

own group and so the genetic background of first generation children can be presumed to be similar to that of their counterparts in the Indian subcontinent. A previous study of Asian children with insulin dependent diabetes in West Yorkshire found that they have the typical genetic susceptibility factors HLA-DR3 and DR4.<sup>10</sup> Thus on this genetic background susceptible individuals seem to be exposed to an environmental triggering factor(s) to which they are not exposed before migration.

What is the nature of this environmental effect? Two types of effect have been considered previously: diet and viral infections. Again one can make a presumption that dietary factors are unlikely as most Asian families in Bradford have maintained their usual diet, although methods of preserving food may have changed. Viral infections may be more likely. Either these children are now being exposed to new agents which are uncommon in India and Pakistan or they are being exposed to viruses in a different way—for example, at a different age. Climatic conditions may be important as childhood diabetes seems rare in hot countries but is increasingly common in cooler northern European countries.

In conclusion, this unique transmigratory population has moved from an area of extremely low incidence of childhood diabetes to an area with a relatively high incidence and children in this community have shown a corresponding increase in disease incidence, now matching that of the indigenous population, thereby providing evidence of an aetiological environmental effect.

We thank Miss S H Benson, Professor A Samad Shera (Karachi), Dr H B Chandalai (Bombay), Dr P A McKinney, and Dr A Franks for advice; Mr M Barrett of Bradford District Metropolitan Council, for demographic data; and

Mrs M E Smith for preparing the manuscript. This work was supported by project grant LW 219 of the Yorkshire Regional Health Authority. AS and RC are supported by the Leukaemia Research Fund.

- 1 Walker A, Cudworth AG. Type 1 (insulin-dependent) diabetic multiplex families: mode of genetic transmission. *Diabetes* 1980;29:1036-9.
- 2 Barnett AH, Eif C, Leslie RDG, Pyke DA. Diabetes in identical twins: a study of 200 pairs. *Diabetologia* 1981;20:87-93.
- 3 Reijonen H, Ilonen J, Michelson B, Akerblom HK. HLA-DQ beta-chain restriction fragment length polymorphism as a risk marker in type 1 (insulin-dependent) diabetes mellitus: a Finnish family study. *Diabetologia* 1990;33:357-62.
- 4 Diabetes Epidemiology Research International. Preventing insulin-dependent diabetes mellitus: the environmental challenge. *BMJ* 1987;295:479-81.
- 5 Toniolo A, Federico G, Manocchio I, Onodera T. Aetiology and pathogenesis of type 1 diabetes: viruses. In: Besser GM, Bodansky HJ, Cudworth AG, eds. *Clinical diabetes*. Philadelphia: Lippincott, 1988:10.1-10.20.
- 6 Helgason T, Jonasson MR. Evidence for a food additive as a cause of ketosis-prone diabetes. *Lancet* 1981;ii:716-20.
- 7 Marmot G. General approaches to migrant studies: the relation between disease, social class and ethnic origin. In: Cruickshank JK, Bevers DG, eds. *Ethnic factors in health and disease*. London: Wright, 1989:12-7.
- 8 Gamble DR. Aetiology and pathogenesis of type 1 diabetes: epidemiology. In: Besser GM, Bodansky HJ, Cudworth AG, eds. *Clinical diabetes*. Philadelphia: Lippincott, 1988:7.1-7.8.
- 9 Menon PSN, Viramani A, Shah P, Raju R, Sethi AK, Sethia S, et al. Childhood onset diabetes mellitus in India: an overview. *International Journal of Diabetes in Developing Countries* 1990;10:11-6.
- 10 Bodansky HJ, Beverley DW, Gelsthorpe K, Saunders A, Bottazzo GF, Haigh D. Insulin dependent diabetes in Asians. *Arch Dis Child* 1987;62:227-30.
- 11 Payne RW, Lane PW, Ainsley AE, Bicknell KE, Digby PGN, Harding SA, et al. *Genstat 5 reference manual*. Oxford: Clarendon Press, 1987.
- 12 Metcalf MA, Baum JD. Incidence of insulin-dependent diabetes in children aged under 15 years in the British Isles during 1988. *BMJ* 1991;302:443-7.
- 13 Bingley PJ, Crake EAM. Rising incidence of IDDM in Europe. *Diabetes Care* 1989;12:289-95.
- 14 Samanta A, Burden AC, Hearnshaw JR, Swift PGF. Diabetes in Asian children. *Lancet* 1990;335:1341.
- 15 Samanta A, Burden AC, Jones GR, Woodlands IG, Clarke M, Swift PGF, et al. Prevalence of insulin-dependent diabetes mellitus in Asian children. *Diabetic Med* 1987;4:65-7.
- 16 Burden AC, Samanta A, Chaundhuri KH. The prevalence and incidence of insulin-dependent diabetes in white and Indian children in Leicester city (UK). *International Journal of Diabetes in Developing Countries* 1990;10:8-10.
- 17 Simmons D. Prevalence and age of onset of type 1 diabetes in adult Asians in the Coventry diabetes study. *Diabetic Med* 1990;7:238-40.

