

The application of geographic information systems and spatial data during Legionnaires' disease outbreak responses

M Bull (matthew.bull@hpa.org.uk)¹, I M Hall¹, S Leach¹, E Robesyn²

1. Microbial Risk Assessment, Emergency Response Department, Health Protection Agency, Porton Down, United Kingdom
2. Surveillance and Response Support Unit, European Centre for Disease Prevention and Control, Stockholm, Sweden

Citation style for this article:

Bull M, Hall IM, Leach S, Robesyn E. The application of geographic information systems and spatial data during Legionnaires' disease outbreak responses. *Euro Surveill.* 2012;17(49):pii=20331. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20331>

Article submitted on 28 March 2012 / published on 6 December 2012

A literature review was conducted to highlight the application and potential benefit of using geographic information systems (GIS) during Legionnaires' disease outbreak investigations. Relatively few published sources were identified, however, certain types of data were found to be important in facilitating the use of GIS, namely: patient data, locations of potential sources (e.g. cooling towers), demographic data relating to the local population and meteorological data. These data were then analysed to gain a better understanding of the spatial relationships between cases and their environment, the cases' proximity to potential outbreak sources, and the modelled dispersion of contaminated aerosols. The use of GIS in an outbreak is not a replacement for traditional outbreak investigation techniques, but it can be a valuable supplement to a response.

Background

Legionnaires' disease (LD) is an atypical type of pneumonia caused by bacteria of the genus *Legionella* [1-3]. The disease mainly affects people over 50 years of age, and generally men more than women [4,5]. Smokers, people with certain occupations, and people with underlying medical conditions may be at a higher risk of infection [1]. The early symptoms of Legionnaires' disease can include an influenza-like illness with muscle aches, tiredness, headaches, dry cough and fever [1,2]. The fatality rate of Legionnaires' disease can vary from 1% to 17% of cases in the general population and may be higher in the risk groups [5-9]. The right skewed incubation period distribution has a median of six days but can range between two and 19 days [10].

Susceptible persons typically become infected when they inhale *Legionella* bacteria in aerosolised form. There is no evidence of person-to-person transmission [11]. *Legionella* organisms are found widely in the environment. They multiply under favourable conditions created by man-made water systems, such as hot and cold water systems, whirlpools, water in air

conditioning cooling systems, and cooling towers, from where they can be aerosolised.

The majority of LD cases are reported as single (sporadic) cases which can occur throughout the year, with most cases occurring in late summer and early autumn [3,4,12]. However, clusters and outbreaks also occur [6-9]. During an LD outbreak descriptive epidemiological and (clinical and environmental) microbiological investigations are often sufficient to identify the outbreak source when it becomes clear that all cases have visited a common location. However, there are instances where there is no obvious, common link between cases. It is in these situations that geographic information systems (GIS) can provide supplementary insight.

A GIS can be described as the integration of software and hardware for the digital capture, management, analysis and visualisation of geographically referenced data. The majority of health data are inherently spatial and have a location, be it an address or a broader administrative unit. GIS enable interpretation of this information spatially, looking for patterns, trends and relationships that might exist between disease (or other occurrences), demography, environment, space and time. GIS therefore have wide-ranging applications in public health, including outbreak response.

If a common source is responsible for an increase in LD cases it is reasonable to assume that those infected with LD have been in relatively close spatial proximity to the same source at some point over the likely incubation period range. By using GIS to analyse the spatial distribution of cases and how they have interacted with their environment, including their proximity to potential sources such as cooling towers, it is possible to identify areas in geographical space that are perhaps common between cases and perhaps suggestive of where the source of an outbreak might be located. In this way GIS can help identify an outbreak source, target additional investigation or corroborate findings

from other types of investigation. The aim of this paper was to review the available peer-reviewed scientific literature to highlight the application and potential benefit of using GIS within LD outbreak investigation. This paper will not review the use of GIS for LD outbreak detection or other analysis of surveillance data.

