of weekly lag periods which were entered simultaneously into the regression models. Wald tests were performed to assess the value of adding each additional lagged term.

The (lagged) associations between the risk of LD and meteorological conditions were examined (i) using quartiles of each weather variable (with the overall coefficient for each quartile determined by linear combination of the coefficient for each component lag period included), and (ii) graphically by using natural cubic spline functions of each weather parameter fitted using the *SPBASE* command in Stata v. 9 (three internal knots placed at equally spaced percentiles, graphed as the predicted risk of LD relative to the mean of quartile 1) [20]. Robust standard errors

were calculated clustering on government office region (GOR) to allow for spatial correlations within the data.

The relationship between each case of LD and each weather variable was examined initially in univariable analyses and then in a multivariable model that included all five meteorological variables regardless of statistical significance. Each variable was entered into the model using a natural cubic spline function with three internal knots, and with the lag structure developed in the single-weather analyses (these were largely lag periods of 2–10 days; however, week-of-lag terms were also considered for the temperature and rainfall models, and were used where there was evidence of a clear improvement in model fit).

An interaction between temperature and RH has been previously suggested in the literature [21], and we tested this using a simplified version of the multivariable model. RH was represented as a binary variable above and below the 25th centile (66·21%; lag 2–10 days), and temperature by a linear threshold model, with the threshold at the 25th centile (11·35 °C; lag 0–9 weeks).

## RESULTS

The number of sporadic, community-acquired cases of LD reported to PHE has increased substantially over the 16-year period covered by the study: the average number of cases per year between 1993 and 2000 was 62·5, and between 2001 and 2008 was 147.

Table 1. Meteorological data for case and control days (where a lag period is specified, the data is an average across that time period)

	Cases		Control days	
	Mean	IQR	Mean	IQR
Daily maximum temp. (°C)				
No lag	15.74	11.14–19.93	15.84	11.36–20.09
Lag 2–10 days	15.90	11.42–20.06	15.97	11.29–20.35
Lag 0–9 weeks	16.33	12.00–20.35	16.28	11.74–20.33
Relative humidity (%)				
No lag	74.52	66.29–83.26	74.53	66.20–83.70
Lag 2–10 days	74.88	69.80–80.72	74.32	68.76–80.53
Rainfall (mm)				
No lag	2.41	0.02–2.84	2.29	0.02–2.89
Lag 2–10 days	2.55	1.02–3.71	2.32	0.77–3.33
Windspeed (knots)				
No lag	8.52	5.56–10.69	8.39	5.51–10.61
Lag 2–10 days	8.43	6.61–9.72	8.37	6.52–9.70
UV (/100 000 mW/m <sup>2</sup> )				
No lag	11.37	2.73–18.06	11.35	2.68–18.06
Lag 2–10 days	11.47	3.25–18.21	11.69	3.25–18.67

IQR, Interquartile range.

The male-to-female sex ratio of cases in the dataset was 3.6:1, and the highest proportion of cases (30.5%) fell within the 50–59 years age group. The majority of cases (41.4%) occurred in late summer, between July and October, consistent with European data [1].

Each weather parameter was divided into quartiles. There were 1676 cases and 45 204 control days (each case was allocated 27 control days, with a small number of exceptions that fell at the beginning or end of the 1993–2008 time period, where meteorological data were available for fewer than 27 days). Summary data is presented in Table 1, with various lag periods.

### Univariable analysis

The univariable model for each variable, with quartiles and cubic splines can be seen in Figure 2, and odds ratios (OR) for 95th vs. 75th centiles, and 75th vs. 50th centiles are shown in Table 2.

Lag effects were examined for 2–10 days and also by week of lag up to 12 weeks for temperature and rainfall. All week-of-lag terms were included where there was evidence of clear improvement in model fit. Only in the case of temperature was there such evidence (for lag effects up to 9 weeks), as shown in Table 3. For maximum daily temperature, we therefore included weekly terms for lags 1–9 weeks. For all other variables we used lags of 2–10 days.

