The shift in seasonality of legionellosis in the USA

T. M. Alarcon Falconi[1]*, M. S. Cruz[2]* and E. N. Naumova[1,2]

[1]School of Engineering, Tufts University, Medford, MA, USA and [2]Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA

Abstract

According to the Centers for Disease Control and Prevention (CDC), from 2000 to 2014, reported cases of legionellosis per 100 000 population increased by 300% in the USA, although reports on disease seasonality are inconsistent. Using two national databases, we assessed seasonal patterns of legionellosis in the USA. We created a monthly time series from 1993 to 2015 of reported cases of legionellosis from the CDC, and from 1997 to 2006 of medical claims of legionellosis-related hospitalisation in older adults from the Centers for Medicaid and Medicare Services (CMS). We split the study time interval into two segments (before and after 2003), and applied a Poisson harmonic regression model to each dataset and each segment. The time series of monthly counts exhibited a significant shift of seasonal peaks from mid-September (9.676 ± 0.164 months) before 2003 to mid-August (8.452 ± 0.042 months) after 2003, along with an alarming increase in the amplitude of seasonal peaks in both CDC and CMS data. The lowest monthly reported cases of legionellosis in 2015 (281) exceed the maximum value reported before 2003 (206). We also observed a discrepancy between CDC and CMS data, suggesting that not all cases of legionellosis diagnosed by hospital-based laboratories were reported to the CDC. Improved reporting of legionellosis is required to better inform the public and organise disease prevention.

Introduction

The ongoing *Flint water crisis* has revealed many environmental and public health problems stemming from poor water infrastructure, including the rise in Legionnaires' disease caused by waterborne bacteria, legionella. While the main focal point of Flint investigations has been the lead poisoning of county residents, there was also a simultaneous outbreak of 87 cases of Legionnaires' disease with 10 deaths in Genesse County, which includes Flint [1, 2]. Legionellosis, which includes Legionnaire's disease, demonstrates seasonal and geographic variability and is on the rise worldwide. Surveillance systems established to track legionellosis disease across Europe, North America, New Zealand, Australia and other countries are reporting increasing trends [3–6]. Between 1991 and 2012, the Australian Department of Health reported a two-thirds increase in legionellosis cases [6]. In the USA, during 2000–2014, crude incidence of legionellosis increased by approximately 300%, from 0.4 to 1.6 reported cases per 100 000 population [7, 8].

Legionnaires' disease and Pontiac fever are the two most common forms of legionellosis, a nationally notifiable disease in the USA. Legionnaires' disease causes severe pneumonia requiring hospitalisation, while Pontiac fever is generally presented as a flu-like illness and often resolves on its own. Thus, it is highly unlikely that hospitalisation records capture Pontiac fever cases. The Centers for Disease Control and Prevention (CDC) require reporting of all common forms of legionellosis, including Legionnaires' disease and Pontiac fever, but publicly available data do not distinguish between the two. To assess the epidemiology and increased incidence of legionellosis, the CDC recommend a combination of passive and active surveillance. There are two CDC national surveillance systems: the National Notifiable Diseases Surveillance System (NNDSS) established in 1976 and the Supplemental Legionnaires' Disease Surveillance System conducted from 2011 to 2015 [9, 10]. Outbreaks of legionellosis are reported through the Waterborne Disease Outbreak Surveillance System (WBDOSS), established in 1971. Monitoring of waterborne legionellosis outbreaks began in 2001. Since then legionella has been recognised as one of the most prominent aetiologic agents associated with waterborne outbreaks, being responsible for 66% of all outbreaks in 2011–2012 [11].

Although the specific causes for the increase in legionellosis cases are still unknown, potential factors include improved diagnosis and reporting [12], ageing of the national water infrastructure [13] and an ageing population [8]. Legionellosis occurs through inhalation of water aerosol contaminated by legionella, a pathogen that can live in organic matter attached to pipe inner surfaces [14, 15]. Amoebae growing in the water supplies act as the site of replication of legionella. However, it is not clear what occurs in plumbing sludge, since legionella can grow in

CAMBRIDGE
UNIVERSITY PRESS

biofilms that have a mixture of bacterial species. Water contamination occurs in the piping distribution systems of hot tubs, pools, showers, baths, humidifiers, mist and cooling equipment [16]. Older people (aged 50+), smokers and people with underlying medical conditions, such as diabetes and immune system disorders, are reported to be susceptible to infection [16]. Environmental risk factors include ageing water infrastructure and poor maintenance of complex water systems in hospitals and long-term care facilities [13, 14, 17]. A recent study indicates that the removal of sinks from patient rooms and introduction of a method of 'water-free' patient care has been associated with a significant reduction of patient colonisation with Gram-negative bacteria, especially in patients with a longer intensive care unit length of stay [18].

Meteorological factors are shown to be associated with the seasonal increase in disease incidence [19, 20]. Legionella bacteria thrives in a specific temperature range, 25–45 °C [14, 15]; and reports from surveillance of legionellosis point out the seasonal nature of infection. The peak of reported legionellosis cases usually occurs in late summer to fall, which is potentially linked to warmer, higher relative humidity and wetter weather conditions [8, 16, 19, 21–24]. Additional testing during the non-typical pneumonia season and the use of cooling towers during the summer are also offered as an explanation of seasonal rise [15]. Characterisation of the seasonality of legionellosis, while well documented, has been typically limited to a simple description of the month with high incidence. The highest number of cases were reported in June–October in the USA [8], July–August in Canada [5], August–November in Europe [3] and March–May and September–November in Australia and New Zealand [4, 6, 16, 25]. Characterisation of seasonal patterns, in terms of peak timing and amplitude, is essential to investigate the underlying reasons for changing epidemiology of infection. A shift in seasonal peak timing and amplitude points to a potential alteration in risk factors that affect disease incidence, including the probability of exposure and environmental triggers that promote pathogen growth. Our early work has demonstrated the potential to gain insight into the aetiology of ill-posed infections by quantifying their seasonal characteristics and comparing to the seasonality of well-documented infections [26].

The surge in legionellosis has been most noted since 2003 [27], which is alarming given the significant morbidity and mortality of legionellosis. Approximately 9% of legionellosis cases are fatal and 40% require intensive care [9]. Elderly patients (aged 65+) with pneumonia caused by *Legionella pneumophila* have a higher frequency of underlying comorbidities and thus high cost of care [28]. We determined that between 1997 and 2006, there were approximately 8000 hospitalisations due to Legionnaires' disease in US residents aged 65+, representing approximately $340 million in annual charges to Medicare [29]. Passive surveillance systems are believed to under-report disease incidence, especially of non-mandatory infections [8]. Our early work has illustrated an approach to link surveillance and hospitalisation records to assess the degree of agreement between two data sources [30]. By comparing the number of reported cases with the number of hospitalised cases for the same infection, we identified municipalities that significantly deviated from the typical experience in the state and might be under-ascertaining cases [30]. Since legionellosis is a nationally notifiable disease, cases reported to the national surveillance system should exhibit strong concordance with hospitalisation records. The Centers for Medicare and Medicaid Services (CMS) maintain records of all the hospitalised Medicare patients

nationwide and include information on approximately 96% of the US elderly population [31, 32].