Literature search strategy

The literature was searched at the end of November 2011. The data was sourced using the Scopus (www.scopus.com) and PubMed (www.ukpmc.ac.uk/) citation databases of peer-reviewed literature using the following search terms:

“((cluster analysis OR space OR spatial OR gis OR geographical) AND (legionnaires disease outbreak OR legionellosis))”

The returned titles and abstracts were reviewed by the first author and full texts were obtained for those publications that appeared relevant to the scope of the review. The selected full texts were then further reviewed by all the authors and selected for inclusion, if they provided details on the application of GIS or some type of spatial analysis within an LD outbreak investigation. Articles were excluded if they did not give practical details of the use of GIS or spatial analysis in the context of a *Legionella* outbreak. Additional published materials that were cited in the articles returned by the initial search, and met the selection criteria, were also sourced for inclusion. Unpublished examples of GIS-based analyses employed in LD outbreak response were not considered because they had not been subjected to peer review.

Results

Literature search

Of the 137 articles retrieved in the literature search, four met the inclusion criteria and were included in this review. A further four articles, cited in these articles were also included, together with an additional article that met the inclusion criteria and was known to the authors, but did not appear in the literature search.

Data collection

It is evident that the body of literature covering the application of GIS within LD outbreak response is fairly small; however from the examples available it is clear that the application of GIS relies on the availability of detailed patient case data and, depending on the type of analysis, other data such as information on potential source locations or demographic and meteorological data.

Patient case data

Typically, the incubation period of Legionnaires' disease is between two and 19 days [10]. It is therefore highly desirable to collect data for each patient case (and possibly controls) covering that period of time before the onset of symptoms. In terms of spatial data,

home location is commonly recorded as a minimum requirement to identify a case's location in geographic space; however it is likely that patients will not be stationary at their home location but travel throughout their environment, for work or recreation, over the time period in question. Table 1 summarises the patient case data collected for a range of outbreaks reported in the reviewed literature. The majority of those studies [6,7,13-18] collected other data in addition to home location to gain a fuller understanding of the spaces occupied by cases (and possibly controls). Collecting case data to this level of detail is a challenge in itself, both in terms of resource availability for the outbreak control team and the physical ability of cases (who may be seriously ill) to recall and provide such detailed information. As such, such detailed data are often absent from an outbreak investigation, and the application of GIS-based analyses is therefore not possible or seriously restricted.

Potential source locations

Where the source of an LD outbreak is not clear from the initial descriptive epidemiological investigation, it is often the case that cooling towers or other aerosol-emitting facilities are found to be the responsible sources [3]. By collecting details of the locations of these potential sources, GIS can be used to assess the relative likelihood that a source could be responsible for an outbreak based on the spatial movements of patient cases in relation to each potential source location. In a number of countries it is a statutory requirement to register a cooling tower with a particular administrative body, either at a local, regional or national level; however in other countries it is not [20]. It is likely that a desktop assessment using mapping tools, and some field reconnaissance may be required to quality-check such registers and to also identify other potential outbreak sources.

Demographic data

Data about the population in an area are often utilised to calculate attack rates, providing relative measures of disease occurrence or effects [6,7,13,15,16].

Meteorological data

A number of studies have also made use of meteorological data for the purpose of atmospheric dispersion modelling [6,7] in an attempt to identify whether a modelled release from a suspected facility (such as a cooling tower) is consistent with the spatial pattern of infection. Climatic variables within atmospheric dispersion models can include wind speed, wind direction, temperature, humidity and atmospheric stability measures.

Data analysis

The real value of a GIS for an LD outbreak investigation is to take spatially implicit 'textual' information, such as addresses and descriptions of travel movements, and make them spatially explicit geometric features (coordinates) with linked attributes. This information

TABLE 1

Summary of case data collected in Legionnaires' disease outbreaks for analyses based on geographic information systems