(Accepted 19 February 1992)

## Pinpointing clusters of apparently sporadic cases of legionnaires' disease

Rajinder S Bhopal, Peter Diggle, Barry Rowlingson

### Abstract

**Objectives**—To test the hypothesis that many non-outbreak cases of legionnaires' disease are not sporadic and to attempt to pinpoint cases clustering in space and time.

**Design**—Descriptive study of a case series, 1978-86.

**Setting**—15 health boards in Scotland.

**Patients**—203 probable cases of non-outbreak, non-travel, community acquired legionnaires' disease in patients resident in Scotland.

**Main measures**—Date of onset of disease and postcode and health board of residence of cases.

**Results**—Space-time clustering was present and numerous groups of cases were identified, all but two being newly recognised. Nine cases occurred during three months within two postcodes in Edinburgh, and an outbreak was probably missed. In several places cases occurred in one area over a prolonged period—for example, nine cases in postcode districts G11.5 and G12.8 in Glasgow during five years (estimated mean annual incidence of community acquired, non-outbreak, non-travel legionnaires' disease of 146 per million residents *v* 4.8 per million for Scotland). Statistical analysis showed that the space time clustering of cases in the Glasgow and Edinburgh areas was unusual ( $p=0.036$ ,  $p=0.068$  respectively).

**Conclusion**—Future surveillance requires greater

awareness that clusters can be overlooked; case searching whenever a case is identified; collection of complete information particularly of date of onset of the disease and address or postcode; ongoing analysis for space-time clustering; and an accurate yet workable definition of sporadic cases. Other researchers should re-examine their data on apparently sporadic infection.

### Introduction

Legionnaires' disease is an environmentally acquired bacterial pneumonia associated with explosive outbreaks. However, as indicated by surveillance data in England and Wales,<sup>1</sup> the United States,<sup>2</sup> and Scotland,<sup>3</sup> most cases are not linked with outbreaks, and for these the source of infection is rarely identified.<sup>4,6</sup> Such cases are commonly referred to as sporadic cases.<sup>2,6,7</sup> Strictly speaking, therefore, they should be single cases with no association in time and space with other cases. However, we can find no published definition of sporadic infection with legionnaires' disease or discussion of the measures required to show that such cases are truly sporadic. By default, cases which are not part of identified outbreaks are usually classified as sporadic.

Present advice is to undertake detailed epidemiological and environmental investigations for cases related to outbreaks, but not for sporadic cases.<sup>7</sup> Usually, detailed investigation is started when two

Division of Epidemiology and Public Health, School of Health Care Sciences, Newcastle upon Tyne NE2 4HH  
Rajinder S Bhopal, professor of epidemiology and public health

Department of Mathematics, School of Engineering, Computing, and Mathematical Sciences, University of Lancaster, Lancaster LA1 4YF  
Peter Diggle, professor of statistics  
Barry Rowlingson, research associate

Correspondence to: Professor Bhopal.

BMJ 1992;304:1022-7

cases with a possible common source are reported, in line with the advice of the Communicable Disease Surveillance Centre of the Public Health Laboratory Service. Full investigation of all single cases would often waste resources.<sup>7</sup> One possible consequence of this pragmatic policy is that small clusters may be missed or their detection may be delayed. For example, in the East Cumbria Health District one workplace related cluster of legionnaires' disease would probably have remained undetected if this policy had been followed (P Tiplady, personal communication). Bhopal and colleagues showed that in Scotland non-outbreak legionnaires' disease varied both in time, with abrupt annual variations in the disease incidence<sup>3,5</sup> and an unusual seasonal pattern,<sup>8</sup> and in space, with geographical variations by health authority<sup>3</sup> and location of cooling towers.<sup>5</sup> This paper presents a new analysis of the Scottish data to determine whether clusters of apparently sporadic cases could be pinpointed.

## Subjects and methods

### SOURCES OF DATA AND CLASSIFICATION OF CASES

Details of the methods have been published.<sup>3,5,10</sup> Briefly, case lists of legionnaires' disease held at the Ruchill Hospital department of laboratory medicine (the acting reference laboratory for legionella infection for Scotland<sup>9</sup>) and at the Communicable Disease (Scotland) Unit (which collates Scottish surveillance data, including voluntary laboratory reports) were amalgamated and a master list of 452 patients was prepared. Addresses, postcodes, clinical features, date of onset of the disease, laboratory results, and other data were obtained from laboratory request forms, hospital records, correspondence with general practitioners and consultants, a questionnaire to patients, and statistical returns. Postcodes were obtained for 446 (99%) of the 452 patients on the list and were double checked with addresses and directories of postcodes. (Many errors were found in the postcodes recorded in the medical records.) Postcodes were converted with the postcode directory to map grid reference numbers, which were accurate to within 10 metres.

Postcodes and their equivalents abroad (for example, zip codes in the United States) are codes attached to addresses to help postal delivery.<sup>11</sup> In Britain the codes consist of letters and numbers. The full code can be used to attach a map grid reference,<sup>11</sup> which in Scotland locates the address to within a 10 m grid square. This can be done automatically with the computerised central postcode directory.<sup>11</sup> A part of the postcode (for example, G12) locates the address to within a geographical area known as a postcode district. Some postcode districts are further subdivided into postcode sectors (for example, G12.7). The population residing in postcode districts and sectors can be estimated from the decennial census in Scotland. The geographical and population sizes of postcode areas varies enormously. About 900 postcode sectors exist in Scotland, which has a population of about five million. In addition, some heavy users of postal services have a unique postcode which identifies the premises.