The objectives of this analysis are to assess and compare the seasonal patterns of legionellosis in the USA using two national databases. We first created a monthly time series from 1993 to 2015 of cases of legionellosis using publicly available yearly reports from the CDC's NNDSS, and from 1997 to 2006 of medical claims of hospitalisation due to legionella in older adults from CMS databases. To examine the trend, seasonal peak timing and amplitude, we split the study time interval into two segments (before and after 2003), and applied a Poisson harmonic regression model to each data and each segment of the time series. Peak timing estimates and their confidence intervals were derived from the model regression parameters using the δ-method [33]. National surveillance provides continuous and ongoing data that enable a detailed analysis of changes in trend and seasonal patterns over time, and allow us to compare such patterns with trends observed in the highly susceptible elderly population.

Methods

CDC: surveillance records

We abstracted reported cases of legionellosis from the CDC's yearly NNDSS summary reports from 1993 to 2015. These publicly available reports feature the official statistics for notifiable diseases, and include counts, rates, age distribution by groups, and counts by state and territory. Monthly legionellosis disease counts are only available for the entire USA, while disease counts by age group are only available by year and not by month. Thus, we focused on disease counts available by month and year, and by year and age group.

We abstracted disease counts by month and created a monthly time series from 1993 to 2015. Year 2004 had information on total cases and reported age, but not on reported cases by month. We interpolated monthly counts for 2004 based on reported monthly cases of adjacent years (2003 and 2005), and on total 2004 cases. First, we averaged monthly disease counts for years 2003 and 2005 to obtain unadjusted estimates of monthly disease counts for 2004. Then, we calculated the adjusted monthly disease counts using equation 1:

$$Y_{i,\ 2004} = \frac{Y_{i,2004_{\mathrm{unadj}}}}{\sum_{i=1}^{12} Y_{i,2004_{\mathrm{unadj}}}} \times Y_{\mathrm{Total},2004}, \qquad (1)$$

where $Y_{i,2004}$ and $Y_{i,2004_{\mathrm{unadj}}}$ are the adjusted and unadjusted 2004 disease counts for month i, respectively; and $Y_{\mathrm{Total},2004}$ is the total number of CDC reported cases for 2004. We then rounded the adjusted 2004 disease counts to the nearest integer.

We used time series and calendar plots to explore disease count distribution by month (Figs 1 and 2). We used monthly proportions of reported cases for a given year to construct a calendar plot illustrating the variability of yearly reported cases. Disease counts were converted to monthly proportions for a given year using equation 2:

$$P_{i,j} = \frac{Y_{i,j}}{\sum_{i=1}^{i=12} Y_{i,j}} \times 100\ \%, \qquad (2)$$

where $P_{i,j}$ is the percentage of disease counts for each month i based on the total for year j, and $Y_{i,j}$ is the disease incidence for month i and year j.
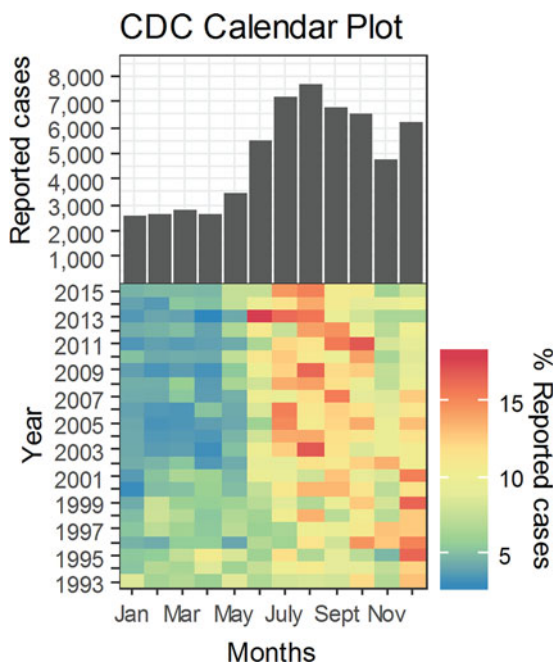
Fig. 1. Histogram and calendar plot of legionellosis cases reported to the Centers for Disease Control and Prevention from January 1993 to December 2015. Cool to warm colour scale represents a low to high scale of monthly per cent of reported cases. Low monthly percentage of yearly reported counts is represented by a dark blue tone with red symbolizing the other end of the scale.

To estimate the contribution of disease cases that occurred in older adults, we reviewed age distributions in publicly available CDC reports. Abstraction of disease counts by year and age group from 1993 to 2015 revealed inconsistencies with age grouping. Age categorisations contained 11 groups in 1993 and 1994: <1, 1–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–39, 40–49, 50–59, 60+. In 1995, age categories were re-classified as: <5, 5–14, 15–24, 25–44, 45–64, 65+, and a new category 'age not stated' was introduced. Starting in 1996, age categories were again re-classified as: <1, 1–4, 5–14, 15–24, 25–39, 40–64, 65+ and 'age not stated'. Table S1 shows the misalignment in age categories between time periods 1993–1994, 1995 and 1996–2015. We examined the contribution of each age category to the total number of reported cases of legionellosis for years that had consistent age categories (1996–2015). Age categories <1, 1–4, 5–14, 15–24 and 25–39 combined constituted no more than approximately 15.5% of the total reported cases and were thus aggregated to a single age category 0–39. Age categories 40–64 and 65+ accounted for most cases each year and were thus kept as their own categories. Figure 3 shows the yearly age distribution of reported legionellosis cases for the three age categories (0–39, 40–64 and 65+). By re-aligning the age categories, we determined the counts associated with older adults to enable proper comparison with CMS data. To calculate reporting rates in disease counts per 100 000 population, we used CDC yearly population estimates [34].

CMS: hospitalisation records

CMS is the national data repository of medical claims and offers a universal, uniform, and near-exhaustive coverage of US Medicare beneficiaries aged 65+. We abstracted hospitalisation

records from CMS using International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) code 482.84 reported in primary or secondary diagnostic fields during a 16-year period, from 1 January 1991 to 31 December 2006 [29]. Adoption of the 482.84 code for legionellosis was not widely used until mid-year 1997, with only one case of legionellosis being recorded between 1991 and 1997. The first hospitalisation record occurred in July 1997 and most likely does not reflect the true occurrence of legionellosis for this year but a transition in coding scheme. Thus, we only used records from January 1998 to December 2006 in this analysis. To construct the time series, we aggregated records of hospitalisation cases of legionellosis for beneficiaries 65+ by month of hospital admission. To calculate hospitalisation rates in disease counts per 100 000 population, we used CDC yearly population estimates for older adults (65+) [34].