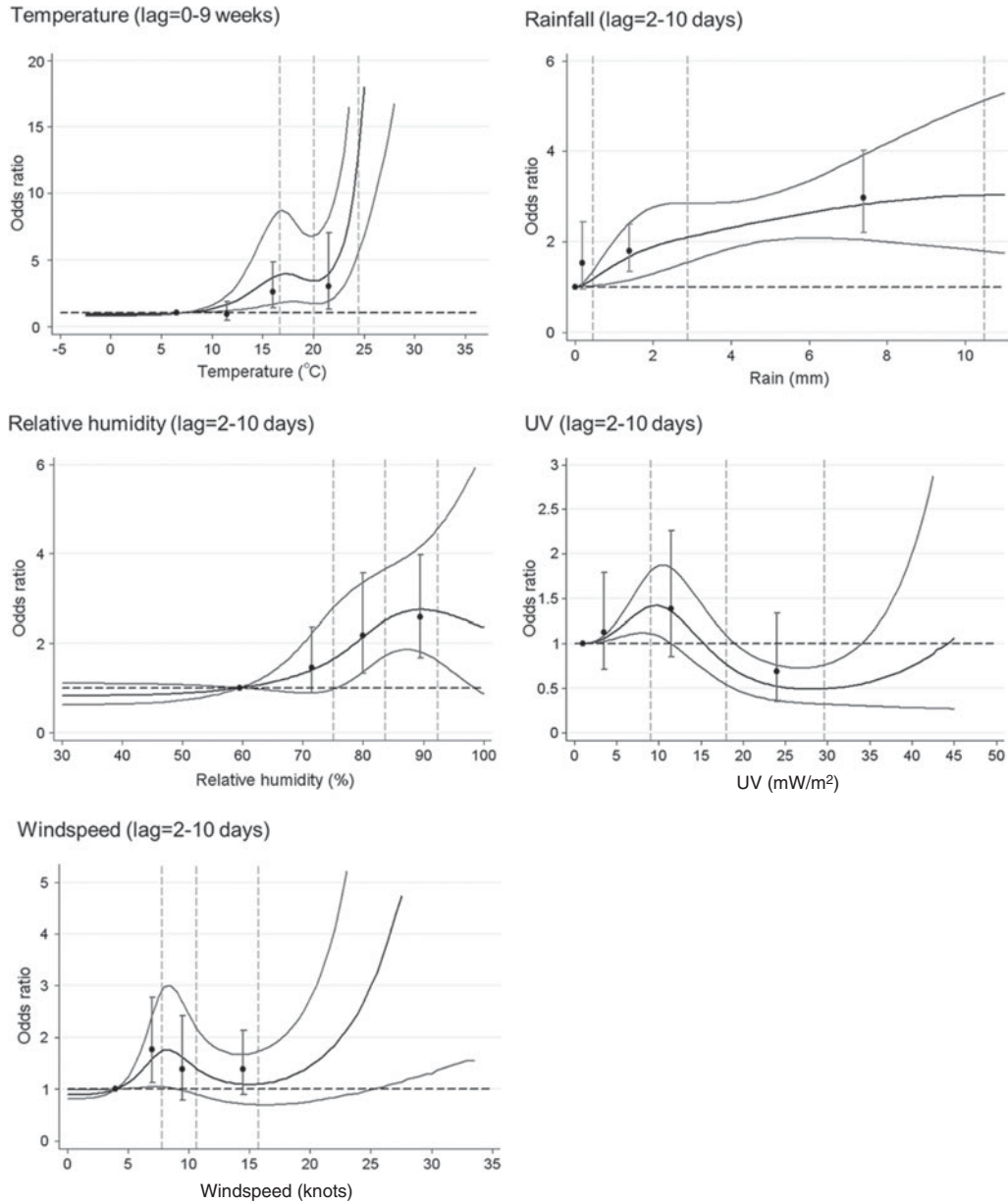
### Multivariable model

For most weather variables, the patterns of association seen in the univariable models were broadly similar in the multivariable model (Table 2). The point estimate for the 95th vs. 75th centile comparison was slightly stronger in the multivariable model for temperature, and appreciably so for UV exposure, but somewhat lower for rainfall. The point estimates for the 75th vs. 50th centile comparison was weaker in the multivariable model for RH and rainfall, but not materially altered for the other variables. Overall, the evidence for windspeed and UV suggested no clear association with risk of LD, while temperature, RH and rainfall all showed some evidence of increased risk at mid to high levels.

The analysis of potential effect modification between temperature and RH suggests that the effect of temperature increased from an OR of 1.08 (for each degree increase above 11.35 °C) for RHs below the 25th centile (<66.2%), to 1.18 for RHs ≥25th centile. The inclusion of the interaction term significantly improved the model fit ( $P < 0.0001$ ).