Outbreak	Author(s)	Cases	Case data collected for GIS-based analyses
Pas-de-Calais, France Nov 2003–Jan 2004	[6] Nguyen et al. (2006)	86 cases	As part of a matched case–control study, data on each location visited and time spent at each location (as well as method of transport) was mapped.
Fredrikstad and Sarpsborg, Norway May 2005	[7] Nygard et al. (2008)	56 cases	The location of cases' home addresses was mapped as well as their movements for the 14 days prior to the onset of symptoms.
Murcia, Spain Jul 2001	[13] Garcia-Fulgueiras et al. (2003)	>800 suspected cases reported, 449 confirmed.	As part of a case–control study (consisting of 85 cases and 170 controls), home locations, work locations and travel movements (as well as method of transport) were mapped for the 14 days before onset of symptoms.
Hereford, England Oct–Nov 2003	[14] Kirrage et al. (2007)	28 cases	The location of cases' home addresses was mapped as well as their movements for the 14 days prior to onset of symptoms.
Barcelona, Spain Oct–Nov 2000	[15] Jansa et al. (2002)	54 cases	The location of cases' home addresses was mapped as well as the locations they had visited during the two days prior to their admission to hospital.
Delaware, United States Jul–Sep 1994	[16] Brown et al. (1999)	29 cases	As part of a case–control study participants were asked to identify areas on a gridded map that they had visited, as well as the number of visits made and the length of time spent in each area. Data were collected for the 14 days prior to the onset of symptoms.
South Wales Sep 2010	[17] Keramarou & Evans (2010)	22 cases	The location of cases' home addresses as well as any other locations visited were mapped for the 14 days prior to the onset of symptoms.
Alcoi, Spain July–Sep 2009	[18] Coscolla et al. (2010)	11 cases	The location of cases' home addresses were mapped as well as their movements.
Alcoi, Spain Sep 1999–Dec 2000	[19] Martinez-Beneito et al. (2006)	36 cases in the 1st outbreak, 11 cases in the 2nd outbreak and 97 cases in the 3rd outbreak	As part of a case–control study home addresses were mapped.

can be plotted onto a map and used within analytical operations. Textual information, such as addresses, can often be sufficient to suggest the source of an outbreak if, for example, all cases report having been to the same location, such as a spa pool. However, in those instances where the source remains unclear, the information provided in the case questionnaire can be mapped. Visualising that data on a map could reveal a pattern of infection that may be suggestive of a source or focus the investigation on a particular area. In addition, a number of analytical techniques have been described in the literature that utilise patient case data, as well as other sources of spatial information, to analyse the spatial relationships that exist between cases and their environment.

Potential source proximity analysis

A common strategy employed is to identify potential sources of an outbreak, such as cooling towers, and then to analyse the spatial relationships between each case and each of these potential sources. Kirrage et al. [14] employed this technique within their investigation of the 2003 outbreak in Hereford, United Kingdom. Having identified the locations of cooling towers within the area of the outbreak, each cooling tower location was 'buffered' by 250 m, 500 m, and 1,000 m. A

composite score quantified the risk of exposure and therefore the likely source of contamination amongst seven sites of interest in and around the city centre. When reviewing the composite scores, two sites were identified as being associated with significantly more cases. Additional epidemiological and microbiological investigation then enabled the rapid identification of a single cooling tower as the source of the outbreak.

Garcia-Fulgueiras et al. [13] adopted a similar approach as part of their investigation into the world's largest LD outbreak to date with more than 800 suspected cases (449 confirmed) in Murcia, Spain. As part of the case–control study a variety of data were mapped, including home and work addresses, travel movements and method of transport. Also, thirty zones were defined around potential sources of contaminated aerosols (such as cooling towers). The authors analysed movements through each of these zones and revealed a strong association, in all eight multivariate analyses described in the paper, between passing through the zone surrounding a hospital cooling tower and being ill with LD.

In the same way that simple counts or scores can be attached to a buffer or zone, attack rates can be

calculated to provide a relative measure of disease occurrence within a population. Nygard et al. [7] used attack rate analysis to help identify a commercial air scrubber as being the responsible source for the 2005 outbreak in Fredrikstad and Sarpsborg, Norway. The assumption behind this technique is that the risk of infection will decrease with distance from a facility that is responsible for an outbreak. This part of their investigation revealed that people living within 1 km of a particular industrial air scrubber were most at risk, and that was the only source for which the risk decreased with increasing distance.