Cases were categorised by travel history (patients who had travelled outside Scotland during the 10 days before their illness were classed as travel related cases), outbreak history (patients not involved in the two known outbreaks of legionellosis<sup>12,13</sup> were classed as non-outbreak cases), and by hospital exposure (patients who had not visited hospital in the 10 days before illness were classed as community acquired cases). A case definition similar to that used by others was used to categorise patients as probable (see appendix), possible, or unlikely cases.<sup>3</sup> Probable cases comprised 364 patients resident in Scotland at the time of illness,

of which 100 were travel associated cases, 60 were associated with either outbreaks or hospital exposure, and one was in a patient who gave only an English address, leaving 203 community acquired and apparently sporadic cases. Two postcodes were missing, and the month of the onset of the disease could not be assigned for five cases. Details of the basis for the diagnosis and the subcategories have been published.<sup>3</sup>

### ANALYSIS

Clustering was sought with four methods of analysis. Firstly, cases were tabulated by Scottish health board area and year of onset and, for Greater Glasgow, Lothian, and Lanarkshire, by year and month. Secondly, point maps were created with a microcomputer based mapping package to locate the residence (as indicated by grid reference number) of community acquired, non-outbreak, non-travel cases.<sup>14</sup> Thirdly, cases suspected as part of a time or space cluster were identified on listings of cases ordered by postcode of residence and date of onset of disease. Finally, a newly developed theory by Diggle *et al*<sup>15</sup> was applied to cases in the Greater Glasgow and Lothian areas to establish the strength of statistical evidence for space-time clustering. (The number of cases in the Lanarkshire area was too small to permit this type of analysis.)

The statistical method of this new theory was developed from the statistical theory of spatial point processes<sup>16,17</sup> and may be viewed as an extension of Knox's test for space-time interaction.<sup>18</sup> Knox's test statistic is the number of pairs of cases which are both closer in space than some predetermined distance threshold  $s_0$  and closer in time than some predetermined time threshold  $t_0$ . Two criticisms of this test are that the values of  $s_0$  and  $t_0$  are often rather arbitrary and that no allowance is made for edge effects which arise either from natural geographical boundaries such as coastlines or because there are unrecorded cases outside the designated study region.

The new method answers the first criticism by interpreting the expected number of close pairs of cases as a function of spatial and temporal separations  $s$  and  $t$ , say  $\lambda K(s, t)$  where  $\lambda$  is the expected number of cases per unit area per unit time over the whole area and time period of the study. The scaling by the space-time incidence rate makes the function  $K(s, t)$  invariant under random thinning, which means that estimates of  $K(s, t)$  remain valid if an unknown proportion of cases are unreported or are otherwise missing from the database for reasons unconnected with their spatial or temporal locations. To answer the second criticism, the simple count of close pairs of cases in Knox's statistic is replaced by a sum of weights chosen to eliminate edge effects.

Another consequence of the definition of  $K(s, t)$  is that if spatial and temporal processes are independent—that is, if there is no space-time interaction—then  $K(s, t)$  factorises as  $K(s, t) = K_1(s)K_2(t)$ , where  $K_1(s)$  and  $K_2(t)$  are the  $K$  functions defined analogously for the purely spatial process of locations<sup>16,17</sup> and for the purely temporal process of dates of onset.<sup>19</sup> Thus the natural estimate of space-time interaction, viewed as a function of spatial and temporal separation, is  $\hat{D}(s, t) = \hat{K}(s, t) - \hat{K}_1(s)\hat{K}_2(t)$ , where  $(\hat{\quad})$  denotes estimates of the corresponding functions.

To construct a formal test of significance for space-time interaction we need to combine the values of  $\hat{D}(s, t)$  from the different values of  $s$  and  $t$ . A natural statistic is:

$$T = \iint \{D(s, t)/SE(s, t)\} ds dt,$$

where  $SE(s, t)$  denotes the standard error of  $\hat{D}(s, t)$ <sup>15</sup> and the range of integration is chosen to cover the relevant ranges of spatial and temporal separations,

which for technical reasons must be small relative to the spatial and temporal dimensions of the study. The sampling distribution of T is approximately normal, but an exact test can be implemented by recomputing T after each of a series of independent random permutations of the dates of onset with the spatial locations held fixed. If we do this  $m-1$  times and the original value of T ranks  $k$ th largest among all  $m$  values, then the  $p$  value of the test is  $p=k/m$ .

## Results

Table I shows the number of cases by year in each health board area. There was annual variation, which was most apparent in Greater Glasgow, Lothian, and Lanarkshire Health Boards. Table II gives details of cases by year and month of onset of disease for Greater Glasgow. Table III lists the cases which were strongly suspected of being part of clusters and indicates the number of hospitals to which these patients were admitted.