### DISCUSSION

This study provides evidence of an association between sporadic, community-acquired cases of LD and some meteorological conditions, independent of



**Fig. 2.** Univariable models for each meteorological parameter: natural cubic spline curves with point estimates and 95% confidence intervals for odds ratios by quartile of the parameter distribution. Vertical dotted lines indicate the 50th, 75th and 95th centiles of the relevant parameter distribution.

season. It suggests that temperature and rainfall, and to lesser extent RH, may be associated with the risk of sporadic disease.

There appeared to be a lag of up to 9 weeks between air temperatures and disease risk, consistent with an effect of temperature on the growth and spread of the bacterium in the environment. *Legionella* bacteria are slow growing, and require the support of other organisms in order to multiply in the environment (e.g. algae, other bacteria, protozoa). Growth of photosynthetic primary producers (e.g.

algae and cyanobacteria) will be followed by heterotrophic bacteria and other organisms, the protozoa that feed on them, and finally the legionellae that grow in the protozoa [22, 23]. The long lag period may therefore reflect the time taken for the populations of other organisms to multiply in the environment and support *Legionella* growth. Additionally, the long lag period may reflect the time taken for air temperatures to warm the environment, including ponds and rivers; most sources of LD are fed by mains water systems, and mains water does not heat up quickly.



Table 2. Odds ratios (OR) and 95% confidence intervals (CI) for risk of Legionnaires' disease for each meteorological parameter in univariable and multivariable models. The multivariable model includes all weather variables fitted simultaneously

Parameter/model	OR (95% CI)	
	95th vs. 75th centile	75th vs. 50th centile
Temperature		
Univariable model	3.91 (2.06–7.40)	0.87 (0.45–1.70)
Multivariable model	4.51 (1.92–10.56)	0.92 (0.41–2.05)
Relative humidity		
Univariable model	1.08 (0.78–1.49)	1.52 (1.21–1.90)
Multivariable model	1.08 (0.76–1.54)	1.20 (0.85–1.68)
Rainfall		
Univariable model	1.44 (0.69–3.01)	1.78 (1.50–2.13)
Multivariable model	1.14 (0.64–2.05)	1.55 (1.13–2.14)
Windspeed		
Univariable model	0.78 (0.53–1.16)	0.80 (0.64–1.00)
Multivariable model	0.75 (0.50–1.15)	0.72 (0.58–0.90)
Ultraviolet		
Univariable model	0.65 (0.45–0.93)	0.54 (0.43–1.67)
Multivariable model	1.09 (0.68–1.74)	0.66 (0.51–0.85)

When RH was investigated in a single-weather model it appeared to show an association with LD, consistent with the findings of previous studies. However, the inclusion of additional weather variables in the model reduced this association, implying that the apparent relationship may have been driven by other weather parameters. This study also provides evidence of an association with rainfall during a case's incubation period, which may equate to the time of aerosol dispersal. This could be a result of rainfall 'stirring up' or 'flushing through' water systems; such conditions have been identified as important during outbreak situations [16, 17].

There appears to be evidence of association between conditions of low windspeed and a risk of LD, but the confidence intervals for the regression model were wide. Biologically it is plausible that any association would be strongest at very low, gentle windspeeds which would allow the aerosolized organism to disseminate but which would not be strong enough to break up the aerosol. However, it is difficult for weather stations to record windspeeds <2 m/s (equivalent to 3.89 knots) [24]. In addition, the approach used in this analysis of aggregating mean data from a large number of weather stations would tend to reduce extreme measurements.

The results from the UV model suggest that there may be an association present in the data, with a

lower risk of disease >15 mW/m<sup>2</sup> (and a generally protective effect at all levels in the multivariable model). This is biologically plausible: high UV levels may damage the aerosolized bacteria and reduce the risk of infection.

### Limitations

The study inevitably has some limitations. Some of the lags identified were very long. This is unusual and it is difficult to disentangle the influence of the weather over a 9-week period from other seasonal effects. This study attempted to deal with the issue by controlling for seasonality through use of a case-crossover design, the addition of a linear 28-day term, and the inclusion of 'day of the week' as a categorical variable. This in itself is not without risk; there is a danger of a downwards bias in the results of the regression due to overly aggressive fitting of time trends.

Using regional meteorological data series may not fully represent the environmental conditions affecting the growth of legionellae since there can be appreciable geographical variations and disparities between day-time and night-time conditions.