Dispersion modelling

Nygard et al. [7] used the AirQUIS Gaussian puff model INPUFF [21] to simulate the dispersion of aerosols emitted from a number of potential sources of infection incorporating data on wind direction, velocity, temperature and atmospheric stability. They had to assume values for particle size of the aerosols, pipe diameter, output velocity and emission rate. Whilst acknowledging some of the limitations of the modelled outputs, the plumes were used to establish the proportion of patients who would have been exposed to the various potential sources by either living or visiting a location within the modelled dispersal region during the incubation period. The results showed the best fit, with the same source highlighted by the attack rate analysis.

Similarly, Nguyen et al. [6] also employed aerosol dispersion modelling in an attempt to simulate the dispersion of aerosols from a suspected cooling tower. An atmospheric dispersion modelling system (ADMS) [22] was used to simulate emissions from the suspected facility during each wave of the outbreak. The outputted maps of aerosol dispersions revealed a good fit between the modelled plumes and the geographical distribution of cases.

As is acknowledged by both Nguyen et al. [6] and Nygard et al. [7], there are a number of difficulties in modelling the airborne dispersal of contaminated aerosols. In essence a plume model attempts to track the concentration or dose of a contaminant through space and time following its release into the atmosphere. A simple Gaussian model makes the simplifying assumption that wind is of a fixed speed and direction for the duration of the release and whilst having an inspiration in turbulent fluid theory, the form of the Gaussian plume is dependent on empirical estimates of downwind dispersion. The effects of buildings and vehicles changing the flow are not included in such a model. Essentially dispersion models serve two major functions: firstly, to estimate the exposed population following a potential release, and secondly, to infer potential release sites from the pattern of observed infections. For the former, insufficient evidence has been compiled to suggest an infectious dose of *Legionella* in humans (or the probability of infection following inhalation of a dose), or how long the bacteria can survive in the atmosphere once aerosolised. As such, converting the contours from a

plume model into exposure and potential infections is difficult without additional strong assumptions being made. The latter use for dispersion modelling is challenging in the majority of outbreaks as the uncertainty regarding the time of infection means that the location of the infection is unclear. Furthermore the total at-risk population in time and space may be unclear. One is often left with simply stating whether the pattern of infection is consistent with a modelled release, rather than making any stronger statements.

Case-based analysis

If no suspect sources are identified, the focus of analysis will have to be on the spatial interactions of each patient case with their general environment. In other words, there is a need to analyse interactions between the places where people live, the places people have visited and the routes they have taken.

Coscolla et al.'s [18] study into the 2009 LD outbreak in Alcoi, Spain involved the collection of detailed patient case data including home location, any other locations visited and routes of transport. Within a GIS each location and route was then buffered by a 500 m radius, with those buffers representing areas in which contact with contaminated aerosols may have taken place and where infection might have occurred. Areas where different cases' buffers intersected were considered to represent locations more likely to contain the source of infection, with the initial hypothesis being that the outbreak originated from a common, static point source (e.g. a cooling tower). However, the authors noted obvious spatial variation in the data with two different neighbourhoods of the city being linked with particular waves of infection over the course of the outbreak. A secondary hypothesis was proposed: that the source of contamination was mobile. An asphalt paving machine was identified as being the responsible infection source. It was used in both neighbourhoods at times consistent with the pattern of infection attached to each wave of the outbreak.

As part of the investigation by Jansa et al. [15] into the 2000 outbreak in Barcelona, Spain, incidence rates by census tracts (geographic boundaries created for the aggregation and reporting of census data) containing approximately 400 people, revealed significant spatial variation. Within the affected area, the incidence rates revealed that the northern part of the district was more heavily affected (6.4/1,000) than the southern area (2.23/1,000). The identified area was subsequently shown to be in closest proximity to the cooling towers identified as responsible for the outbreak by further environmental and microbiological investigation.

