

Modeling Legionnaires' Disease Outbreaks

Estimating the Timing of an Aerosolized Release Using Symptom-onset Dates

Joseph R. Egan, Ian M. Hall, David J. Lemon, and Steve Leach

Background: Over the last 30 years, there have been a number of reported Legionnaires' disease outbreaks resulting from the release of causative organisms from aerosol-producing devices.

Methods: We model a Legionnaires' disease epidemic curve as the convolution of an infection-time distribution (representing the aerosolized release) and an incubation-period distribution. The model is fitted to symptom-onset data from specific outbreaks to estimate the start and end dates of the release. We also develop this retrospective "back-calculation" model into a prospective "real-time" model that can estimate the final size of an ongoing outbreak, in addition to the timing of its release.

Results: In the retrospective analysis, the estimated release end dates were generally earlier than reported end dates. This suggests that, in many outbreaks, the release might have already ended by the time the source was reportedly cleaned or closed. Prospective analysis showed that valid estimates of the release start date could be achieved early in the outbreak, the total number of cases could be reasonably determined shortly after the release had ended, and estimates of the release end date could be satisfactorily achieved in the latter stages of the outbreak.

Conclusions: This model could be used in the course of a Legionnaires' disease outbreak to provide early estimates of the total number of cases, thus helping to inform public-health planning. Toward the end of the outbreak, estimates of the release end date could help corroborate standard epidemiologic, environmental, and microbiologic investigations that seek to identify the source.

(*Epidemiology* 2011;22: 188–198)

Submitted 7 May 2010; accepted 21 September 2010.

From the Microbial Risk Assessment, Emergency Response Department, Health Protection Agency, Porton Down, Salisbury, Wiltshire, United Kingdom.

Supported by the Department of Health for England (Health Protection Agency grant numbers 104307, 104308, 103504).

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Joseph R. Egan, Microbial Risk Assessment, Emergency Response Department, Health Protection Agency, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom. E-mail: joseph.egan@hpa.org.uk

Copyright © 2011 by Lippincott Williams & Wilkins

ISSN: 1044-3983/11/2202-0188

DOI: 10.1097/EDE.0b013e31820937c6

Legionnaires' disease is a form of pneumonia caused by the inhalation of *Legionella pneumophila* bacteria suspended in aerosolized water.¹ Symptoms include fever, nonproductive cough, muscle aches, and headaches, with suggested treatments including macrolide and fluoroquinolone antibiotics.^{1,2} In recent outbreaks, approximately 70% of cases required hospitalization^{3–6} (with a range of about 40%–90%^{7–10}), with overall case-fatality rates of about 5%^{3,5,6,8,10} (with a range of 0%–10%^{4,7,9}). Legionnaires' disease was first identified in 1976 following a common-source outbreak at an American Legion meeting in Philadelphia, PA.^{11,12} Since the discovery of Legionnaires' disease, there have been many community-acquired outbreaks linked to environmental sources such as cooling towers and whirlpool spas.^{3–10,13–24} Colonization of such aerosol-producing systems by *L. pneumophila* can result from a lack of regular or effective disinfection.¹ A Legionnaires' disease outbreak is usually halted following closure or thorough cleaning of the source, commonly identified via epidemiologic, environmental, and microbiologic investigations. However, if the travel histories of cases during the 2–10 days before symptom onset² (an approximation for the largest likely incubation period) provide little evidence for a specific source, then often a "brute-force" tactic is adopted, with many potential aerosol-producing systems cleaned or closed. This can take days or even weeks to complete.³ Samples taken from potential sources before the implementation of any preventive methods can provide evidence of *L. pneumophila* contamination, especially when matched to clinical isolates.^{4,13} In this study, we model symptom-onset dates of cases in an attempt to corroborate whether the end of an aerosolized release coincided with any reported decontamination procedures.

An increase in cases reporting pneumonia-like illness is likely to be the first sign of an *L. pneumophila* aerosolized release,⁴ following which epidemiologic, environmental, and microbiologic evidence is used to define a "community outbreak."² The date of symptom onset for each case is usually known, although the date of infection is often much less clear or unknown. In this study, the "release" is defined to be the period of time during which the source is releasing contaminating aerosol; it is assumed to be equal to the period during which cases are infected. We propose that cases' incubation periods and infection dates conform to theoretical probability distributions that can be convolved to form a probability distribution of

cases' symptom-onset dates. We test this model against publicly available Legionnaires' disease outbreak data (the authors extracted the data from published graphs for the majority of outbreaks), for which there is reasonable evidence of source identification and its cleaning or closure date. In addition to retrospectively analyzing outbreaks, the model is developed to provide real-time estimates of the release timing and outbreak size. Such early estimates of the total number of cases could benefit public-health planning (eg, gauging the required number of hospital beds). Our model is an extension of those developed for bioterrorist anthrax attacks,^{25,26} where the release is considered instantaneous rather than continuous.

METHODS

Retrospective Analysis at the End of an Outbreak Infection Times

In April 2000, Legionnaires' disease was confirmed for 125 people who visited or passed within 500 m of the Melbourne Aquarium, Australia.⁵ The histogram in Figure 1A shows infection dates for 121 cases, made possible because the majority of cases were in the vicinity of the aquarium only once in the weeks before symptom onset (G. Tallis, written communication, March 2006). Although the majority of infection dates for the Melbourne outbreak were known, in many Legionnaires' disease outbreaks the time of infection for each individual cannot be explicitly identified when there is potential for multiple exposure times. We therefore assume the most parsimonious process for the time-dependent hazard of infection, in which a release with explicit start and end times gives equal probability to an individual becoming infected at any time during the

release. This is described by a uniform probability density function, termed "uniform release window," with parameters x and y representing the start and end dates, respectively (see the fit in Fig. 1A for an example).

Incubation Periods

During the Melbourne outbreak, incubation periods were obtained for 114 cases (Fig. 1B histogram, G. Tallis, written communication, March 2006). The incubation period was calculated by subtracting the date of visit, or the date when cases were known to have passed by the aquarium, from the symptom-onset date. Two-parameter gamma, log-normal, and Weibull probability density functions were fitted to the Melbourne dataset, using the "mle" function in the R stats4 package.²⁷ The maximized log-likelihood for the gamma, log-normal, and Weibull probability density functions were -272.66 , -272.83 , and -278.27 , respectively. Therefore, on the basis of the models that were analyzed, the best fit to the data was provided by the gamma probability density function with shape parameter $a = 4.96$ (95% confidence interval = 3.82 – 6.32) and scale parameter $b = 1.27$ (95% confidence interval = 0.99 – 1.68), termed "gamma incubation period" (Fig. 1B fit). A χ^2 goodness-of-fit test confirmed the quality of the fit ($P = 0.75$, 7 degrees of freedom). Incubation periods were similarly identified for 136 cases during a Legionnaires' disease outbreak in Bovenkarspel, Netherlands.⁹ Unfortunately, we were unable to obtain the raw data from the Bovenkarspel outbreak, which would have allowed a statistical comparison with the Melbourne dataset; however, the sample medians were close (6 and 7 days for Melbourne and Bovenkarspel, respectively). In addition, a Le-