This study also had to rely on the assumption that individuals contracted their disease within their region of residence. This was made more likely by the exclusion of any case that had been away from home

Table 3. Odds ratios (OR) and 95% (confidence intervals (CI) for risk of disease at 24.5 °C vs. 16.73 °C (95th vs. 50th centile) at 1- to 12-week temperature lags, and P values for Wald tests of additional terms of lag. Based on natural cubic spline models with three internal knots

Weekly lags in model	OR (95% CI)	P value (week tested)
1	0.95 (0.67–1.37)	0.00 (week 1)
1–2	1.96 (1.27–3.02)	0.00 (week 2)
1–3	3.34 (2.37–4.71)	0.16 (week 3)
1–4	3.46 (1.91–6.28)	0.00 (week 4)
1–5	3.97 (2.26–6.98)	0.02 (week 5)
1–6	4.87 (3.39–6.99)	0.05 (week 6)
1–7	4.56 (2.51–8.29)	0.20 (week 7)
1–8	3.07 (1.91–4.94)	0.01 (week 8)
1–9	3.40 (1.86–6.20)	0.00 (week 9)
1–10	4.22 (2.05–8.68)	0.08 (week 10)
1–11	3.84 (1.49–9.86)	0.88 (week 11)
1–12	3.46 (1.20–9.97)	0.38 (week 12)

overnight during their incubation period; however, the study could not account for individuals who had made day-trips outside the region, although this is likely to have had only a minor effect with regard to the misclassification of weather conditions. There are other factors which may influence the survival of *Legionella* in the environment, and which were beyond the scope of this analysis. As an example, air pollution may influence the survival of legionellae during their transmission in aerosol.

#### Other literature

Much of the previous literature on sporadic cases of LD and weather factors is recent. In 2006 there was a sudden, unexpected increase in the number of sporadic cases occurring in Northern Europe [25, 26]. Investigators established that there was no change in the circulating strain of *Legionella* that could explain the increase in case numbers, and no new sources of infection were identified; it was instead hypothesized that weather conditions might have been responsible [11, 21, 27]. The findings in this paper are broadly comparable with the results of those studies.

A recent paper by Dunn *et al.* found an association between LD, RH and windspeed, but the association did not remain after controlling for season and year [28]. In contrast, our study suggests that associations do remain, even after close control for time-related variables.

#### Public health implications

The associations identified in this analysis suggest that current control measures do not adequately curb the growth and distribution of legionellae during periods of high risk. There may be an opportunity to better target prevention measures to counter these raised risks. For example, it might be possible to monitor air temperature using a 9-week moving average, with the aim of identifying periods of high risk. A reminder could then be issued to public health professionals and water treatment companies of the importance of ensuring their systems have proper control measures in place. This type of alert is already carried out on an informal basis when risk periods are identified through environmental testing.

The work presented in this study may have implications for the future burden of disease. If the associations with temperature demonstrated here reflect a causal relationship, then the influence of climate change on the number of cases of LD should be considered. This study suggests that temperature may affect the growth and replication of bacteria in the environment; rising temperatures may result in a greater bacteria load within water systems which, when released into the environment, may result in an increased likelihood of exposure to infected aerosols. In addition, raised temperatures can affect people's behaviour and alter their exposure to infection: air conditioning systems are more often used during warm periods, people shower more frequently, and windows are left open.

#### CONCLUSIONS

In conclusion, this study presents evidence to show that there is an association between selected meteorological conditions and the occurrence of sporadic cases of community-acquired LD for cases detected in residents of England and Wales. It provides evidence of association between the risk of LD and temperature, with an apparently long time lag and possibly modified by RH, and between LD and rainfall at short time lags. These associations may be useful in targeting public health interventions during periods of high risk, as identified by weather conditions.

#### DECLARATION OF INTEREST

None.