Three of the four cases in Ayrshire and Arran Health Board were in patients who lived in adjoining postcode sectors (KA8.8 and KA8.9), and two patients were ill in August 1982 (table III). The two cases in Dumfries and Galloway, in which date of onset of illness was in August and September 1980, were in patients who lived in adjacent postcode sectors (DG1.4 and DG2.9). Postcodes were missing for two of the four cases in Grampian, but there was no obvious spatial clustering of the other cases. The two cases in Tayside in 1980 were in patients who were ill at about the same time (May and June) and lived close together (in postcode sectors DD2.3 and DD3.6). Notably, both of the other Tayside cases, though separated in time (1983 and 1985) were in patients who lived in postcode sector DD4.8.

### GREATER GLASGOW

The distribution of cases in Greater Glasgow is described in tables I to III and figure 1 (which for

clarity omits unclustered cases). There was evidence of clusters in 1978, 1983, 1984, and 1985. Though there was temporal clustering in 1979 (three cases with onset of illness in May and in November and nine in total between July and September), the cases were not in the same or adjacent postcode sectors. The pattern of distribution of cases was of the type previously described—that is, around the city centre.<sup>5</sup> In 1980 and 1981 there was no obvious spatial clustering.

In 1984 a major community acquired outbreak was associated with the G31 postcode district.<sup>12</sup> At about that time, in an adjoining postcode district (G4.0) which had had two apparently clustered cases in 1983 (table III), there were five cases with illness in March, June, August (two cases), and October. There had been awareness of this cluster in Glasgow but the cases were judged to be endemic. In September the G31 outbreak was declared to be over.<sup>12</sup> Over the subsequent 14 months, however, there were six cases in this geographical area, but the dates of illness were separated in time: October (postcode sector G31.4) and November (G31.3) 1984 and January (G31.1), February (G31.3), April (G31.3), and November (G31.3) 1985. In late 1985 there was a major nosocomial outbreak linked with a cooling tower at the Glasgow Royal Infirmary (in postcode G4.0 (fig 1)).<sup>13</sup> There had been 13 apparently sporadic cases between January and April 1985, of which five were in patients who lived in the G21/G22 postcode districts, which are north and north east of the infirmary (the prevailing wind is south westerly). The January and February cases were, at the time, judged to be endemic. Districts G21 and G22 had 17 cases of non-outbreak, non-travel legionnaires' disease between 1978 and 1986: one in 1978, five in 1979, one in 1981, two in 1983 (both in November), one in 1984, six in 1985, and one in 1986. The patients were admitted to five hospitals. The mean annual disease incidence in districts G21/G22 was 25.6 per million residents.

Nine cases were in patients who lived in the postcode sectors G11.5 and G12.8: one was ill in August 1980 (sector G11.5), two in November 1983 (G11.5, G12.8), three in August 1984 (G11.5, G12.8), one in October 1984 (G12.8), and one each in March (G11.5) and April 1985 (G11.5). In the two postcode sectors G11.5 and G12.8 the mean annual incidence for non-outbreak disease was 146 per million residents. By comparison the mean annual incidence of non-outbreak, non-travel infection in Scotland between 1978 and 1986 was 4.83 per million whereas in the City of Glasgow it was 15.5 per million.

For a geographical area approximating to the boundary of the City of Glasgow statistical analysis based on the exact test described above showed that the space-time interaction was, overall, unlikely to occur by chance, the  $p$  value being 0.036.

### LOTHIAN

In Lothian Health Board space-time clustering had also occurred, as indicated in tables I and III and in figure 2 (which omits non-clustered cases). In addition to the four cases in 1982 in patients living in postcode sectors EH4.2, EH4.6, and EH4.7, single cases occurred in 1980 and 1984 in this area. In the nine cases in patients living in sectors of EH6 and EH7 postcode districts the date of onset of disease was between May and December 1983 and these cases were, in retrospect, probably part of an outbreak. One case had occurred earlier in sector EH7.6 in September 1982. Two cases occurred in sector EH4.5, but they were separated by a three year interval.

Statistical analysis for a geographical area approximating to the boundary of the City of Edinburgh again showed that space-time clustering was unlikely to have occurred by chance, though the  $p$  value (0.068) was

TABLE I—Number of non-travel, non-outbreak, community acquired cases of legionnaires' disease in each health board by year

Health board*	Year										Total
	1978	1979	1980	1981	1982	1983	1984	1985	1986		
Argyll and Clyde		1		1	1		1	1			5
Ayrshire and Arran				1	2	1					4
Borders	1						1	1			3
Dumfries and Galloway			2								2
Fife				1	1			1	1		4
Forth Valley			1		2				1		4
Grampian							3				3
Greater Glasgow	7	16	5	5	3	16	36	29	2		119
Lanarkshire	2	2	3	1	1	1	4	4			18
Lothian	2		2		11	16	2		2		35
Orkney				1							1
Tayside			2						1		4
Western Isles		1									1
Total	12	20	15	10	21	35	47	37	6		203

\*No cases in Highland or Shetland.

TABLE II—Number of non-travel cases of legionnaires' disease in Greater Glasgow Health Board by year and month\*

Year	Month												Total
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
1978				1	1		1			1	3		7
1979					3		2	4	3		3		15*
1980	1						1	2			1		5
1981	1			1		1		1		1			5
1982						1	1			1			3
1983	2	2				1	1	1	3		6	1	16
1984		1	2			2	1	3	2	7	6	12	36
1985	4	7	1	2	1	1	1		1	3	4	4	29
1986										1	1		2
Total	8	10	3	4	5	5	7	9	12	14	24	17	118

\*For one case neither the month of onset nor the date of serological testing was known.