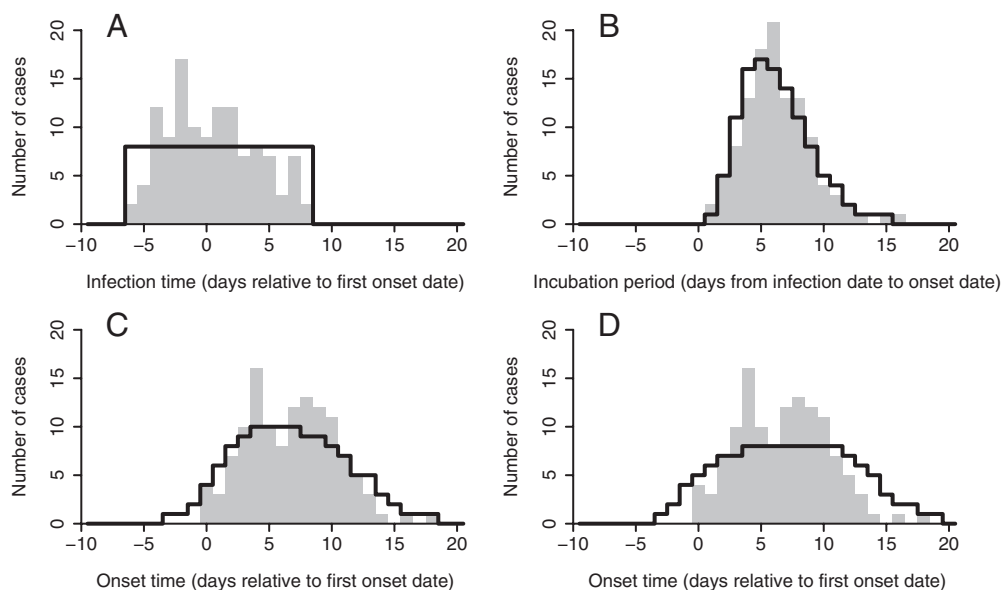
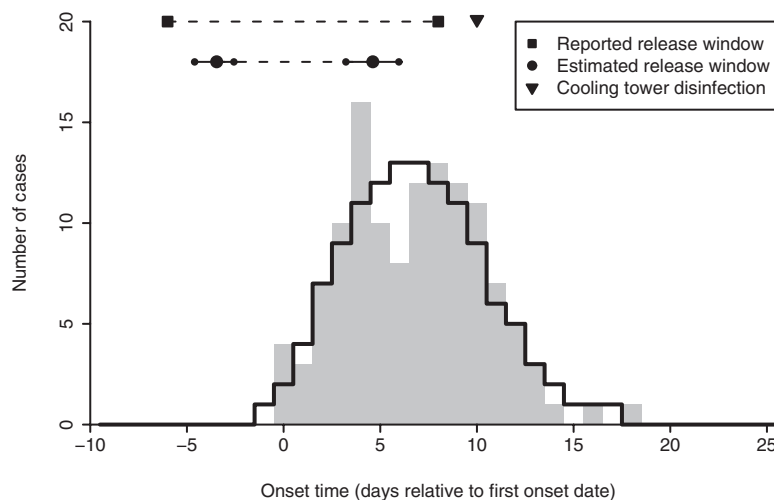


FIGURE 1. Melbourne outbreak histograms (gray) and model fits (black, rounded to the nearest integer). A, Uniform release window fit. B, Gamma incubation period fit. C, Histogram convolution. D, Uniform release window and gamma incubation period convolution fit.

FIGURE 2. Melbourne outbreak epidemic curve (gray) and back-calculation model fit (black, rounded to the nearest integer), assuming a uniform release window. The large squares and circles represent the reported and estimated uniform release window, respectively. The small circles represent the 95% confidence intervals for both the estimated release start and end dates. The triangle represents the date of the cooling tower's (ie, probable source's) disinfection.



gionnaires' disease outbreak in Miyazaki, Japan²⁸ yielded data on 76 incubation periods; reassuringly, a Kolmogorov-Smirnov test revealed no evidence against this dataset being drawn from the gamma incubation period ($P = 0.51$).

Onset Times

In the situation where individual infection dates and incubation periods are known, their summation yields individual symptom-onset dates. The reported epidemic curve (histograms in Fig. 1C, D) can be approximated by convolving the histograms of the infection dates and incubation periods (Fig. 1C fit). A second approximation, and the basis of our approach, would be by convolving the distributions fitted to the data displayed in the histograms, that is, convolving a uniform release window with the gamma incubation period (Fig. 1D fit). In the Melbourne example, the number of available infection dates (121) and incubation periods (114) are a subset of the number of available symptom-onset dates (124), thus incorporating a small error in both approximations. Because symptom-onset dates are usually reported during an outbreak, we reverse the latter convolution process and back-calculate the start and end dates of a uniform release window (x and y), assuming that the gamma incubation period parameters (a and b) remain valid across all outbreaks. A mathematical description of this "back-calculation model" is provided in eAppendix 1 (<http://links.lww.com/EDE/A451>). In their statistical analysis of the HIV/AIDS pandemic, Brookmeyer and Gail²⁹ first proposed the use of a time-dependent infection rate model, via a convolution with the incubation-period distribution, to estimate characteristics of the infection source.

Prospective Analysis During an Outbreak

The total number of cases becomes known shortly after the end of a Legionnaires' disease outbreak. However, during the course of the outbreak, its size is yet to be established. Therefore, real-time estimates of the start and end dates of a release need to be considered in parallel with estimates of the

final outbreak size. This is a more complex scenario than an instantaneous release of, say, anthrax, where the 2 key parameters of interest are the time at which the release occurred and the eventual size of the outbreak. Not only does the continuous-release scenario require an additional (time) parameter, but early estimates might be needed before the release has even ended. Using symptom-onset times of the early cases, the best-fitting set of release parameters can be calculated. A mathematical description of this "real-time model" is provided in eAppendix 2 (<http://links.lww.com/EDE/A451>). It should be noted that the general real-time model captures the special-case back-calculation model as the outbreak comes to an end.

RESULTS

Retrospective Analysis at the End of an Outbreak Reported Outbreak Data

Figure 2 shows the back-calculation model fit to the Melbourne⁵ epidemic curve and the estimated uniform release window. A χ^2 goodness-of-fit test revealed no statistically significant differences between the reported and estimated symptom-onset dates ($P = 0.70$, 10 degrees of freedom). The release is estimated to start 3 days before the first symptom-onset date, and to end 8 days later. The approximate 95% confidence interval was $(-1-1)$ day for both the start- and end-date estimates. Although the back-calculation model captures 88% of the known infection dates, 6 cases were infected in the 2 days preceding the lower confidence interval of the estimated start date and 9 cases were exposed in the 2 days after the upper confidence interval of the estimated end date. Intuitively, to reproduce a close fit to the epidemic curve, the back-calculation model is forced to over- and underestimate the start and end dates, respectively (compare Fig. 2 and Fig. 1D).