To characterise the level of agreement between CMS hospitalisation cases and CDC reported cases for the age category of 65+, we calculated an index of reporting discrepancy (IRD) for each year using equation 3:

$$\text{IRD} = \log \frac{N_{\text{CDC\_65+}}}{N_{\text{CMS}}}, \tag{3}$$

where $N_{\text{CDC\_65+}}$ and N_{CMS} are the CDC and CMS cases, respectively. The index approaches 0 when the number of CDC and CMS cases are similar. A negative IRD indicates that reported CDC cases are lower than CMS cases; and when CDC cases exceed CMS cases, IRD values are positive.

Statistical model

Visualisation of the CDC time series showed two time periods with different patterns for trend and seasonality (Fig. 1). Trend and the amplitude of seasonal variations appear constant through December 2002; the increase in both trend and amplitude become notable starting January 2003. This apparent difference in trend marks two time segments, period 1 and period 2, best captured by a segmented regression model with the break point at 2003, in combination with harmonic terms describing the cyclic seasonal pattern of counts represented by the Poisson distribution. Thus, we defined the segmented Poisson harmonic regression model [35] in equation 4:

$$\begin{aligned}
\ln(E[Y(t)]) = {}& \beta_{\text{inter}} + \beta_{\text{trend}_{p1}} t_{p1} + \beta_{\cos_{p1}} \cos(2\pi\omega t_{p1}) \\
& + \beta_{\sin_{p1}} \sin(2\pi\omega t_{p1}) + \beta_{\text{trend}_{p2}} t_{p2} \\
& + \beta_{\cos_{p2}} \cos(2\pi\omega t_{p2}) + \beta_{\sin_{p2}} \sin(2\pi\omega t_{p2}),
\end{aligned} \tag{4}$$

where $Y(t)$ is the disease incidence at month t; β_{inter} is the intercept reflecting counts at the break point (T_{break}); β_{trend} are the coefficients for trend with t_{p1} and t_{p2} representing time in month before and after T_{break}, respectively; similarly $\beta_{\cos}$ and $\beta_{\sin}$ are the coefficients of the harmonic terms for each of two segments or time periods, *p1* and *p2*, respectively; and $\omega = 1/M$, where $M = 12$ is the length of the annual cycle in month.

We used equation 4 to model all available CDC data (1993–2015) and overlapping CDC and CMS data (1998–2006). Thus, we produced the three regression models: (1) CDC data from 1993 to 2015, (2) CDC data from 1998 to 2006 and (3) CMS data from 1998 to 2006.
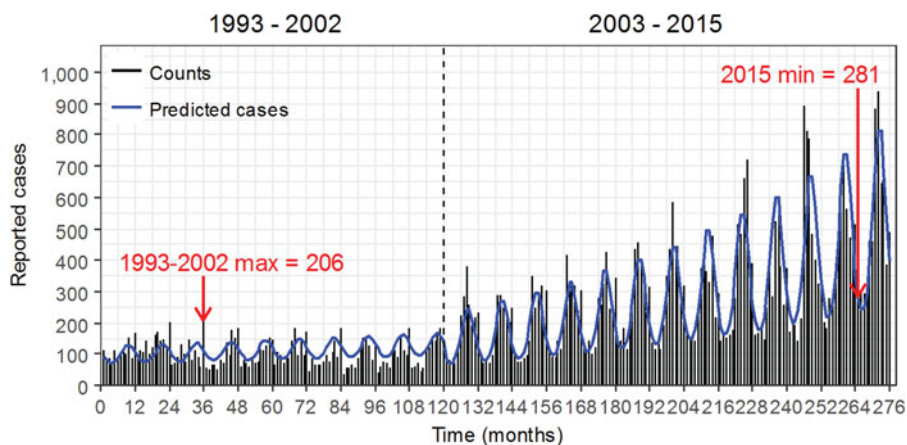
Fig. 2. Monthly time series of reported cases of legionellosis from the Centers for Disease Control and Prevention with model results superimposed as a solid blue line. The dashed vertical line at December 2002 marks two periods.
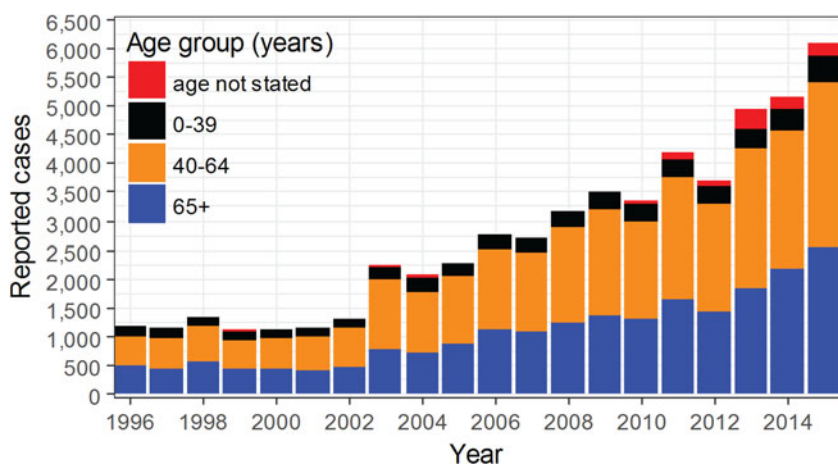
Fig. 3. Trend by age groups of reported legionellosis cases from the Centers for Disease Control and Prevention from years 1996 to 2015.

Calculation of peak timing and amplitude

Estimates for peak timing and its confidence interval were derived from the model regression parameters for periods 1 and 2 using the δ-method [33]. Table 1 shows the corresponding equations. Using the regression model and peak timing estimates, we calculated values for peak incidence in each period. Values for nadir were estimated with the time reference of 6 months before or after the peak timing estimate. Amplitude was calculated as the ratio between peak and nadir values for the beginning of the time series (T_{start}) and at the end of the time series (T_{end}). Amplitude for the breakpoint, T_{break}, was calculated as the average of the two amplitudes surrounding the breakpoint. This approach allowed us to compare the changes in amplitude from the beginning, through the breakpoint, to the end of the time series for each dataset.