TABLE III—Clusters of apparently sporadic cases of legionnaires' disease by area of residence, date of onset of disease and numbers of hospitals and consultants involved

Health board and postcode sector of apparent cluster*	No of cases	Date of onset	No of hospitals	No of individual consultants in charge of each group
Ayrshire and Arran: KA8.8, KA8.9	2	Aug 1982	2	2
Dumfries and Galloway: DG1.4, DG2.9	2	Aug, Sep 1980	1	Unknown†
Greater Glasgow:				
G5.0	3	Oct, Nov 1978	2	3
G21.1, G21.3	2	Nov 1983	2	2
G11.5, G12.8	2	Nov 1983	2	2
G4.0	2	Aug, Nov 1983	1	1
G33.5	2	Sep, Dec 1983	1	2
G4.0	5	Mar, Jun, Aug (2 cases), Oct 1984	2	5
G31.4, G31.3, G31.1	6	Oct, Nov 1984; Jan, Feb, Apr, Nov 1985	2	4
G33.3	3	Jul, Oct, Nov 1984	2	3
G13.4, G13.3	2	Sep, Dec 1984	2	2
G72.8	3	Nov, Dec 1984	2	3
G5.8, G5.9	2	Feb 1985	2	2
G21.2, G22.6, G22.7, G21.4	5	Jan, Feb, Oct 1985	3	6
Lanarkshire:				
ML1.1, ML4.2	2	Oct, Nov 1980	2	2
G67.1, G67.2	2	Nov, Dec 1984	2	2
G71.6, G71.8	2	Feb, Oct 1985	2	2
Lothian:				
EH9.1	2	Aug, Sep 1982	1	1
EH4.2	2	Jul, Aug 1982	2	2
EH4.6, EH4.7	2	Aug, Sep 1982	2	Unknown†
EH6.5, EH6.8, EH7.4, EH7.6	9	May, Aug, Sep, Nov, Dec 1983	5	7‡

\*Postcodes in a row are contiguous. †Consultant for one case was not known.  
‡Consultant for two cases was not known.

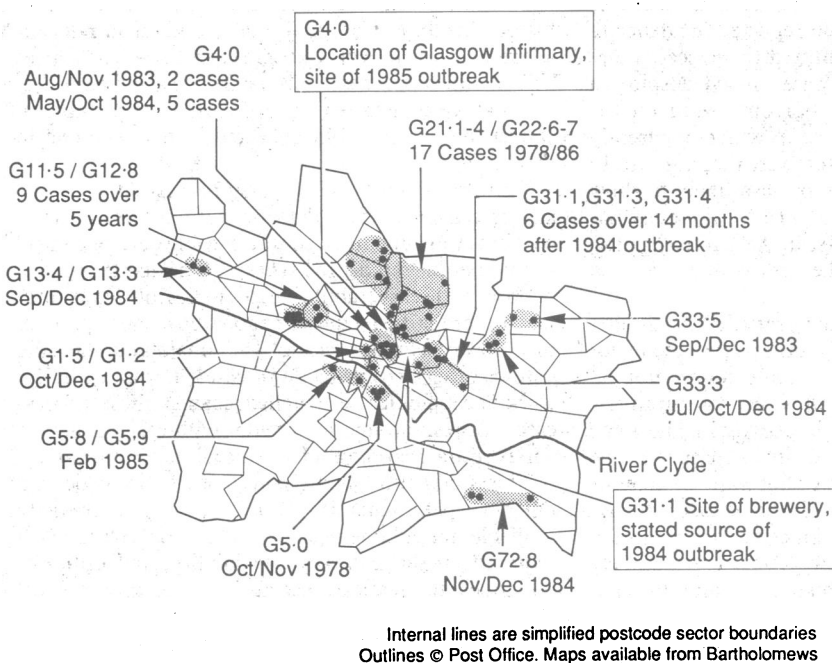


FIG 1—Community acquired, non-travel, and apparently sporadic cases of legionnaires' disease (●) in Glasgow suspected to constitute a space-time cluster

higher than that conventionally taken to be significant.

Since the same null hypothesis is being tested on independent data from two separate regions—namely, the Greater Glasgow and Lothian Health Boards—we can obtain a combined significance level by referring the value of  $C = -2(\log(p_1) + \log(p_2))$  to critical values of the  $\chi^2$  distribution with four degrees of freedom. This gives  $C = 12.02$ , corresponding to a p value of 0.017.

#### LANARKSHIRE

In Lanarkshire Health Board there was also clustering of cases (tables I and III), and the pairs of postcodes in which cases occurred were adjacent to each other.

#### COMPARISON OF CLUSTERING FOR TRAVEL AND NON-TRAVEL CASES

The above analysis highlights numerous time-space clusters for non-outbreak, non-travel infection. Of

these, only two had been previously recognised. For travel related infection only two patients were close in terms of time of illness and place of residence: they lived in postcode sectors G46.6 and G46.7 and were ill in July 1979; they had travelled to different countries.