The Table provides a summary of estimates for 18 other Legionnaires' disease outbreaks.^{3,4,6-10,13-24} Note that epi-

TABLE. Reported and Estimated Outbreak Release Parameters

Outbreak Identifier	Outbreak Location	No. Cases	Reported First Symptom Onset Date ^a	Reported End Date ^b (Days)	x ^c (Uniform Estimated Start) ^c (95% CI)	x ₁ (Logistic Estimated Location) ^d (95% CI)	x ₂ (Logistic Estimated Scale) ^e (95% CI)	y ^f (Uniform Estimated End) ^f (95% CI)	y ^g (Logistic Estimated End) ^g (95% CI)	Difference Uniform End (Days) ^h	Difference Logistic End (Days) ^h	df Uniform	df Logistic	P ⁱ Uniform	P ^j Logistic	AIC Uniform	AIC Logistic
A ⁴	Murcia, Spain	413	26/06/01	12	0 (−1 to 0)	4 (1 to 37)	1 (0 to 2)	7 (6 to 7)	5 (3 to 6)	−5	−7	13	11	<0.05	<0.05	2162.6	2121.2
B ⁸	Miyazaki, Japan	272	24/06/02	30	−1 (−2 to 0)	8 (5 to 10)	3 (1 to 4)	32 (30 to 33)	30 (28 to 31)	2	0	31	24	<0.05	0.22	2023.6	1958.4
C ⁹	Bovenkarspel, Netherlands	182	25/02/99	3	−2 (−3 to −1)	−1 (−2 to 37)	1 (0 to 2)	5 (4 to 6)	5 (2 to 5)	2	2	10	9	<0.05	<0.05	952.3	950.0
D ¹⁶	Bovenkarspel, Netherlands	151	25/02/99	4	−2 (−3 to −2)	−2 (−3 to 14)	1 (0 to 2)	6 (5 to 7)	6 (3 to 6)	2	2	10	9	0.07	0.05	811.6	812.8
E ⁷	Pamplona, Spain	146	26/05/06	7	−2 (−3 to 0)	1 (−1 to 32)	0 (0 to 1)	−2 (−3 to 0)	−2 (−2 to 1)	−9	−9	7	6	0.71	0.54	689.3	691.5
F ⁵	Melbourne, Australia	124	17/04/00	8	−3 (−5 to −3)	−3 (−5 to 11)	0 (0 to 2)	5 (3 to 6)	5 (1 to 6)	−3	−3	10	9	0.70	0.51	667.5	669.6
G ¹³	Barrow, United Kingdom	114	12/07/02	20	−1 (−2 to 0)	41 (10 to 100)	4 (2 to 5)	17 (15 to 18)	14 (12 to 15)	−3	−6	14	8	<0.05	0.37	707.4	656.1
H ³	Cerdanyola, Spain	114	15/07/02	32	−1 (−3 to 0)	13 (10 to 23)	3 (1 to 5)	30 (28 to 32)	27 (23 to 29)	−2	−5	13	13	<0.05	<0.05	820.3	748.5
I ⁶	Genova, Italy	98	01/08/95	19	−3 (−5 to −2)	4 (1 to 49)	3 (1 to 5)	20 (17 to 23)	17 (11 to 19)	1	−2	10	10	<0.05	<0.05	673.1	656.8
J ¹⁰	Kapellen, Belgium	93	04/11/99	4	−2 (−3 to −1)	1 (−3 to 37)	2 (0 to 3)	5 (4 to 7)	4 (2 to 6)	1	0	8	6	<0.05	<0.05	479.5	482.1
K ¹⁷	Stafford, United Kingdom	62	07/04/85	13	−2 (−3 to −1)	2 (−1 to 43)	2 (0 to 4)	12 (10 to 14)	11 (8 to 11)	−1	−2	5	5	<0.05	0.14	364.1	359.68
L ¹⁴	Sarpsborg, Norway	56	12/05/05	11	−2 (−6 to 0)	−3 (−4 to 32)	1 (0 to 2)	−2 (−6 to 0)	−1 (−2 to 2)	−13	−12	4	3	0.06	0.06	298.2	300.38
M ¹⁸	Barcelona, Spain	54	31/10/00	14	−3 (−5 to −2)	−3 (−5 to 36)	1 (0 to 4)	5 (3 to 7)	5 (1 to 7)	−9	−9	4	3	0.09	<0.05	289.8	292.1
N ¹⁹	Vic-Gurb, Spain	54	14/10/06	16	−5 (−8 to −3)	−3 (−8 to 5)	2 (0 to 6)	22 (19 to 26)	22 (18 to 26)	6	6	6	4	0.33	0.65	387.9	389.3
O ²⁰	Amsterdam, Netherlands	31	27/06/06	14	−4 (−6 to −2)	−4 (−7 to 35)	0 (0 to 4)	5 (1 to 10)	5 (0 to 9)	−9	−9	2	1	0.10	<0.05	197.0	198.8
P ²¹	Zaragoza, Spain	29	02/06/04	13	−4 (−7 to −2)	−3 (−7 to 43)	1 (0 to 8)	14 (11 to 18)	14 (10 to 18)	1	1	1	1	0.59	0.25	188.6	190.7
Q ²⁴	Torrevecchia, Spain	28	26/11/05	11	−2 (−5 to 0)	−2 (−4 to 33)	0 (0 to 3)	−1 (−5 to 2)	0 (−2 to 2)	−12	−11	1	3	<0.05	<0.05	137.2	139.2
R ¹⁵	Sheboygan, United States	28	10/08/86	17	−2 (−7 to 0)	−3 (−6 to 33)	0 (0 to 3)	−2 (−7 to 6)	0 (−2 to 6)	−19	−17	1	3	0.09	0.14	165.7	166.6
S ²²	Hereford, United Kingdom	27	08/10/03	35	−4 (−9 to −1)	21 (14 to 98)	5 (2 to 9)	34 (29 to 41)	29 (25 to 34)	−1	−6	1	4	<0.05	0.06	208.6	189.9
T ²³	Christiansburg, United States	22	29/09/96	10	−4 (−7 to −2)	42 (3 to 79)	4 (1 to 8)	12 (9 to 17)	9 (7 to 12)	2	−1	1	1	<0.05	0.07	140.4	135.0

^aDate format: DD/MM/YY.^bCorresponds with cleaning or closure of the source reported in the literature relative to reported first symptom-onset date.^cMaximum likelihood estimate is relative to reported first symptom-onset date. 95% confidence intervals are shown in parentheses as whole days relative to the maximum likelihood estimate.^dSee eAppendix 1 (<http://links.lww.com/EDE/A451>) for definition of the location and scale parameters.^eDifference between the estimated end and the reported end.^fP represents the P value of a χ^2 test between the reported and modeled epidemic curve.^gdf indicates degrees of freedom; AIC, Akaike Information Criterion.