Results

The monthly distribution of reported legionellosis cases from the CDC is shown in Figure 1 as a histogram aligned with the calendar plot through a shared axis. The histogram shows the monthly reported cases across the entire time period (1993–2015). The majority of cases were reported from June onward with the highest values in August and a distinct increase in December. The calendar plot shows the relative contribution of monthly cases to the total number of cases reported within a given year. Similarly to the histogram, the calendar plot shows the majority of cases

between June and October. Closer inspection of the calendar plot and monthly contribution of cases (Table 2) reveal that the frequency of reported cases in December was high in period 1 (1993–2002) and declined by 3.7% in period 2 (2003–2015). In period 2, most cases were reported in July and August. Between periods 1 and 2, the frequency of reported cases increased by 3.7% in July and 3.5% in August (Table 2). Consistently through the 21 years of data, January to May have the least number of cases of legionellosis.

Figure 2 shows time series of monthly legionellosis counts with the two distinct time periods indicated by the vertical dashed line. There is a clear change in trend and an amplitude increase in period 2 (2003–2015) compared with period 1 (1993–2002). Predicted cases derived from the regression model indicate peaks and valleys in both time periods with a steady trend and amplitude in period 1 and an increase in trend and amplitude in period 2.

Results from the regression models are shown in Table 3 as the estimates of per cent change in disease counts per month and peak timing for periods 1 and 2. Trend in reported cases is more pronounced in period 2 with an average 0.835% change in disease counts per month as compared with 0.258% for period 1. The average peak timing of reported cases shifted from mid-September (9.676 months, 95% CI 9.512–9.841) in period 1 to mid-August (8.452 months, 95% CI 8.410–8.495) in period 2 with more defined peaks as evidenced by narrower confidence intervals (Table 3, Fig. 5).

Table 1. Description of equations used in estimating the peak timing

Description	Equation
Phase shift, Ψ as a distance of peak from beginning of time series in radians	$-\arctan\left(\dfrac{\beta_{\sin}}{\beta_{\cos}}\right)$ (5)
Variance of phase shift, Var(Ψ)	$\dfrac{\{(\beta_{\cos}\times\sigma_{\beta_{\sin}})^2 + (\beta_{\sin}\times\sigma_{\beta_{\cos}})^2 - (2\times\beta_{\sin}\times\beta_{\cos}\times\sigma_{\beta_{\sin}\beta_{\cos}})^2\}}{(\beta_{\sin}^2 + \beta_{\cos}^2)^2}$ (6)
Confidence interval of phase shift in radians	$\Psi \pm \left(1.96\times\sqrt{\text{Var}(\Psi)}\right)$ (7)
Peak timing in months, where M is the length of one cycle	$M\times\left(-\dfrac{\Psi}{2\pi}\right),\qquad \beta_{\cos}>0 \text{ and } \beta_{\sin}>0,$ $M\times\left(1-\dfrac{\Psi}{2\pi}\right),\qquad \beta_{\cos}>0 \text{ and } \beta_{\sin}<0,$ (8) $M\times\left(\dfrac{1}{2}-\dfrac{\Psi}{2\pi}\right),\qquad \beta_{\cos}<0$
Confidence interval for peak timing in months	Substitute equation 7 for Ψ in equation 8

Table 2. Summary of monthly contributions of cases during two time periods

Month	1993–2002 Mean (s.d.)	2003–2015 Mean (s.d.)	δ	*p*-value*
Jan	4.937 (1.493)	4.158 (0.463)	−0.779	0.089
Feb	6.117 (1.330)	3.938 (0.528)	−2.179	**<0.001***
Mar	6.266 (0.770)	4.310 (0.858)	−1.956	**<0.001***
Apr	6.323 (1.997)	3.911 (0.688)	−2.411	**0.001***
May	6.155 (1.632)	5.549 (1.394)	−0.606	0.348
Jun	7.326 (1.307)	9.588 (3.005)	2.262	**0.038***
July	9.167 (1.865)	12.903 (2.321)	**3.736**	**<0.001***
Aug	10.124 (2.542)	13.635 (2.258)	**3.512**	**0.002***
Sept	9.798 (2.094)	12.542 (2.204)	2.745	**0.006***
Oct	11.077 (2.123)	11.312 (2.643)	0.235	0.821
Nov	8.973 (3.110)	8.085 (1.279)	−0.888	0.360
Dec	13.737 (2.142)	10.068 (1.840)	**−3.669**	**<0.001***

*Significant (by analysis of variance).

Table 3. Summary of trend and peak timing estimates for reported legionellosis cases by the CDC during 1993–2015

| Term	Estimate (95%CI) Period 1	Period 2
Trend as % change in disease counts per month	0.258 (0.223–0.293)	0.835 (0.817–0.854)
Peak timing in months	9.676 (9.512–9.841)	8.452 (8.410–8.495)
Peak timing	Sep 20 (Sep 15–25)	Aug 14 (Aug 13–15)
	CI-low: 0.164 (4.93d)	CI-low: 0.042 (1.31d)
	CI-high: 0.164 (4.93d)	CI-high: 0.042 (1.31d)

Table 4. Reported CDC cases and CMS hospitalisation cases by year and 65+ age category

| Period		Year	Total	CDC Rate of total per 100 000	65+	Rate of 65+ per 100 000	% 65+	Total	CMS Rate of total per 100 000	% $\frac{CMS}{CDC_{Total}}$	IRD*
Period 1		1993	1280	0.492							
		1994	1615	0.614							
		1995	1241	0.466	518	1.534	41.7				
		1996	1198	0.445	516	1.511	43.1				
		1997	1163	0.427	454	1.320	39.0	330	0.959	28.4	0.139
	Period 1a	1998	1355	0.491	569	1.644	42.0	1091	3.151	80.5	−0.283
		1999	1108	0.397	440	1.264	39.7	916	2.632	82.7	−0.318
		2000	1127	0.399	429	1.223	38.1	710	2.025	63.0	−0.219
		2001	1168	0.410	421	1.193	36.0	623	1.765	53.3	−0.170
		2002	1321	0.459	476	1.340	36.0	595	1.675	45.0	−0.097
Period 2	Period 2a	2003	2232	0.769	796	2.220	35.7	407	1.135	18.2	0.291
		2004	2093	0.715	728	2.011	34.8	325	0.898	15.5	0.350
		2005	2301	0.779	864	2.357	38.4	930	2.538	40.4	−0.022
		2006	2834	0.950	1111	2.989	39.2	993	2.672	36.9	0.049
		2007	2716	0.902	1082	2.860	39.8				
		2008	3181	1.046	1241	3.200	39.0				
		2009	3522	1.148	1386	3.498	39.4				
		2010	3346	1.082	1296	3.202	38.7				
		2011	4202	1.348	1646	3.979	39.2				
		2012	3688	1.174	1442	3.341	39.1				
		2013	4954	1.566	1844	4.127	37.2				
		2014	5166	1.620	2185	4.729	42.3				
		2015	6079	1.891	2559	5.358	42.1				
Total			58 890		22 023			6920			

We examined the contribution of each age category to the total number of reported cases of legionellosis for years that had consistent age categories (Table S1). For CDC 1996–2015 data, adults older than 40 years old (40–64 and 65+) represented the largest age group in reported cases for all years. The marked rise in reported cases after 2002 is consistently concentrated among individuals over 40 years old (Fig. 3).