#### HOSPITALS AND CONSULTANTS

Table III shows that in many, but not all, instances the patients in a cluster were admitted to more than one hospital. The nine patients in postcode sectors G11.5 and G12.8 (1980-5) were admitted to three hospitals, the 17 in the G21/G22 district (1978-86) were admitted to five hospitals, and the nine in the EH6/EH7 district in 1983 were admitted to five hospitals. It was rare for one consultant to be in charge of more than one case in each cluster group.

#### Discussion

Legionnaires' disease is rare, and the recorded mean annual incidence between 1978 and 1986 was about eight per million in Scotland<sup>3</sup> and about three per million in England and Wales<sup>1</sup> and the United States. When two cases of this infection occur in a small geographical area the possibility that they are linked by a common source needs to be excluded. Such cases may not be part of identifiable outbreaks but also may not be truly sporadic.

As is often the case with rare diseases,<sup>20</sup> even without formal statistical analysis it was possible to conclude that space-time clustering was present. The detection and analysis of space-time clustering is a highly topical and developing field. In this study we were able to assess the value of a new extension<sup>15</sup> of Knox's method<sup>18</sup> in comparison with descriptive epidemiology. The combined results from Glasgow and Edinburgh confirmed significant space-time clustering of cases. For other places the number of cases were too few to permit formal statistical analysis.

The value of formal statistical analysis is that a significant result justifies the subsequent, less formal search for possible clusters for detailed epidemiological study. The challenge remains to develop statistical techniques which can detect individual clusters and exclude false clusters with high and quantifiable probability.

The space-time clusters were recognised solely on the basis of residential postcode. If other information of patients' whereabouts had been available, including their workplace, other clusters would probably have been found. The most notable cluster was that of nine cases in patients in Edinburgh (postcode districts EH6 and EH7) between July and September 1983. An outbreak had, apparently, gone unnoticed. These nine patients were admitted to five hospitals and were cared

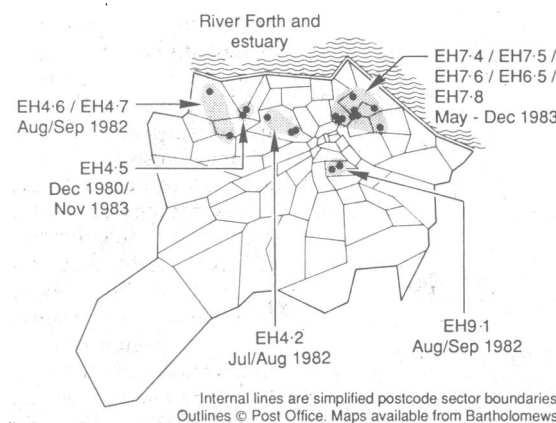


FIG 2—Community acquired, non-travel, and apparently sporadic cases of legionnaires' disease (●) in Edinburgh suspected to constitute a space-time cluster

for by seven separate consultant teams. The most unexpected finding was of cases occurring in one locality but widely separated in time, as shown best in postcode districts G12 and G11 in Glasgow and also in other places, such as in Tayside Health Board.

Six cases occurred in the postcode district G31 in the 14 months after the outbreak there was judged to be over<sup>12</sup> (in one the address was missing on the laboratory form), clearly demonstrating the great difficulties in classifying cases as they occur. These cases were identified singly (by four consultants), and the diagnosis was usually separated by weeks or months, so the cases were categorised as sporadic or endemic. In retrospect, there is some suspicion that they may have been part of the 1984 outbreak.<sup>12</sup> The five cases in postcode districts G21 and G22 in 1985, in retrospect worthy of detailed investigation, were judged at the time to be sporadic cases despite public anxiety (D McEwan, conference of Institution of Environmental Health Officers and Legal Research Institute, Warwick, 1986). With equivocal evidence of an outbreak it may be tempting to classify cases as sporadic or endemic, for the costs of declaring an outbreak in terms of time, manpower, and publicity are high.

The Scottish surveillance system for legionella infections is unlikely to be unique or negligent in having missed clusters. Indeed, the longstanding interest in Scotland in the epidemiology of legionnaires' disease,<sup>3,21</sup> the presence of a centralised reference facility,<sup>9</sup> and the relatively high reported incidence of disease (all factors which permitted this research) are indications that the clusters were missed despite an attentive surveillance system but one based on the standard approach—that is, one in which apparently single cases of disease were not actively investigated. In other countries sporadic cases are also likely to show clustering; for example, variation among states has been reported for sporadic cases in the United States.<sup>2</sup> These findings have both theoretical and practical implications.

As Bhopal and colleagues have argued,<sup>3,5</sup> the finding of clustering contradicts the hypothesis that the major source of sporadic cases is the home water supply,<sup>4</sup> as then the risk would probably be comparatively uniform over time and space in areas with the same water supply—for example, the City of Glasgow.<sup>3</sup> The evidence points to either complex hot water systems or cooling towers as the source of clustered cases. As hot water systems have not been linked with community acquired outbreaks<sup>1</sup> cooling towers are the most likely source of infection in the clustered, non-outbreak, cases.<sup>5</sup>

The pattern of infection described here, and particularly the occurrence of cases in one vicinity but separated by long periods of time, suggests the hypothesis that many apparently sporadic cases result from low intensity, intermittent exposure to a common source and are hence part of mini-outbreaks. Legionnaires' disease is underdiagnosed, and the sporadic cases which were not part of an identified cluster may be single cases only because other associated cases were not recognised.