Similarly, attack rate analysis was utilised by Nguyen et al. [6] as part of a wide range of analytical methods investigating the 2003–04 outbreak in Pas-de-Calais, France. Their analysis revealed the attack rate was highest in the Harne commune in which the suspected cooling tower was located.

Martinez-Beneito et al. [19] applied a spatial statistical methodology to investigate three consecutive outbreaks in the industrial city of Alcoi, Spain between September 1999 and December 2000. 36 cases were identified in the first outbreak, 11 in the second outbreak, and 97 in the third outbreak. The authors identified a group of controls who were staying in hospital in the same period as cases in the first outbreak and who were of the same sex and roughly the same age. Residential postcodes were obtained and a spatial point process model was constructed with the aim of identifying whether the geographical distribution of the cases could be considered to be random. Ripley's K function [23], a descriptive statistic for identifying deviation from spatial homogeneity, was estimated for cases and controls and a difference between these statistics was then calculated and tested for statistical significance. Results of significance tests suggested higher aggregation of cases than of controls in all outbreaks. Risk surface maps were also estimated for each outbreak. These were given based on the difference between the observed probability of being a case at a particular location and the expected probability of being a case within the city. Thus areas of high risk were highlighted on which attempts to find a source for each outbreak should focus.

Brown et al.'s [16] study looked into a method for calculating dose of exposure. The outbreak was strongly linked to a hospital in Wilmington, Delaware, United States. Attack rate analysis revealed that the highest relative risk existed among hospital staff and those living within a census tract adjacent to the hospital. In total 29 cases met the study's case definition criteria for LD, and 21 of these were included in the case-control study, with three controls being matched to each case. A standardised questionnaire to interview cases and their controls was used that focussed on the area near the hospital where the attack rate was highest. Interviewees were provided with a gridded map of the area. They were asked to mark possible locations for their exposure in the two weeks before onset of illness. Further information was recorded about the number of visits made and the length of time spent in each grid cell. Separate regression models were used to determine the change in frequency and duration of potential exposure in each grid cell and the change in risk associated with a change in distance from the hospital. Risk of illness was found to decrease with increasing distance from the hospital, but to increase for each additional hour spent in grid cells within 0.125 miles of the hospital. The median dose of modelled potential exposure was higher for cases than controls.

Discussion

The use of GIS in LD outbreak investigation is not a replacement for traditional descriptive epidemiological and microbiological investigative techniques, but it should be viewed as a valuable addition to the public health professional's toolbox. However, it is important to keep in mind that each outbreak is a unique event,

and as such not all analytical techniques reviewed in this article will be appropriate in all circumstances. The body of peer reviewed literature covering the application of GIS for LD outbreak investigation is currently relatively small, so the extent to which GIS is used more generally across public health organisations for this purpose is unclear.

Four types of spatial data have been identified in this review as being potentially useful to an outbreak response: case data (i.e. locations visited in incubation period including their home); potential sources in the locality (i.e. a registry of cooling tower locations and field investigation of other sources); information about the broader demography of the population (i.e. how many people live in the administrative regions identified or a control group to compare to cases) and finally meteorological data (i.e. wind speed and direction if dispersion modelling is being performed). To facilitate a response, mechanisms for collecting and storing such data should be in place before an outbreak occurs. These mechanisms should be considered an important aspect of LD outbreak preparedness and have the potential of speeding up and improving substantially the use of these techniques.

Two broad families of statistical analysis were identified from the literature: one using case data to infer zones for further/higher priority field analysis; the other focussing on known potential sources and checking whether the pattern of infection of cases is consistent with a release emanating from there. A third type of analysis that overlaps with these two approaches, dispersion modelling, can be useful if the release occurred over a short time period, but the results of such analysis are likely to be compromised by the uncertainty in infection time of each case and the infectious dose. If resources allow, a carefully designed case-control study that includes appropriate controls might better support source hypothesis testing than using dispersion modelling.