Table 4 shows the annualised age distribution for CDC records for time period 1 (1993–2002) and period 2 (2003–2015), and for CDC and CMS records for the two overlapping time periods (1a: 1998–2002 and 2a: 2003–2006). Legionellosis cases identified in hospital settings exceed the number of cases reported to the CDC for the age category of 65+ in all years except for 1997, 2003, 2004 and 2006. This lower reporting of CDC 65+ cases as compared with CMS cases is reflected in the negative IRD values, which range from −0.022 to −0.318. Since 1998, there has been a decrease in negative numbers with IRD becoming positive in 2003, 2004 and 2006. From 1998 to 2002, the ratio of CMS hospitalisation cases to total CDC legionellosis cases ranged from 45.0% to 82.7%. In 2003 and 2004, this ratio was at its minimum (<20%), but rose again in 2005 with a slight decrease in 2006.

Similarly to the CDC data, CMS records exhibited a change in the trend and seasonal patterns of legionellosis between the two time periods. CMS counts show a downward trend from 1998 to 2002 and an upward trend starting in 2003 (Fig. 4). The downward trend is also captured in model results with a negative per cent change in legionellosis counts per month (Table 5). In period 1a, 1998–2002, the average monthly counts of hospitalisation due to legionellosis declined by 1.925%. The peak timing was in late September (9.921 months), although the confidence interval ranged from late August to early November (95% CI 8.774–11.067). CDC data for period 1a also showed a peak timing in mid-September (9.397 months, 95% CI 9.206–9.589). In period 2a, we see better synchronisation of upward trends and peak timing between CMS and CDC records: both CMS and CDC records peak in August, 8.383 and 8.820 months, respectively, although the confidence interval of the peak timing for hospitalisation records is wider. Figure 5 summarises peak timing along with their confidence intervals for CDC periods 1 and 2, and CDC and CMS periods 1a and 2a. The widest confidence interval corresponds to period 1a hospitalisation records. Confidence intervals of peak timing for CDC records are narrower than for
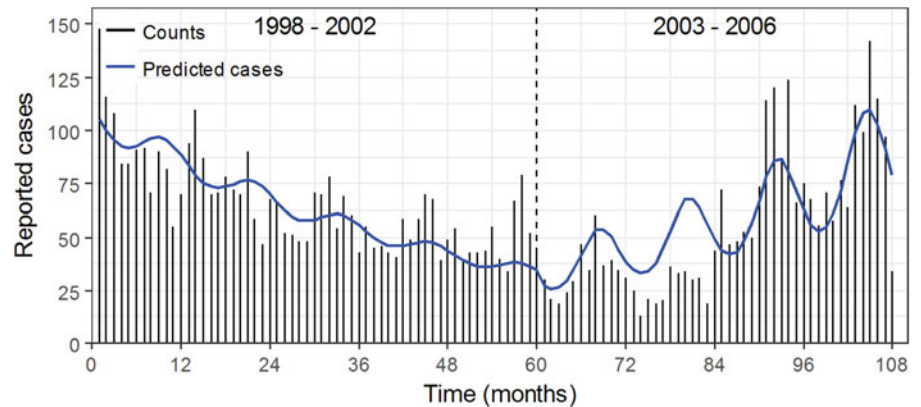
Fig. 4. Monthly time series of reported cases of legionellosis in older adults (65+) from the Centers for Medicare and Medicaid from January 1998 to December 2006 with model results superimposed. The dashed line at month 60 or December 2002 separates periods 1a and 2a.
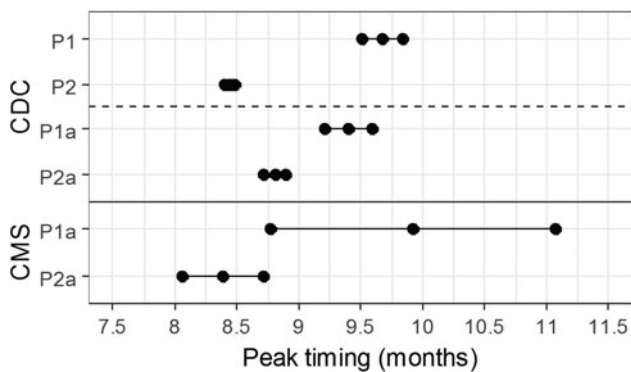
Fig. 5. Peak timing estimates for pre- and post-periods for three models. The *k* timing estimates for pre- and post-periods for three models.

CMS records in all time periods. For both data sources, confidence intervals become narrower in the second time period (2 and 2a). Models for CDC data explained approximately 81% of data variability (85% for the 1993–2015 model and 78% for the 1998–2006 model). Variability explained for the CMS model was approximately 60%.

Table 6 shows the predicted annual legionellosis counts and rates per 100 000 population, and the amplitude for the beginning and end of each time period as well as the breakpoint in 2003 for CDC records (1993–2015), and for years that had overlapping CDC and CMS records (1998–2006). The predicted disease counts and rates for CDC records, representing general population, show a sharp increase between 2003 and 2015 with cases and rates increasing by approximately 255% (1670–5925 counts) and 219% (0.578–1.843 counts per 100 000 population), respectively. Amplitude shows an increase of 97% (1.771–3.482) between 1993 and 2015. Amplitude also shows an increase between the beginning and the end of years that have overlapping CMS and CDC records; between 1998 and 2006, amplitude increased by 95% (2.082–4.054) for CDC reported cases and by 100% (1.034–2.070) for CMS hospitalisations due to legionellosis in older adults.

Discussion

Our study clearly indicates a notable shift of seasonal peaks in legionellosis from mid-September (9.676 ± 0.164 months) before 2003 to mid-August (8.452 ± 0.042 months) after 2003 along with a substantial increase in seasonal amplitude of 97% between 1993 and 2015. The shift in seasonal peaks before and after 2003 was also observed between 1998 and 2006 along with a seasonal amplitude increase of 100% in US older adults and 95% in the general population. While the seasonality of legionellosis has been documented, characterisation of seasonality has been typically limited to a simple description of the month with high incidence, which is often inconsistent, imprecise and lacks ability to formally compare seasonal variations. We applied an approach that extracts information on seasonality in a rigorous and systematic manner and enabled us to estimate seasonal characteristics, such peak timing and amplitude, and to compare the results across time periods and affected populations. The shift in disease peak towards mid-August is indicative of changes in epidemiology of disease transmission and must be communicated to health professionals, physicians, researchers, public, and policy and decision makers.