Small clusters of cases, and sometimes outbreaks, are hard to identify promptly in clinical practice. In urban settings patients admitted on different days are usually cared for by different clinical teams, in different wards, and sometimes in different hospitals, a point shown in table III. Excellent communication between admitting teams and hospital and community staff is then important to detect outbreaks, but communication is often poor.<sup>22</sup> Clusters can be identified by laboratory staff when they are unusually striking<sup>12</sup> and when the patients' clinical histories, dates of onset of disease, and postcodes of residence are given in laboratory request forms. Incompleteness of labora-

tory forms is a major problem; in this study only about half of the forms contained addresses.<sup>23</sup>

#### DETECTING SMALL CLUSTERS

Passive surveillance and routine analysis are insufficient for the timely detection of small clusters. For this there are four main requirements. Firstly, awareness among clinicians and those involved in surveillance that outbreaks or clusters of cases may be easily missed or their detection may be delayed<sup>12,22</sup> and that sporadic cases may herald an outbreak. A single case should be perceived as part of an extant or imminent outbreak until shown otherwise. A high rate of sporadic infection is probably associated with clustering and outbreaks,<sup>24</sup> and some outbreaks are preceded by a single case. For example, Timbury *et al* reported that one case had occurred six weeks before an outbreak<sup>13</sup>; investigation of that single case might have prevented the later outbreak. This nosocomial outbreak<sup>13</sup> was also preceded by a cluster of cases in the nearby community. The significance of the observation is unclear, but it supports the view that epidemics are more likely when the disease is endemic.<sup>24</sup> Case searching should, perhaps, follow the diagnosis of single cases, not just when there is a high suspicion of an outbreak.<sup>7</sup>

Secondly, whenever legionnaires' disease is suspected or diagnosed information may be needed on the clinical details, date of onset, occupation, travel history, history of hospitalisation, and on all relevant addresses (home, work, hospital, and hotel, as appropriate) and postcodes. Once patients have died or left hospital these data may be difficult or impossible to collect so they should be collected on admission or on diagnosis.

Thirdly, though research may be best done regionally or centrally, early detection of clusters is a local concern. Consultants in communicable disease control should be able rapidly to list cases by postcode and date of onset of disease, tabulate cases in flexible ways, and, when there is suspicion of spatial clustering, prepare maps. For each new case of legionnaires' disease the date of onset and postcode of residence and, ideally, workplace should be matched against those of past cases. A desktop microcomputer with easily available database and mapping software is all that is required.

This study showed the power of the postcode as a unit of spatial analysis, one which is potentially available for all medical records. More research is required on the accuracy, availability, and utility of postcodes in medical records.<sup>10,11,23</sup> As geographical information systems become more widely used the quality of data collected on postcodes will probably rise.

Fourthly, particularly in new research on the source of sporadic infection, a definition of a sporadic case of legionnaires' disease is essential. The environmental circumstances of cases which conform to the adopted definition would be of particular interest to researchers.

A more vigorous approach is required for the preliminary investigation of apparently sporadic cases of legionnaires' disease to avoid missing clusters, to develop understanding of the sources and transmission of disease in non-outbreak cases, to enhance the quality of the surveillance data, and ultimately to control and prevent the disease. The present policy, now enshrined in general textbooks,<sup>25</sup> of awaiting two cases in which infection has been from a common source before instigating formal epidemiological and environmental investigation, should be reviewed.

We thank medical records officers, medical colleagues, the Communicable Disease (Scotland) Unit, and the Information Services Division of the Common Services Division for help in tracing missing addresses; Dr J Newell for statistical



advice; Dr R J Fallon for access to laboratory records and ongoing advice; Dr M White for critical comments; Dr A C Gatrell for help in collaboration with Lancaster University; the audio-visual department of Newcastle University Medical School for preparing the figures; Mrs K Ferguson for clerical support; and Mrs L Hutchinson for preparing the manuscript. The study was supported by the Greater Glasgow Health Board and the Pneumonia Research Trust.

## Appendix

### CASE DEFINITION FOR LEGIONNAIRES' DISEASE

"Probable" cases had a clinical history of an acute pneumonia or acute lower respiratory tract infection and one or more of the following: culture of the organism and for *L. pneumophila* only: (a) a fourfold rise (or fall) in antibody titre to at least 64 but no other clinical diagnosis; (b) a static titre of 256 or more but no other clinical diagnosis; (c) positive direct fluorescent antibody test result for respiratory secretions or tissue with use of specific reagents; (d) legionella antigen detected in urine. Patients with other findings were categorised as possible or unlikely cases.