The nature of the outbreak, as well as data availability, will influence the selection of a GIS-based investigative approach. The analytical options, based on data availability, are summarised in Table 2. The techniques that test presumptive sources against the observed distribution of cases can identify a single source, or a number of sources, that are more likely to have been responsible for the outbreak than others. As such these types of analyses can help focus additional investigation, particularly if there are a large number of potential outbreak sources initially being considered. Even in the absence of detailed case data, home locations alone have been successfully utilised, in conjunction with demographic data and potential source location data, to map rates of disease occurrence at varying distances from potential sources. It should be stressed that these techniques are reliant on good quality information about potential source locations, and if the actual source of the outbreak is absent from your

TABLE 2

Summary of analytical techniques used in Legionnaires' disease outbreak investigations, given data availability

Data type	Examples of possible analyses	Considerations
Patient case data (patient home locations only)	<p>Home locations can be plotted for outbreak visualisation.</p> <p>Density analyses (such as kernel density analysis) may be used to highlight areas in space with a high density of cases. Area(s) of higher density may suggest that the outbreak source is within relatively close spatial proximity.</p> <p>If case–control study data are available, comparative analyses between cases and controls can be performed. Basic cluster analyses can be utilised to identify whether spatial clustering is greater in cases than controls (e.g. Martinez-Beneito et al. [19]). Clustering of cases may suggest that an outbreak source is located within relatively close spatial proximity or may identify a region for further (field) investigation.</p>	<p>Simply plotting patient case home locations can provide a spatial context to an outbreak.</p> <p>Using only home location can bias any analyses, as home location is not necessarily the location of infection.</p>
Patient case data (patient home locations + travel histories)	<p>Case travel histories can be plotted for outbreak visualisation (e.g. Coscolla et al. [18]).</p> <p>Density analyses (such as kernel density analysis) may be used to highlight areas in space with a high density of spatial interactions between cases. Area(s) of higher density may suggest that the outbreak source is within relatively close spatial proximity.</p>	<p>Case data that includes travel histories give a more complete record of the spaces occupied by each case (where infection may have taken place).</p> <p>Clear overlaps may be identified but bias may be introduced (each case's travel history must be carefully weighted so that their contributions are equal and reported movements are fairly accounted). Note that without comparator information travel routes may simply highlight popular commuter routes</p>
Patient case data + potential source location data	<p>Zones or buffers can be established around each of the potential source locations. Overlay analysis can then be used to identify which cases live, work or have travelled within each zone. You would expect the responsible source to display a high number of cases living, working or travelling within its zone, compared to other sources (e.g. Kirrage et al. [14]).</p> <p>If case–control study data are available then comparative analyses between cases and controls can be performed (e.g. Garcia-Fulgueiras et al. [13]).</p>	<p>Centrally archived lists of sources may be obsolete (new unregistered sources or decommissioned sources might exist in the locality).</p> <p>Without well designed case–control/cohort study or demographic data, inference on patient data is likely to be biased (i.e. some areas may be visited rarely by certain groups).</p>
Patient case data + demographic data	<p>Demographic data allows for attack rate analysis, providing a relative measure of disease occurrence within a population. Attack rate analysis can be undertaken using populations attached to small-area administrative units and can potentially highlight areas with higher levels of disease occurrence. Such areas should be within close spatial proximity to an outbreak source (e.g. Nguyen et al. [6] and Jansa et al. [15]).</p> <p>Cluster analysis can be used to identify abnormal grouping of cases in space and time, with new techniques being developed that can measure the degree of association between cases.</p> <p>If case–control study data are available then comparative analyses between cases and controls can be performed (e.g. Brown et al. [16]).</p>	<p>Demographic data are normally based on home locations, however, daytime population figures may be significantly different due to the movements of working populations.</p> <p>Knowledge that cases are clustered in space and time may not reduce an area of interest for potential sources, but may potentially confirm other investigations.</p> <p>Case control data may actually be more appropriate/detailed than general population data that reflect only residence.</p>
Patient case data + potential source location data + demographic data	<p>Radial attack rate analysis buffers each potential source at multiple distances and calculates the attack rate within each buffer. For the responsible facility you would expect to observe a pattern where the attack rate decreases with an increase in distance from the facility (e.g. Nygard et al. [7]).</p>	<p>Demographic data are normally based on home locations, however, daytime population figures may be significantly different due to the movements of working populations.</p> <p>Case control data may actually be more appropriate/detailed than general population data.</p>
Patient case data + potential source location data + demographic data + meteorological data	<p>Dispersion modelling allows you to identify whether a modelled plume from a potential source is consistent with the observed pattern of infection (e.g. Nguyen et al. [6] and Nygard et al. [7]).</p>	<p>Dispersion models can provide intuitive outputs if, for example, the release is over clear short time window, there is only one possible source, or people have not moved.</p> <p>A lack of information on dose response and general uncertainty over infection time for cases means that, in many situations, dispersion models will not inform the outbreak control team's hypotheses.</p>