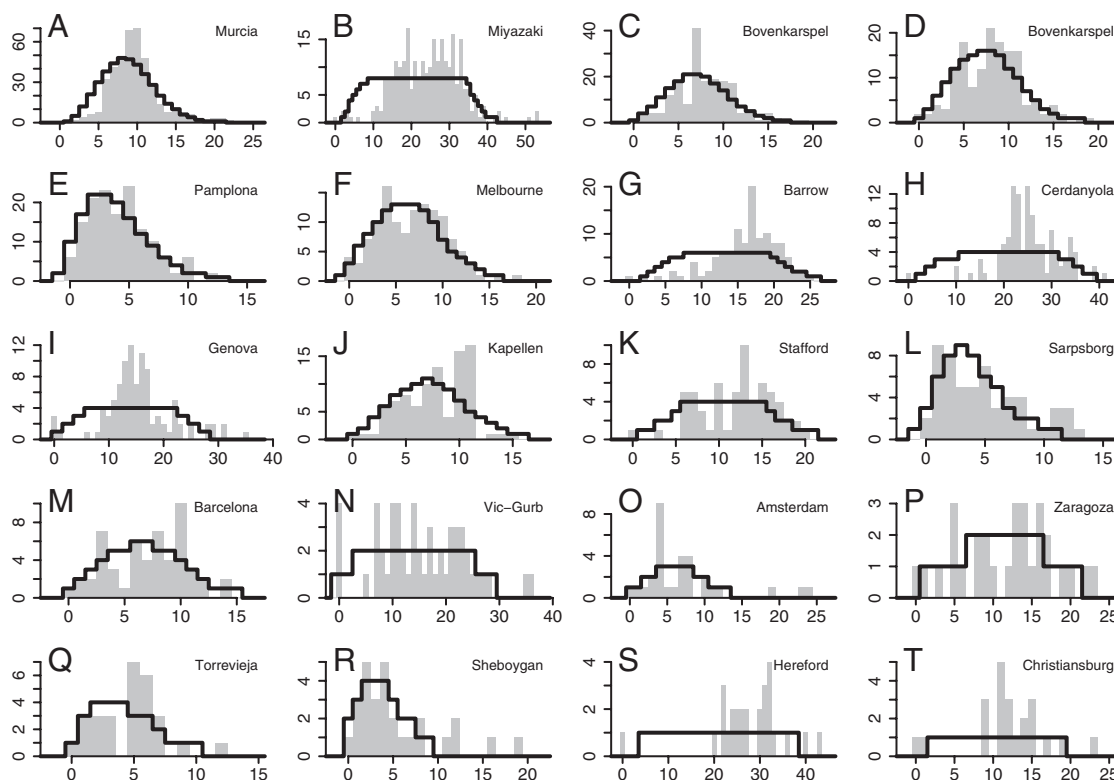


FIGURE 3. Epidemic curves (gray) and back-calculation model fits (black), assuming a uniform release window for the outbreaks described in the Table. X-axis represents symptom-onset time (days relative to first onset date) and y-axis represents the number of cases (rounded to the nearest integer for the fits). Note the differing scales among the various outbreaks.

demographic curves C and D relate to the same outbreak (see below). Nine of the 20 modeled epidemic curves provided either reasonable (D, L, M, O, and R: $0.05 \leq P < 0.20^{30}$) or good (E, F, N, and P: $P \geq 0.20$) fits to the data, based on a χ^2 goodness-of-fit test. The other 11 fits were poor ($P < 0.05$). A visual comparison provided in Figure 3 shows that 8 of these 11 fits (panels A, B, G, H, I, K, S, and T) consistently overestimate the early period of the outbreak and underestimate the peak. There are 3 possible explanations for this behavior—first, that very early cases are sporadic cases unrelated to the outbreak in question; second, that following outbreak detection, active case finding increases ascertainment of later cases; or third, that the assumption of a uniform release window is too simplistic. The third explanation is perhaps the most likely and is considered further in the next section. The poor fits in panels J and Q similarly coincided with the peak being underestimated, although the early stages of the outbreak are captured well. The final poor fit (panel C) is a result of day 8 of the outbreak having a particularly high number of cases when compared with the previous and subsequent days. An earlier report of the same outbreak provides an epidemic curve with approximately 30 fewer cases; the back-calculation model actually provides a better fit to this truncated dataset (panel D, $P = 0.07$, 10 degrees of

freedom). One speculative (but unconfirmed) explanation for the discrepancy between the 2 reported epidemic curves is that the additional 30 cases were those for whom information was unobtainable and were, therefore, all allocated the same symptom-onset date.

The estimated end date is earlier than the reported end date in 6 of the 9 reasonable or good epidemic curve fits, and later than the reported end date in only 3. This suggests that in many outbreaks, the release might have already ended by the time the source had been reportedly cleaned or closed. For 3 of these outbreaks (E, L, and R), the estimated start and end dates are the same, suggesting an instantaneous release of organisms. Indeed, particularly short releases that ended before the reported source cleaning or closure have been hypothesized for these outbreaks,^{7,14,15} but such a hypothesis has lacked the supporting statistical evidence provided in this study.

Sensitivity Analysis

To account for the discrepancies between the reported and modeled epidemic curves, we considered a slightly more complex release of organisms. The 2-parameter uniform release window was replaced by a 3-parameter “logistic release window” that allowed more flexibility in the early stages of the outbreak (eAppendix 1, <http://links.lww.com/EDE/A451>, for a

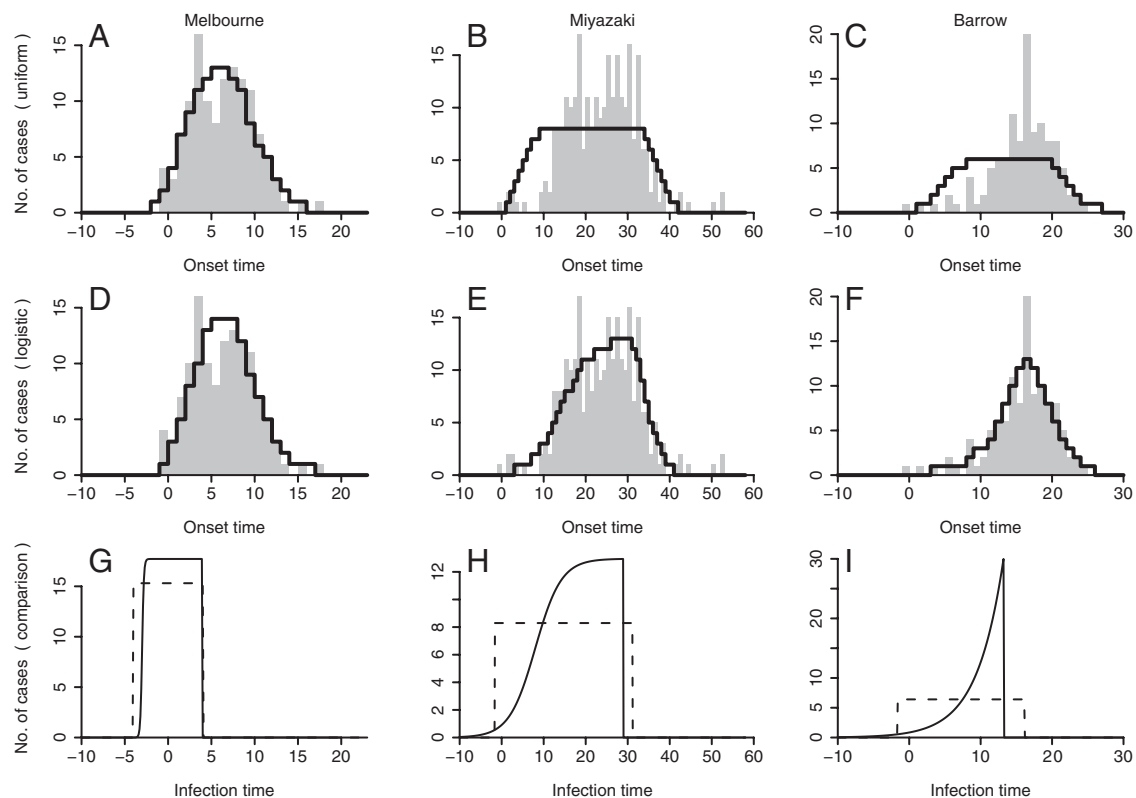


FIGURE 4. Epidemic curves (gray) and back-calculation model fits (black, rounded to the nearest integer) assuming a uniform release window (A–C) and a logistic release window (D–F). Uniform (dashed) and logistic (un-dashed) release windows (G–I) corresponding to the back-calculation model fits. Melbourne (A, D, and G), Miyazaki (B, E, and H), and Barrow (C, F, and I) outbreaks. Note the differing scales among the various outbreaks.

mathematical description). We hypothesized that exponential growth of *L. pneumophila* during the early stage of the release window, followed by a period during which the number of organisms remained at a stable carrying capacity (ie, the maximum sustainable population), and ending abruptly by source cleaning or closure (or some other factor), might provide biologic justification for such a release window.