The study confirms that the incidence of legionellosis has been increasing in the USA since 2003. The rapid increase began in 2003 and continued through 2015 to reach 6079 cases and a rate of 1.89 cases per 1 000 000 population, a 360% and 312% increase, respectively. Reported cases of legionellosis to the CDC remained stable between 1993 and 2002 with an average of approximately 1260 cases reported each year at an average rate of 0.46 cases per 100 000 population. Our modelling predicts approximately 1551 cases in 2002 at a rate of 0.539 per 100 000 population, and approximately 5925 cases in 2015 at a rate of 1.843 per 100 000 population. This corresponds to an increase of approximately 282% in predicted cases and 242% in predicted rates.

The steady increase starting in 2003 observed across all age groups has been attributed to changes in test type and test performance for pneumonia-like diseases following the outbreak of SARS in 2003 [27]. Improved surveillance has been noted as a potential reason for the continued increase of legionellosis incidence across Europe [19, 36]. The Australian Department of Health has documented a sharp increase in reported legionellosis cases, from 122 to 382, between 1991 and 2012 [6]. Monitoring of waterborne legionellosis outbreaks began in the USA in 2001, which given a 1 year implementation window, could have resulted in the increase in cases reported to the CDC in 2003.

However, changes in seasonal characteristics of legionellosis are less likely to be the result of only improved reporting. Cases of legionellosis reported to the CDC showed the expected seasonal pattern of high incidence in late summer and fall and low incidence in winter, but peak timing, a fundamental characteristic

Table 5. Summary of trend and peak timing estimates for the reported legionellosis cases by the CDC and for the hospitalisation cases of older adults (65+) due to legionellosis from CMS during 1998–2006

	CMS		CDC	
Term	Period 1a	Period 2a	Period 1a	Period 2a
Trend as % change in disease counts per month	−1.925 (−2.077 to −1.771)	2.004 (1.796–2.212)	0.574 (0.458–0.690)	1.311 (1.189–1.433)
Peak timing in months	9.921 (8.774–11.067)	8.383 (8.053–8.712)	9.397 (9.206–9.589)	8.820 (8.734–8.906)
Peak timing	Sep 28 (Aug 24–Nov 2)	Aug 12 (Aug 2–22)	Sep 12 (Sep 6–18)	Aug 25 (Aug 23–28)
	CI-low: 1.146 (34.38d)	CI-low: 0.330 (10.22d)	CI-low: 0.191 (5.74d)	CI-low: 0.086 (2.67d)
	CI-high: 1.146 (34.38d)	CI-high: 0.330 (10.22d)	CI-high: 0.191 (5.74d)	CI-high: 0.086 (2.67d)

Table 6. Predicted disease counts with rates, and amplitude at the beginning (T_{start}), breakpoint (T_{break}) and end (T_{end}) for each of two time periods

		Predicted annual counts (rates per 100 000)[b]			Amplitude		
Years T_{start}–T_{end}	Data source[a]	T_{start}	T_{break}	T_{end}	T_{start}	T_{break}	T_{end}
1993–2015	CDC	1174 (0.452)	1670 (0.578)	5925 (1.843)	11.769	2.627	3.482
1998–2006	CDC	1107 (0.401)	1620 (0.561)	2852 (0.956)	2.082	3.068	4.054
	CMS	1142 (3.298)	462 (1.294)	970 (2.610)	1.034	1.552	2.070

[a]CDC data include the entire population and CMS data only older adults (65+).
[b]To estimate reporting rates per 100 000, the total US population was used for CDC rates, and the population of older adults (65+) was used for CMS rates.

of seasonality, changed. The peak timing shifted by a month starting in 2003 from mid-September to mid-August, which exceeds the typical legionellosis incubation period of 1–10 days [14]. Table 2 shows that the highest number of cases was observed in December before 2003 and in August after 2003. We examined whether the shift in peak timing was solely due to the reduced number of cases in December as part of a sensitivity analysis (Table S2). We repeated the peak estimation for time series with December counts removed, and confirmed a peak in August for both time periods. This result suggests that the shift in peak timing from mid-September to mid-August after 2003 is driven primarily by the lower number of reported cases in December. Finding records of legionellosis outbreaks in December prior to 2003 proved challenging.

The observed shift in peak timing could be related to environmental drivers and summer travel. Fluctuations in seasonal peaks (September in 2011, October in 2012 and August in 2013) were observed in a study of surveillance records of legionellosis in England and Wales potentially due to weather conditions [23, 37]. In 2013, approximately 31% of the reported cases of Legionnaire's diseases in England and Wales were associated with travel abroad, with India having the highest incident rate (7.56 cases per million visits), followed by Malta (6.54 cases per million visits) [37]. According to the World Health Organization (WHO), approximately 20% of the cases of legionellosis detected in Europe are travel-related [14]. In the USA between 1980 and 1998, approximately 21% of cases of Legionnaire's disease from the CDC's passive surveillance system were travel-related [38]. The study unfortunately did not provide information destination and time of travel, and the publicly available CDC data do not have enough information to analyse such patterns properly. A systematic analysis of outbreaks and incidence of legionellosis is needed to determine the reasons for the shift peak timing on refined spatial and temporal scales.

There needs to be more attention to the shift in seasonal peaks to better understand underlying environmental drivers. Over time, we observed that the seasonal peaks became stronger: amplitude almost doubled and became narrower with an overall duration of about 2 months (Fig. 5). Alarmingly, the minimum monthly cases of legionellosis reported to the CDC in 2015 was 281, which exceeds the maximum monthly cases reported any time before 2003. These changes in seasonality of reported legionellosis suggest underlying causes that go beyond increased reporting. Changes in environmental factors that promote pathogen growth could be one of the causes. The Mid-Atlantic region experienced a sharp increase in legionellosis between May and September of 2003 along with record-breaking rainfall: the average monthly rainfall was 54% higher than historical averages for those months. In five Mid-Atlantic States, a 1 cm increase in rainfall was associated with a 2.6% increase in legionellosis incidence between 1990 and 2003 [22]. An increase in temperature and rainfall as well as higher relative humidity have been associated with higher risk of Legionnaires' disease [19, 39]. These environmental factors could boost proliferation, transmission and potentially pathogen mutation. Shen *et al.*, highlighted complex interactions among *L. pneumophila*, disinfectant residual and biofilms. Biofilms that develop under long-term disinfected conditions provide less protection for *L. pneumophila* as compared with those that develop under disinfectant-free conditions [40]. The efficacy of disinfectants depends on environmental conditions that often change seasonally or due to infrastructure failures. During the Flint water crisis, as the concentration of free chlorine in water delivered to Flint residents decreased, the risk of acquiring Legionnaire's disease increased by 80% per 1 mg/L decrease in free chlorine [41]. Better understanding of the mechanisms that increase exposure and transmission of *L. pneumophila* might lead to opportunities for disease control and prevention.