dataset, it will not be considered within your analysis and subsequently will not be identified.

Alternative approaches examine case data in isolation, looking to identify areas in geographic space that display a higher concentration of spatial interaction amongst cases. Some techniques, such as kernel density, attempt to smooth case point data out across space. In this context, it should be noted that without a comparator population from a control group such analysis is difficult to interpret and is simply an alternative visualisation of the point data. However, even whilst lacking quantitative power, it might highlight areas of particular interest. Others aggregate the observed case numbers to small administrative units for which attack rates can be calculated. Upon identifying areas in space that are seemingly common between cases, investigation can be targeted to look for potential outbreak sources within that vicinity. Care should be taken during the interpretation of outputs that the regions identified are not unduly biased by commuting or other similar behaviours in the underlying population. These techniques can be applied to very detailed case data covering the entire travel histories of patient cases over the course of the likely incubation period. However, as above, they can also be applied in situations where only residential address information is available to provide additional insight into an outbreak.

Outbreaks of Legionnaires disease may be seen as a proxy or analogous to other disease outbreaks, such as Q fever. However, caution should be taken when applying these techniques by judging whether the methodology is appropriate to the specific disease. This is especially true given the fact that the applicability and usefulness of the techniques depend very much on characteristics such as the incubation period of the disease, release and/or transmission characteristics, susceptibility, symptomatology, detection and diagnostics.

A GIS can clearly supplement an outbreak response by quickly visualising both case and potential outbreak source information, as well as providing spatial analytical capabilities to interrogate that data. In order to utilise GIS for these purposes it is important to have clear data collection protocols in place ahead of time, and an awareness of the technical and legal issues around storing and managing such information (particularly patient-identifiable data). The usefulness of GIS to outbreak investigations are largely dependent on the availability of good quality case data, and any enhancements to the way such information is collected would ultimately enhance the application of the spatial analytics used to assist in outbreak responses.

Acknowledgments

The authors would like to thank the working group and the project reviewers of the ECDC funded Legionnaires' disease outbreak investigation toolbox project for their input.

This work was funded by the European Centre for Disease Prevention and Control (ECDC, <http://www.ecdc.europa.eu>) through the call for tender OJ/2010/04/14 – PROC/2010/017.