## REFERENCES

1. **Joseph CA, Ricketts KD.** Legionnaires' disease in Europe 2007–2008. *Eurosurveillance* 2010; **15**(8).
2. **Fliermans CB, et al.** Ecological distribution of *Legionella pneumophila*. *Applied and Environmental Microbiology* 1981; **41**: 9–16.
3. **Committee of Inquiry.** First report of the Committee of Inquiry into the outbreak of legionnaires' disease in Stafford in April 1985. 1986.
4. **EWGLI.** European Guidelines for Control and Prevention of Travel Associated Legionnaires' Disease, 2002.
5. **Berendt RF.** Survival of *Legionella pneumophila* in aerosols: effect of relative humidity. *Journal of Infectious Diseases* 1980; **141**: 689.
6. **Dennis PJ, Lee JV.** Differences in aerosol survival between pathogenic and non-pathogenic strains of *Legionella pneumophila* serogroup 1. *Journal of Applied Bacteriology* 1988; **65**: 135–141.
7. **Hambleton P, et al.** Survival of virulent *Legionella pneumophila* in aerosols. *Journal of Hygiene (Cambridge)* 1983; **90**: 451–460.
8. **Antopol SC, Ellner PD.** Susceptibility of *Legionella pneumophila* to ultraviolet radiation. *Applied and Environmental Microbiology* 1979; **38**: 347–348.
9. **Fisman DN, et al.** It's not the heat, it's the humidity: wet weather increases legionellosis risk in the greater Philadelphia metropolitan area. *Journal of Infectious Diseases* 2005; **192**: 2066–2073.
10. **Hicks LA, et al.** Increased rainfall is associated with increased risk for legionellosis. *Epidemiology and Infection* 2006; **135**: 811–817.
11. **Karagiannis I, et al.** Warm, wet weather associated with increased Legionnaires' disease incidence in The Netherlands. *Epidemiology and Infection* 2009; **137**: 181–187.
12. **Ferre M, et al.** A community outbreak of Legionnaires' disease associated with a cooling tower in Vic and Gurb, Catalonia (Spain) in 2005. *European Journal of Clinical Microbiology and Infectious Diseases (Wiesbaden)* 2009; **28**: 153–159.
13. **Sala MR, et al.** Community outbreak of Legionnaires' disease in Vic-Gurb, Spain in October and November 2005. *Eurosurveillance* 2007; **12**(3).
14. **Anderson P, et al.** Legionnaires' disease in Reading – possible association with a cooling tower. *Community Medicine* 1985; **7**: 202–207.
15. **Public Health Laboratory Service.** Report from the PHLS Communicable Disease Surveillance Centre. *British Medical Journal* 1986; **293**: 681–682.
16. **Hoyle F, Wickramasinghe NC, Watkins J.** Legionnaires' disease: seeking a wider source [Letter]. *Lancet* 1985; **325**: 1216–1217.
17. **Thacker SB, et al.** An outbreak in 1965 of severe respiratory illness caused by the legionnaires' disease bacterium. *Journal of Infectious Diseases* 1978; **138**: 512–519.
18. **British Atmospheric Data Centre.** (<http://badc.nerc.ac.uk/home/index.html>). Accessed 14 September 2010.
19. **Ruckerl R, et al.** Air pollution and inflammation (interleukin-6, C-reactive protein, fibrinogen) in myocardial infarction survivors. *Environmental Health Perspective* 2007; **115**: 1072–1080.
20. **StataCorp.** Stata v. 9. Stata Corporation, College Station, Texas, USA, 2010.
21. **Ricketts KD, et al.** Weather patterns and Legionnaires' disease: a meteorological study. *Epidemiology and Infection* 2009; **137**: 1003–1012.
22. **Fliermans CB.** Ecology of *Legionella*: from data to knowledge with a little wisdom. *Microbial Ecology* 1996; **32**: 200–228.
23. **Taylor M, Ross K, Bentham R.** *Legionella*, protozoa, and biofilms: interactions within complex microbial systems. *Microbial Ecology* 2009; **58**: 538–547.
24. **Met Office.** Met Office Surface Data Users Guide ([http://badc.nerc.ac.uk/data/ukmo-midas/ukmo\\_guide.html#5-5](http://badc.nerc.ac.uk/data/ukmo-midas/ukmo_guide.html#5-5)). Accessed 14 September 2010.
25. **Anon.** National increase in cases of Legionnaires' disease. *CDR Weekly* 2006; **16**.
26. **Joseph CA, Van Der Sande M.** Unexplained summer increase in non-travel related legionellosis in the UK and Netherlands. *Eurosurveillance* 2006; **11**(42).
27. **Conza L, et al.** Meteorological factors and risk of community-acquired Legionnaires' disease in Switzerland: an epidemiological study. *BMJ Open* 2013; **3**.
28. **Dunn C, et al.** Meteorological conditions and incidence of Legionnaires' disease in Glasgow, Scotland: application of statistical modelling. *Epidemiology and Infection* 2013; **141**: 687–696.