With the growing segment of ageing and highly susceptible population, the observed trends of legionellosis are alarming

due to its high morbidity and mortality [16]. While age composition of reported cases of legionellosis to the CDC remained mostly consistent over time, the increase in incidence since 2003 has not been the same for all age categories. Between 1996 and 2015, people 40 years and older had the largest disease burden, with older adults (65+) representing approximately 39% of all reported cases. Yet, most of the increase in incidence has been observed in people 40 years and older. Due to inconsistent age categorisation, we only analysed data after 1996. Frequent changes in age categorisation in CDC data (prior to 1996) create unnecessary confusion, which can be prevented by using a granular age scale for publicly available reports. Furthermore, the detailed analysis of age distribution at granular scale, enabled by new methodology most applicable for big data, allows researchers to early detect changes in patient profiles while still maintaining confidentiality [29, 32].

The increase in legionellosis since 2003 has been well documented, but public reports of disease trends rarely include information of quality of model fit, proper test statistics and interpretation of modelling results [6, 11]. The increasing amplitude of seasonal peaks in reported legionellosis is startling. Yet, the assessment of seasonality of legionellosis in existing surveillance systems has been especially limited to oversimplified information on months with highest and lowest number of records, which mask temporal variations. Organisations responsible for surveillance, like the CDC in the USA, should strive for data harmonisation and proper reporting of seasonal characteristics of disease. Proper reporting will enable reliable comparisons across locations and times, and improve the analysis of temporal variations to enable forecasting of disease incidence, age distribution and emerging trends.

CMS hospitalisation records show somewhat similar patterns to CDC surveillance records with an increase in legionellosis incidence starting in 2003. From 2003 to 2006, predicted cases of legionellosis increased by approximately 110% (462–970 cases) and predicted rates by approximately 102% (1.295–2.610 cases per 100 000 population). Peak timing was not apparent prior to 2003, but then became more defined with narrower confidence intervals. CMS legionellosis cases declined during 1998–2002, which is inconsistent with the flat reporting depicted by the CDC records. This declining trend could be the result of changes in Legionnaire's disease diagnostics. Between 1980 and 1998, diagnosis by urine antigen testing increased from 0% to 69% while diagnosis by serologic testing decreased – the frequency of isolates other than *L. pneumophila* serogroup 1 decreased from 38% to 4% [38]. The authors of that study warned that the decrease in culture-based diagnosis limits the recognition of disease cases associated with other isolates and impairs outbreak investigation. CMS records exhibited a dip in cases in 2003 and 2004 with a noticeable lower number of predicted legionellosis cases as compared with reported cases (Fig. 4). These discrepancies are attributed to incomplete records in the available CMS dataset for 2003 and 2004, which is likely to represent a fraction of actual cases. While the similarities between seasonal patterns in legionellosis depicted by CMS and CDC provide strong support for the detected patterns, improving detection methods and quality of reporting is needed to better characterise seasonal change in pathogen ecology.

Worldwide, the primary cause of infection is inhalation of aerosols or aspiration of water contaminated with legionella bacteria, primarily *L. pneumophila*. Potting soil and compost contaminated with *Legionella longbeachae* also pose a risk for human infection, particularly in Australia and New Zealand [42]. A legionellosis outbreak associated with industrial cooling towers in New Zealand showed that nine out of 13 cases had evidence of infection with either *L. pneumophila* serogroup 1 or *L. longbeachae* serogroup 1 [43]. Since both organisms were also isolated from the cooling towers, waterborne transmission of *L. longbeachae* was suggested but remains unproven. The authors of the study recommend not discounting waterborne transmission in investigations of *L. longbeachae* infection.

Comparison of CDC reporting for people 65 years and older with CMS hospitalisation cases reveals that in 6 out of 10 years (1997–2006) CMS records exceeded the CDC records by 46–522 cases, indicating potential under-reporting of legionellosis to the CDC. Inspection of the age-not-stated category in CDC shows that the differences between CMS and CDC 65+ were not misplaced in the age-not-stated category. This discrepancy in the number of cases between CDC and CMS suggests that not all cases of legionellosis diagnosed by the hospital-based laboratories are reported to the CDC. Those discrepancies have been decreasing since 1998, suggesting that reporting consistency across agencies is improving.

Conclusion

The development of predictive strategies for a complex infection such as legionellosis requires better understanding of microbial ecology in water systems, improved reporting of legionellosis and detailed analysis of seasonal changes. With the growing volume of electronic records, consistently and systematically collected over long time periods, the understanding and prediction of legionellosis trends and temporal variations could be substantially improved. Accurate characterisation of changes in legionellosis trend and seasonality and timely synchronisation and harmonisation of records collected by various agencies are essential to strengthen disease monitoring and inform potential interventions in a meaningful way.

References

1. **Michigan Department of Health and Human Services**. Genesee County Public Health Department. Legionellosis Outbreak-Genesee County, June, 2014 – March, 2015 Summary Analysis. 2016.
2. **Michigan Department of Health and Human Services**. Genesee County Public Health Department. Legionellosis Outbreak-Genesee County, May, 2015 – November, 2015 Summary Analysis. 2016.
3. **European Centre for Disease Prevention and Control**. Legionnaires' disease in Europe, 2014. 2016.
4. **Graham FF** *et al.* (2012) Changing epidemiological trends of legionellosis in New Zealand, 1979–2009. *Epidemiology and Infection* **140**, 1481–1496.

5. **Ontario Agency for Health Protection and Promotion (Public Health Ontario)**. Epidemiology of legionellosis in Ontario, 2013. Surveillance period: January 1, 2013 to December 31, 2013. Toronto, ON, 2014.

6. **NNDSS Annual Report Writing Group** (2015) Australia's notifiable disease status, 2012: annual report of the National Notifiable Diseases Surveillance System. *Communicable Diseases Intelligence* **39**, E46–E136.

7. **Adams DA** *et al.* (2016) Summary of notifiable infectious diseases and conditions – United States, 2014. *MMWR Morbidity and Mortality Weekly Report* **63**, 1–152.

8. **Centers for Disease Control and Prevention (CDC)** (2011) Legionellosis – United States, 2000–2009. *MMWR Morbidity and Mortality Weekly Report* **60**, 1083–1086.

9. **Dooling KL** *et al.* (2015) Active bacterial core surveillance for legionellosis – United States, 2011–2013. *MMWR Morbidity and Mortality Weekly Report* **64**, 1190–1193.