A logistic release window improves the fits of the 2 smallest outbreaks from poor to reasonable (Table), although the 3 early cases in these outbreaks (particularly the Hereford case) could possibly be sporadic. More convincing are the Miyazaki and Barrow modeled epidemic curves, where a logistic release window improves the fits from poor to good (Table). Figure 4 shows that the estimated logistic release window of the Miyazaki outbreak had reached a stable carrying capacity before the source was cleaned or closed, but for Barrow, the estimated logistic release window was still in the growth phase. This suggests that the outbreak might have been much larger without public-health intervention.

Of the remaining 7 originally poor fits, only the Stafford outbreak was improved to a reasonable fit by incorporating a logistic release window. In addition, a comparison of Akaike's Information Criterion (AIC) values in the Table

shows that none of the 9 originally reasonable or good fits benefitted sufficiently from a logistic release window to justify the extra parameter. As an example, Figure 4 shows that the 2 modeled release windows for Melbourne are very similar, indicating that a uniform release window is sufficient to capture the epidemic curve. This analysis suggests that, although a logistic release window can occasionally provide a better fit, a uniform release window is more justifiable for the majority of Legionnaires' disease outbreaks. Finally, it should be noted that the estimated end dates were earlier with a logistic release window for the 5 improved fits, providing further evidence that the release might have ended before the reported cleaning or closure of the source.

Prospective Analysis During an Outbreak

Simulation Study

We first simulated an outbreak of 100 cases resulting from a 10-day release to analyze the performance of the real-time model before testing it with reported data. Infection times of the cases were randomly generated from a uniform release window and then added to deviates from the gamma incubation period (parameterized in the Methods) to provide symptom-onset times. Note that in this simulation, the start

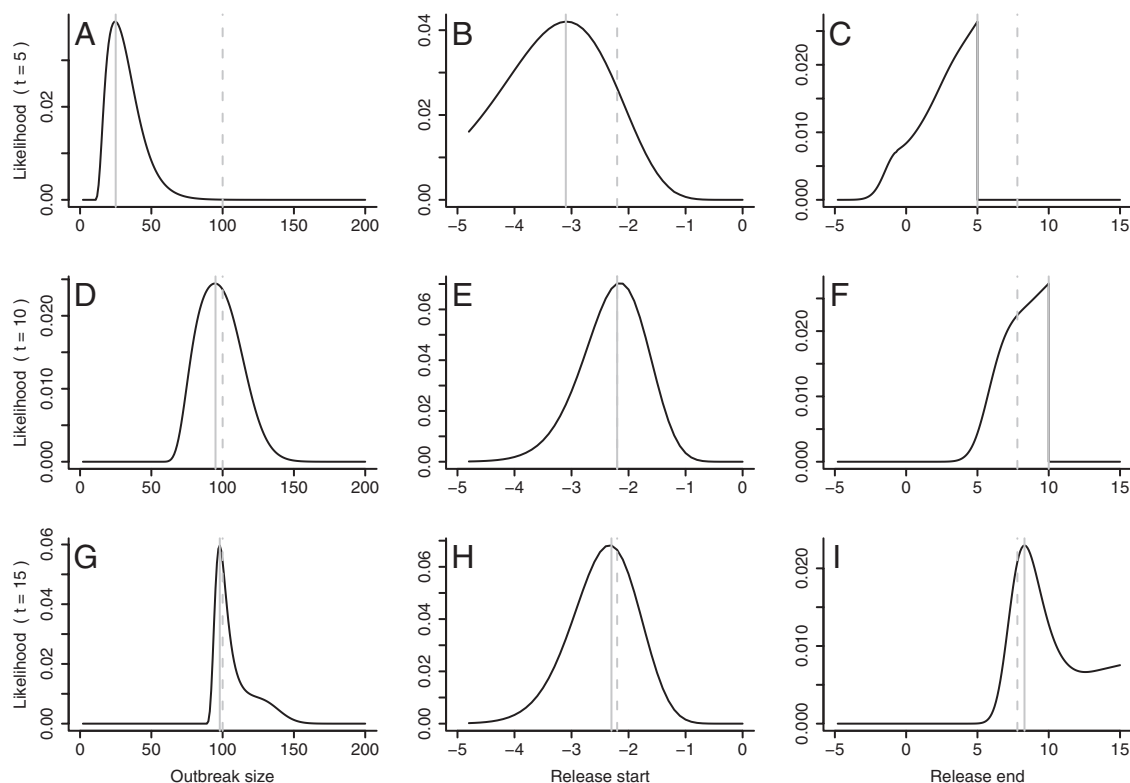


FIGURE 5. One SIMULATION. Real-time model likelihood profiles of the model parameters (A–C) 5 days, (D–F) 10 days, and (G–I) 15 days after symptom onset of the first case ($t = 0$). Gray lines represent actual (dashed) and maximum likelihood estimate (un-dashed) values. The simulation generates 100 cases over a 10-day uniform release window.

and end dates were given by ($x = -2.2$, $y = 7.8$), corresponding with the first symptom-onset time at $t = 0$. Figure 5B shows that 5 days after the appearance of the first symptomatic case (at which point 13 cases had displayed symptoms), the release start date was overestimated by approximately 1 day. The release was estimated to end on the day that the estimates were being made (Fig. 5C), correctly suggesting that the release was still ongoing. This forced an underestimate of the final outbreak size (Fig. 5A). After 10 days of case data (when 58 cases had displayed symptoms), the start date was accurately estimated (Fig. 5E). At this point, the release had ended and the outbreak size estimate moved very close to the actual value. However, the real-time model mistakenly suggested that the release was still ongoing (Fig. 5F). By the 15th day of case data (at which time 91 symptomatic cases had occurred), the real-time model had received sufficient data to accurately capture the end date (Fig. 5I). Also note how the likelihood profile of the size estimate is much narrower around the actual value, highlighting the decreasing uncertainty in this parameter as the end of the outbreak is approached.