References

1. Fields BS, Benson RF, Besser RE. Legionella and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev.* 2002;15(3):506-26.
2. Stout JE, Yu VL. Legionellosis. *New Engl J Med.* 1997;337(10):682-7.
3. World Health Organization (WHO). Legionella and the prevention of legionellosis. Geneva: WHO; 2007. ISBN 92 4 156297. Available from: http://www.who.int/water_sanitation_health/emerging/legionella.pdf
4. European Centre for Disease Prevention and Control (ECDC). Annual epidemiological report on communicable diseases in Europe 2009. Revised Edition. Stockholm: ECDC; 2010 Jun. Available from: http://ecdc.europa.eu/en/publications/Publications/0910_SUR_Annual_Epidemiological_Report_on_Communicable_Diseases_in_Europe.pdf
5. Benin AL, Benson RF, Besser RE. Trends in legionnaires disease, 1980-1998: declining mortality and new patterns of diagnosis. *Clin Infect Dis.* 2002; 35(9):1039-46.
6. Nguyen TMN, Ilef D, Jarraud S, Rouil L, Campese C, Che D, et al. A community-wide outbreak of Legionnaires' disease linked to industrial cooling towers: How far can contaminated aerosols spread? *J Infect Dis.* 2006;193(1):102-11.
7. Nygard K, Werner-Johansen O, Ronsen S, Simonsen O, Kanestrom A, Ask E, 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(1):61-9.
8. Den Boer JW, Yzerman PFE, Schellekens J, Lettinga KD, Boshuizen HC, Van Steenberghe JE, et al. A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. *Emerg Infect Dis.* 2002; 8(1):37-43.
9. Greig JE, Carnie JA, Tallis GF, Ryan NJ, Tan AG, Gordon IR, et al. An outbreak of Legionnaires' disease at the Melbourne Aquarium, April 2000: investigation and case-control studies. *Med J Aust.* 2004;180(11):566-72.
10. Egan JR, Hall IM, Lemon DJ, Leach S. Modeling Legionnaires' disease outbreaks: estimating the timing of an aerosolized release using symptom-onset dates. *Epidemiology.* 2011;22(2):188-98.
11. Ricketts KD, Charlett A, Gelb D, Lane C, Lee JV, Joseph CA. Weather patterns and Legionnaires' disease: a meteorological study. *Epidemiol Infect.* 2009;137(7):1003-12.
12. Health Protection Agency (HPA). Legionnaires' disease. London: HPA. [Accessed 25 Nov 2011]. Available from: <http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/LegionnairesDisease/>
13. Garcia-Fulgueiras A, Navarro C, Fenoll D, Garcia J, Gonzalez-Diego P, Jimenez-Bunuales T, et al. Legionnaires' disease outbreak in Murcia, Spain. *Emerg Infect Dis.* 2003; 9(8):915-21.
14. Kirtage D, Reynolds G, Smith GE, Olowokure B. Investigation of an outbreak of Legionnaires' disease: Hereford, UK 2003. *Respir Med.* 2007;101(8):1639-44.
15. Jansa JM, Cayla JA, Ferrer D, Gracia J, Pelaz C, Salvador M, 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(9):831-8.
16. Brown CM, Nuorti PJ, Breiman RF, Hathcock AL, Fields BS, Lipman HB, et al. A community outbreak of Legionnaires' disease linked to hospital cooling towers: An epidemiological method to calculate dose of exposure. *Int J Epidemiol.* 1999;28(2):353-9.
17. Keramarou M, Evans MR, for the South Wales Legionnaires' Disease Outbreak Control Team. A community outbreak of Legionnaires' disease in South Wales, August-September 2010. *Euro Surveill.* 2010;15(42): pii=19691. Available from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19691>
18. Coscolla M, Fenollar I, Escribano I, Gonzalez-Candelas F. Legionellosis Outbreak Associated with Asphalt Paving Machine, Spain, 2009. *Emerg Infect Dis.* 2010; 16(9):1381-7.

19. Martinez-Beneito MA, Abellan JJ, Lopez-Quilez A, Vanaclocha H, Zurriaga O, Jorques G, et al. Source detection in an outbreak of Legionnaires' disease. In: Case studies in spatial point process models. Baddeley A, Gregori P, Mateu J, Stoica R, Stoyan D (eds). Springer: 2005; p. 169-182.
20. Ricketts KD, Joseph C, Lee J, Wewalka G, European Working Group for Legionella Infections. Survey on legislation regarding wet cooling systems in European countries. Euro Surveill. 2008;13(38):pii=18982. Available from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18982>
21. Petersen WB, Lavdas LG. INPUFF 2.0 – A multiple source Gaussian puff dispersion algorithm. U.S. Research Triangle Park, NC: Environmental Protection Agency; 1986. Report No.: EPA-600/8-86-024.
22. Cambridge Environmental Research Consultants (CERC). ADMS industrial air pollution model. Cambridge: CERC. [Accessed 25 Nov 2011]. Available from: <http://www.cerc.co.uk/environmental-software/model-documentation.html>
23. Ripley BD. Spatial statistics. Wiley Series in probability and mathematical statistics. New York: John Wiley & Sons; 1981.