10. **Centers for Disease Control and Prevention (CDC)**. Active Bacterial Core surveillance (ABCs). Background. Available at https://www.cdc.gov/abcs/overview/background.html (Accessed 26 February 2018).

11. **Beer KD** *et al.* (2015) Surveillance for waterborne disease outbreaks associated with drinking water – United States, 2011–2012. *MMWR Morbidity and Mortality Weekly Report* **64**, 842–848.

12. **Farnham A** *et al.* (2014) Legionnaires' disease incidence and risk factors, New York, New York, USA, 2002–2011. *Emerging Infectious Diseases* **20**, 1795–1802.

13. **Garrison LE** *et al.* (2016) Vital signs: deficiencies in environmental control identified in outbreaks of Legionnaires' disease – North America, 2000–2014. *MMWR Morbidity and Mortality Weekly Report* **65**, 576–584.

14. **Bartram J** *et al.* (eds.) (2007) *Legionella and the Prevention of Legionellosis.* Geneva: World Health Organization.

15. **Fields BS, Benson RF and Besser RE** (2002) Legionella and Legionnaire's disease: 25 years of investigation. *Clinical Microbioly Reviews* **15**, 506–526.

16. **Phin N** *et al.* (2014) Epidemiology and clinical management of Legionnaires' disease. *The Lancet Infectious Diseases* **14**, 1011–1021.

17. **Soda EA** *et al.* (2017) Vital signs: health care-associated Legionnaires' disease surveillance data from 20 states and a large metropolitan area – United States, 2015. *MMWR Morbidity and Mortality Weekly Report* **66**, 584–589.

18. **Hopman J** *et al.* (2017) Reduced rate of intensive care unit acquired gram-negative bacilli after removal of sinks and introduction of 'water-free' patient care. *Antimicrobial Resistance and Infection Control* **6**, 59.

19. **Beaute J** *et al.* (2016) Short-term effects of atmospheric pressure, temperature, and rainfall on notification rate of community-acquired Legionnaires' disease in four European countries. *Epidemiology and Infection* **144**, 3483–3493.

20. **Cunha BA, Burillo A and Bouza E** (2016) Legionnaires' disease. *The Lancet* **387**, 376–385.

21. **Fisman DN** (2007) Seasonality of infectious diseases. *Annual Review of Public Health* **28**, 127–143.

22. **Hicks LA** *et al.* (2007) Increased rainfall is associated with increased risk for legionellosis. *Epidemiology and Infection* **135**, 811–817.

23. **Ricketts KD** *et al.* (2009) Weather patterns and Legionnaires' disease: a meteorological study. *Epidemiology and Infection* **137**, 1003–1012.

24. **Simmering JE** *et al.* (2017) Weather-dependent risk for Legionnaires' disease, United States. *Emerging Infectious Diseases* **23**, 1843–1851.

25. **Li JS, O'Brien ED and Guest C** (2002) A review of national legionellosis surveillance in Australia, 1991 to 2000. *Communicable Diseases Intelligence* **26**, 461–468.

26. **Chui KK** *et al.* (2009) Geographic variations and temporal trends of *Salmonella*-associated hospitalization in the U.S. elderly, 1991–2004: a time series analysis of the impact of HACCP regulation. *BMC Public Health* **9**, 447.

27. **Neil K and Berkelman R** (2008) Increasing incidence of legionellosis in the United States, 1990–2005: changing epidemiologic trends. *Clinical Infectious Diseases* **47**, 591–599.

28. **Sopena N** *et al.* (2007) Community-acquired *Legionella* pneumonia in elderly patients: characteristics and outcome. *Journal of the American Geriatrics Society* **55**, 114–119.

29. **Naumova EN** *et al.* (2016) Hospitalizations due to selected infections caused by opportunistic premise plumbing pathogens (OPPP) and reported drug resistance in the United States older adult population in 1991–2006. *Journal of Public Health Policy* **37**, 500–513.

30. **Mor SM** *et al.* (2009) Cryptosporidiosis in the elderly population of the United States. *Clinical Infectious Diseases* **48**, 698–705.

31. **Fisher ES** *et al.* (1990) Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. *American Journal of Public Health* **80**, 1487–1490.

32. **Cohen SA, Chui KK and Naumova EN** (2011) Measuring disease burden in the older population using the slope-intercept method for population log-linear estimation (SIMPLE). *Statistics in Medicine* **30**, 480–488.

33. **Naumova EN and MacNeill IB** (2006) Seasonality assessment for biosurveillance systems. In: Auget JL, Balakrishnan N, Mesbah M and Molenberghs G (eds). *Advances in Statistical Methods for the Health Sciences.* Boston, MA: Birkhauser, pp 437–450.

34. **United States Department of Health and Human Services (USDHHS), Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS)**. Bridged-Race Population Estimates, United States July 1st resident population by state, county, age, sex, bridged-race, and Hispanic origin. Compiled from 1990–2015. Available on CDC WONDER Online Database.

35. **Chandy SJ** *et al.* (2014) The impact of policy guidelines on hospital antibiotic use over a decade: a segmented time series analysis. *PLoS ONE* **9**, e92206.

36. **Rota MC** *et al.* (2013) Legionnaires' disease in Italy: results of the epidemiological surveillance from 2000 to 2011. *Eurosurveillance* **18**, pii=20497.

37. **Naik FC and Phin N** (2014) Legionnaires' disease in England and Wales 20133. London, UK.

38. **Benin AL, Benson RF and Besser RE** (2002) Trends in Legionnaires disease, 1980–1998: declining mortality and new patterns of diagnosis. *Clinical Infectious Diseases* **35**, 1039–1046.

39. **Fisman DN** *et al.* (2005) It's not the heat, it's the humidity: wet weather increases legionellosis risk in the greater Philadelphia metropolitan area. *Journal of Infectious Diseases* **192**, 2066–2073.

40. **Shen Y** *et al.* (2017) Effect of disinfectant exposure on *Legionella pneumophila* associated with simulated drinking water biofilms: release, inactivation, and infectivity. *Environmental Science & Technology* **51**, 2087–2095.

41. **Zahran S** *et al.* (2018) Assessment of the Legionnaires' disease outbreak in Flint, Michigan. *PNAS* **115**, E1730–E1739.

42. **Whiley H and Bentham R** (2011) *Legionella longbeachae* and legionellosis. *Emerging Infectious Diseases* **17**, 579–583.

43. **Thornley CN** *et al.* (2017) *Legionella longbeachae* detected in an industrial cooling tower linked to a legionellosis outbreak, New Zealand, 2015; possible waterborne transmission? *Epidemiology and Infection* **145**, 2382–2389.