The panels in Figure 5 represent the results from just 1 simulation. To capture the general behavior of the real-time model, we simulated 500 outbreaks, as above (100 cases over

a 10-day uniform release window). The outbreak size estimates showed a pattern similar to Figure 5, in that underestimates necessarily occurred while the release was ongoing (Fig. 6A), but quickly focused on the actual size once the release had ended (Fig. 6D, G). Also following a similar pattern to Figure 5, the estimated start dates were generally accurate even after only a few days of data (Fig. 6B, E, H). However, the estimated end dates were less predictable—before the release had ended, the real-time model correctly estimated that the release was ongoing in approximately two-thirds of simulations, but incorrectly estimated that the release had ended in the remaining one-third (Fig. 6C). Shortly after the release had ended, the real-time model estimated that the release was still ongoing in approximately two-thirds of simulations and underestimated the end date in the remaining one-third (Fig. 6D). Only after a few more days of case data did the real-time model consistently capture the simulated end date (Fig. 6I).

Reported Outbreak Data

We applied the real-time model to truncated Legionnaires' disease outbreak data from Melbourne, where a uniform release window was shown earlier to provide a good fit. As a misspecification, we also tested the real-time model

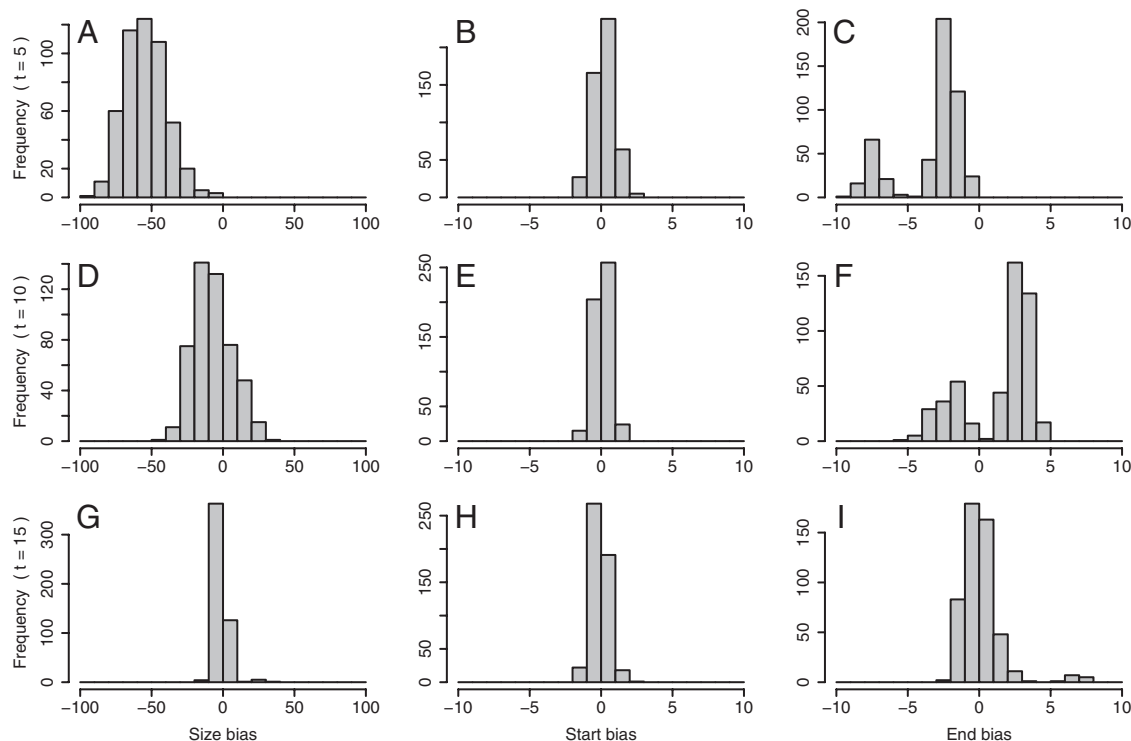


FIGURE 6. Five-hundred SIMULATIONS. Real-time model histograms of the model parameters bias (A–C) 5 days, (D–F) 10 days, and (G–I) 15 days after symptom onset of the first case. Each simulation generates 100 cases over a 10-day uniform release window. Bias is defined as the actual value minus the estimated value. Mean (SD) number of cases after 5 days = 22.9 (6.3), 10 days = 67.2 (8.0), and 15 days = 94.5 (2.8).

against the Barrow data, which was shown earlier to provide a poor fit with a uniform release window (but a good fit with a logistic release window). The most striking differences between the inferences of the 2 outbreaks were the confidence intervals of the size and start date estimates (Fig. 7). For Melbourne, the final outbreak size confidence intervals were fairly wide in the middle stages of the outbreak, although this was not the case for Barrow. In contrast, there was large uncertainty in the estimated start date for the Barrow outbreak, whereas the start of the release for Melbourne was captured early with fairly narrow confidence intervals. Both of these features can be explained by the low numbers of cases in the early stages of the Barrow outbreak. First, the real-time model is forced to infer a small outbreak because a large outbreak would be expected to “take-off” more quickly. Second, low case numbers provide only a “weak signal” to infer the start of the release, resulting in such uncertain estimates. Strategies to overcome these problems are addressed in the Discussion section.

DISCUSSION

Retrospective Analysis at the End of an Outbreak

Our study began with an analysis of the incubation period of Legionnaires' disease based on 114 cases from an outbreak in

Melbourne. A gamma probability density function provided a good fit to the data, giving a mean (SD) of 6.3 (2.8) days. The closely related *Francisella tularensis* and *Coxiella burnetii*³¹ (the causative agents of Tularaemia and Q fever, respectively) show evidence of a dose-dependent incubation period.^{32–34} However, we could find no experimental human data for *L. pneumophila*, and the results from guinea pig experiments were inconclusive with regard to the dose-dependence of the incubation period.^{35–38} Therefore, given the closely agreeing median incubation periods from 2 other Legionnaires' disease outbreaks, we believe that our derived gamma incubation period is a reasonable model for any Legionnaires' disease outbreak incubation-period distribution.

We modeled the epidemic curve resulting from a continuous *L. pneumophila* release by convolving the gamma incubation period with a uniform release window (representing the outbreak-dependent infection dates). Symptom-onset data were used to fit the start and end dates of the release for a number of outbreaks. We found that the estimated end date was generally earlier than the reported end date, suggesting that in many outbreaks, the release might have ended before the reported cleaning or closure of the source. Perhaps, this was because media alerts of the outbreak resulted in maintenance alterations of the responsible aerosol-producing device, or because meteorologic conditions that were favorable to the

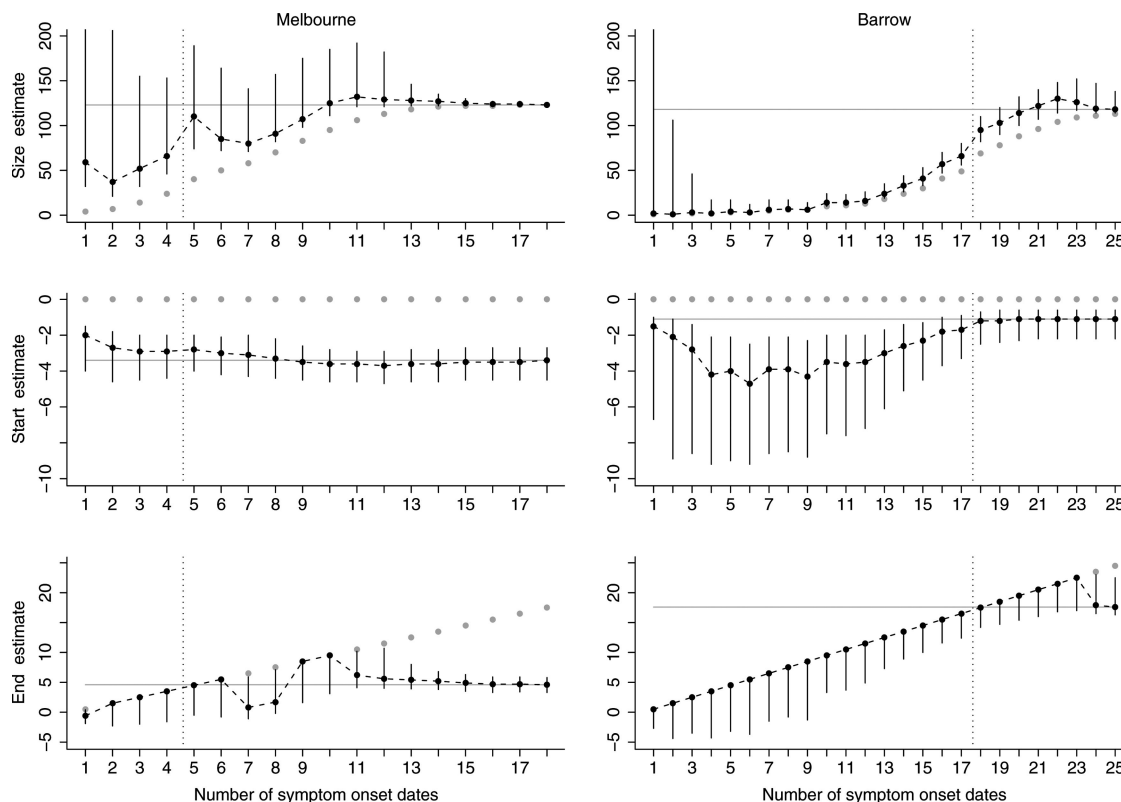


FIGURE 7. Real-time model time series of the release parameter estimates (black points) and 95% confidence intervals (vertical black lines) against the final estimated values (horizontal gray lines) for the Melbourne (left panels) and Barrow (right panels) outbreaks, assuming a uniform release window. In the top panel, gray points represent lower bounds; in the middle and bottom panels, gray points represent upper bounds. Dotted vertical black lines represent final estimated end date following the first symptom-onset date. The lower bound on the size estimate is given by the reported number of cases that have become symptomatic by that day. The upper bound on the release start date is given by the first symptom-onset date. The upper bound on the release end date is the date on which the estimate is made.

dispersion of *L. pneumophila* organisms suddenly changed. Indeed, this latter explanation had been previously suggested for the outbreak in Sheboygan, which our analysis corroborated as an approximate instantaneous release. It seems possible that a “burst” of *L. pneumophila* organisms might have coincided with brief but favorable meteorologic conditions in Sheboygan,¹⁵ and also for the 2 other outbreaks whose releases were estimated to be of a very short duration.^{7,14}

The uniform release window provided reasonable or good epidemic curve fits in 45% of the outbreaks analyzed, whereas the logistic release window provided reasonable or good fits in 60%. However, a more complex logistic release window proved to be a case of overfitting (ie, unjustifiable complexity) in the majority of outbreaks. Thirty percent of the outbreaks were fitted poorly by both uniform and logistic release windows—likely a reflection of the vast number of confounding effects (eg, variations in meteorologic conditions, population densities, age- and health-related susceptibilities, etc) that accompany a Legionnaires’ disease outbreak. It is perhaps inevitable that some Legionnaires’

disease epidemic curves will not conform to simple models. Still, it is promising that the majority of outbreaks were captured with such parsimonious models, given all of the associated variability.

Prospective Analysis During an Outbreak

For our simulation study, we assumed a uniform release window both for the simulated data and for the estimation of the release parameters. The real-time model gave accurate estimates of the start of the release after only a few days of symptom-onset data. The size of the outbreak was unavoidably underestimated while the release was ongoing, but accurate estimates followed shortly after the end of the release and continued to improve with time (and thus with more data). Both shortly before and shortly after the release had ended, estimates of the end date tended to suggest that the release was still ongoing. Such estimates improved only toward the end of the outbreak, when sufficient data had been gathered. These conclusions held for varying size outbreaks and varying release durations (results not shown). During an

actual outbreak, a simple calculation of $n/(y - x)$, where n represents the outbreak size estimate, would give an estimate of the daily number of infections during the release. Therefore, given an outbreak where the release was estimated to be ongoing, it would be possible to inform public-health planners with an estimate of the total number of cases, if the release was to end in the next 1, 2, 3... days. A forecast of when such cases would be likely to display symptoms could also be provided.

We found results similar to the simulation study when applying the real-time model to symptom-onset data from the Melbourne outbreak. However, testing the real-time model with data from the Barrow outbreak (shown to have more likely resulted from a logistic release window) substantially underestimated the outbreak size, with misleading confidence. A naive solution to this problem would be to use the current real-time model and ignore the very early cases once it was clear that an outbreak was "taking-off." A more robust solution might be to replace a uniform release window (embedded within the real-time model) with a logistic release window. However, estimating an extra parameter with limited data might result in greater uncertainty for all parameter estimates. In addition, incorporating a logistic release window would require slower numerical integration solutions rather than the fast analytical solutions outlined in the eAppendix (<http://links.lww.com/EDE/A451>), although introducing a Markov Chain Monte Carlo algorithm might help to counter this effect. Ultimately, given that a uniform release window is parsimonious and generally applicable, the current real-time model would be a reasonable first candidate for future Legionnaires' disease outbreaks.

CONCLUSIONS

To summarize, by assuming an abrupt end to a release of *L. pneumophila*, we have developed a general model that can characterize a Legionnaires' disease outbreak. This can provide useful estimates for public-health planning and can help corroborate source identification. In reality, it is likely that reporting delays and recollection bias will occur during an outbreak, and case data may not appear in the same chronological order as the actual dates of symptoms onset. Future work could investigate the application of the real-time model to case data that appear during the course of an outbreak. The research outlined in this paper uses only symptom-onset dates, ie, temporal data. A natural extension would be to incorporate spatial data, such as cases' home and work locations, to estimate the location of the release as well as its timing. Indeed, we have recently assisted in the development of such a tool for the instantaneous release scenario of an anthrax attack.³⁹ Developing the spatial "instantaneous-release" model into a more general "continuous-release" model with techniques outlined in this paper could prove an interesting area of future research. Any new model would

have to be compared with recent aerosol-dispersion investigations that have shown to be effective in identifying the source location of a Legionnaires' disease release.¹⁴ Finally, it is important to emphasize that such modeling tools should always be viewed as an addition to traditional epidemiologic, environmental, and microbiologic investigative approaches, which have an established track record in source identification during a Legionnaires' disease outbreak.

REFERENCES

1. Fields BS, Benson RF, Besser RE. Legionella and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev.* 2002;15:506–526.
2. Legionella Case Definitions. Health Protection Agency. Available at: <http://www.hpa.org.uk/HPA/Topics/InfectiousDiseases/InfectionsAZ/1234859694179>. Accessed May 10, 2010.
3. Sabria M, Alvarez J, Dominguez A, et al. A community outbreak of Legionnaires' disease: evidence of a cooling tower as the source. *Clin Microbiol Infect.* 2006;12:642–647.
4. Garcia-Fulgueiras A, Navarro C, Fenoli D, et al. Legionnaires' disease outbreak in Murcia, Spain. *Emerg Infect Dis.* 2003;9:915–921.
5. Greig JE, Carnie JA, Tallis GF, et al. An outbreak of Legionnaires' disease at the Melbourne Aquarium, April 2000: investigation and case-control studies. *Med J Aust.* 2004;180:566–572.
6. Castellani-Pastoris M, Ciceroni L, Monaco RL, et al. Molecular epidemiology of an outbreak of Legionnaires' disease associated with a cooling tower in Genova-Sestri Ponente, Italy. *Eur J Clin Microbiol Infect Dis.* 1997;16:883–892.
7. Castilla J, Barricarte A, Aldaz J, et al. A large Legionnaires' disease outbreak in Pamplona, Spain: early detection, rapid control and no case fatality. *Epidemiol. Infect.* 2008;136:823–832.
8. Okada M, Kawano K, Kura F, et al. The largest outbreak of Legionellosis in Japan associated with spa baths: epidemic curve and environmental investigation [in Japanese]. *Kansensh-ogaku Zasshi.* 2005;79:365–374.
9. Den Boer JW, Yzerman EP, Schellekens J, et al. A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. *Emerg Infect Dis.* 2002;8:37–43.
10. De Schrijver K, Dirven K, Van Bouwel K, et al. An outbreak of Legionnaire's disease among visitors to a fair in Belgium in 1999. *Public Health.* 2003;117:117–124.
11. Fraser D, Tsai T, Orenstein W, et al. Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med.* 1977;297:1189–1197.
12. McDade J, Shephard C, Fraser D, Tsai TR, Redus MA, Dowdle WR. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N Engl J Med.* 1977;297:1197–1203.
13. Outbreak of legionnaires' disease in Barrow-in-Furness—update. *Commun Dis Rep CDR Wkly.* 2002;12:5–6.
14. Nygård K, Werner-Johansen Ø, Rønsen S, et al. An outbreak of Legionnaires' disease caused by long-distance spread from an industrial air scrubber in Sarpsborg, Norway. *Clin Infect Dis.* 2008;46:61–69.
15. Addiss DG, Davis JP, LaVenture M, et al. Community-acquired Legionnaires' disease associated with a cooling tower: evidence for longer-distance transport of *Legionella pneumophila*. *Am J Epidemiol.* 1989;130:557–568.
16. den Boer JW, Bosman A, Van Steenberghe J, et al. Een legionella-epidemie onder bezoekers van een beurs in Bovenkarspel. 2001. Available at: <http://www.rivm.nl/bibliotheek/rapporten/213690004.pdf>. Accessed January 4, 2011.
17. O'Mahony MC, Stanwell-Smith RE, Tillett HE, et al. The Stafford outbreak of Legionnaires' disease. *Epidemiol Infect.* 1990;104:361–380.
18. Jansà JM, Caylà JA, Ferrer D, et al. An outbreak of Legionnaires' disease in an inner city district: importance of the first 24 hours in the investigation. *Int J Tuberc Lung Dis.* 2002;6:831–838.
19. Sala MR, Arias C, Oliva JM, et al. Community outbreak of Legionnaires disease in Vic-Gurb, Spain in October and November 2005. *Euro Surveill.* 2007;12:223.
20. van den Hoek J, IJzerman E, Coutinho R. Legionella outbreak in

- Amsterdam: a cooling tower as the source. *Ned Tijdschr Geneesk*. 2006;150:1808–1811.
21. Alonso Pérez de Agreda JP. Community outbreak of Legionnaires' disease related to hospital cooling towers in Zaragoza, Spain, May–June 2004. *Eurosurveillance*. 2004;8. Available at: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2526>. Accessed May 10, 2010.
 22. Kirrage D, Reynolds G, Smith GE, Olowokure B. Investigation of an outbreak of Legionnaires' disease: Hereford, UK 2003. *Respir Med*. 2007;101:1639–1644.
 23. Benkel DH, McClure EM, Woolard D, et al. Outbreak of Legionnaires' disease associated with a display whirlpool spa. *Int J Epidemiol*. 2000;29:1092–1098.
 24. Luna HV, Aledón RM, Morán FG, Rodrigo SG. Outbreak of Legionnaires' disease in Torrevieja, Spain, November–December 2005. *Eurosurveillance*. 2005;10. Available at: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2863>.
 25. Walden J, Kaplan K. Estimating time and size of bioterror attack. *Emerg Infect Dis*. 2004;10:1202–1205.
 26. Brookmeyer R, Blades N. Prevention of inhalational anthrax in the US outbreak. *Science*. 2002;295:1861.
 27. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2008. Available at: <http://www.R-project.org>.
 28. Sasaki T, Matsumoto N, Nakao H, et al. An outbreak of Legionnaires' disease associated with a circulating bathwater system at a public bathhouse. Part I: a clinical analysis. *J Infect Chemother*. 2008;14:117–122.
 29. Brookmeyer R, Gail MH. A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J Am Stat Assoc*. 1988;83:301–308.
 30. Arsham H. Kuiper's P-value as a measuring tool and decision procedure for the goodness-of-fit test. *J Appl Stat*. 1988;15:131–135.
 31. Oyston PC, Sjøstedt A, Titball RW. Tularemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nat Rev Microbiol*. 2004;2:967–978.
 32. Saslaw S, Eigelsbach HT, Prior JA, Wilson HE, Carhart S. Tularemia vaccine study. Part II: respiratory challenge. *Arch of Intern Med*. 1961;107:702–714.
 33. Sawyer WD, Dangerfield HG, Hogge AL, Crozier D. Antibiotic prophylaxis and therapy of airborne tularemia. *Bacteriol Rev*. 1966;30:542–550.
 34. Tigertt WD, Benenson AS, Gochenour WS. Airborne Q fever. *Bacteriol Rev*. 1961;25:285–293.
 35. Twisk-Meijssen MJ, Meenhorst PL, van Cronenburg BJ, et al. The course of *Legionella pneumonia* in guinea pigs after inhalation of various quantities of *L. pneumophila*. *Immunobiology*. 1987;176:108–124.
 36. Baskerville A, Fitzgeorge RB, Broster M, Hambleton P, Dennis PJ. Experimental transmission of legionnaires' disease by exposure to aerosols of *Legionella pneumophila*. *Lancet*. 1981;2:1389–1390.
 37. Muller D, Edwards ML, Smith DW. Changes in iron and transferrin levels and body temperature in experimental airborne legionellosis. *J Infect Dis*. 1983;147:302–307.
 38. Breiman RF, Horwitz MA. Guinea pigs sublethally infected with aerosolized *Legionella pneumophila* develop humoral and cell-mediated immune responses and are protected against lethal aerosol challenge. A model for studying host defense against lung infections caused by intracellular pathogens. *J Exp Med*. 1987;165:799–811.
 39. Legrand J, Egan JR, Hall IM, Cauchemez S, Ferguson NM. Estimating the location and spatial extent of a covert anthrax release. *PLoS Comput Biol*. 2009;5:e1000356.