- 1 Committee of Inquiry. *Second Report of the committee of inquiry into the outbreak of legionnaires' disease in Stafford in April 1985*. London: HMSO, 1987.
- 2 England AC III, Fraser DW, Plikaytis BD, Tsai TF, Storch G, Broome CV. Sporadic legionellosis in the United States: the first thousand cases. *Ann Intern Med* 1981;94:164-70.
- 3 Bhopal RS, Fallon RJ. Variation in time and space of non-outbreak legionnaires' disease in Scotland. *Epidemiol Infect* 1991;106:45-61.
- 4 Stout JE, Yu VL, Muraca P. Legionnaires' disease acquired within the homes of two patients. Link to the home water supply. *JAMA* 1987;257:1215-7.
- 5 Bhopal RS, Fallon RJ, Buist EC, Black RJ, Urquhart JD. Proximity of the home to a cooling tower and the risk of non-outbreak legionnaires' disease. *BMJ* 1991;302:378-83.
- 6 Addis DG, Davis JP, Wand PJ, McKinney RM, Gradus HS, Martins RR. Two cases of community-acquired legionnaires' disease: evidence for association with a cooling tower. *J Infect Dis* 1989;159:572-5.

- 7 Bartlett CRL, Macrae AD, Macfarlane JT. *Legionella infections*. London: Edward Arnold, 1986.
- 8 Bhopal RS, Fallon RJ. Seasonal variation of legionnaires' disease in Scotland. *J Infect* 1991;22:153-60.
- 9 Bhopal RS. A survey of laboratory services for the diagnosis of legionnaires' disease in Scotland. *Health Bull (Edinb)* 1990;48:288-93.
- 10 Bhopal RS. A framework for investigating geographical variation in diseases, based on a study of legionnaires' disease. *J Public Health Med* 1991;13:281-9.
- 11 Gatrell AC, Dunn CE, Boyle PJ. The relative utility of the Central Postcode Directory and Pinpoint Address Code in applications of geographical information systems. *Environment and Planning A* 1991;23:1447-58.
- 12 Ad hoc Committee. Outbreak of legionellosis in a community. *Lancet* 1986;ii:380-3.
- 13 Timbury MC, Donaldson JR, McCartney AC, Fallon RJ, Fleigh JD, Lyon D, et al. Outbreak of legionnaires' disease in Glasgow Royal Infirmary: microbiological aspects. *J Hyg (Camb)* 1986;97:393-403.
- 14 Nimmo AW, Alexander E, Innes G, Paterson JG. Computerised mapping of health data. *Health Bull (Edinb)* 1984;42:199-208.
- 15 Diggle PJ, Chetwynd AG, Haggkvist R. *Second-order analysis of space-time clustering*. Lancaster: Lancaster University, 1991. (Department of Mathematics technical report.)
- 16 Ripley BD. *Spatial statistics*. New York: Wiley, 1981.
- 17 Diggle PJ. *Statistical analysis of spatial point patterns*. London: Academic Press, 1983.
- 18 Knox G. Detection of low intensity epidemicity. *British Journal of Preventive and Social Medicine* 1963;17:121-7.
- 19 Diggle PJ. A kernel method for smoothing point process data. *Applied Statistics* 1985;34:138-47.
- 20 Rothman KJ. A sobering start for the cluster buster's conference. *Am J Epidemiol* 1990;132(suppl 1):S6-13.
- 21 Fallon RJ, Abraham WH. Legionella infections in Scotland. *J Hyg (Camb)* 1982;89:439-48.
- 22 Rashed K, Mason A, Gibson J, Francis J, Fairfax A, Doggett P. Legionnaires' disease at Stafford: management of an epidemic. *Lancet* 1986;i:197-9.
- 23 Bhopal RS. Surveillance of disease and information on laboratory request forms: the example of legionnaires' disease. *J Infect* 1991;22:97-8.
- 24 Helms CM, Viner JP, Weisenburger DD, Chiu LC, Renner ED, Johnson W. Sporadic legionnaires' disease: a clinical observation on 87 nosocomial and community-acquired cases. *Am J Med Sci* 1984;288:2-12.
- 25 McFarlane JT. Acute respiratory infections in adults. In: Brewis RAL, Gibson GJ, Geddes DM, eds. *Respiratory medicine*. London: Baillière-Tindall, 1990:906.

(Accepted 10 February 1992)

## Prevalence of HIV antibody in forensic cases

David W Sadler, Derrick J Pounder,  
George E D Urquhart, Morwyn Porter-Boveri

Departments of Forensic  
Medicine and Medical  
Microbiology, University  
of Dundee, Royal  
Infirmary, Dundee  
DD1 9ND

David W Sadler, lecturer in  
forensic medicine  
Derrick J Pounder, professor  
of forensic medicine  
George E D Urquhart,  
consultant virologist

Department of Public  
Health Medicine, Tayside  
Health Board, Dundee  
Morwyn Porter-Boveri,  
consultant

Correspondence to:  
Professor Pounder.

BMJ 1992;304:1027-8

Most people infected with HIV-I will develop AIDS, but the incubation period varies (mean 10 years, range 2-20).<sup>1</sup> Death reports on people known to be infected with HIV-I but without AIDS permit better estimates of the impact of the HIV epidemic.<sup>2</sup> HIV testing of the dead sidesteps many of the ethical and social policy issues raised in the testing of the living.<sup>3</sup> We report the prevalence of HIV antibody in subjects whose deaths had been selected for medicolegal investigation.

### Subjects, methods, and results

This forensic medicine department provides a medicolegal necropsy service for Tayside (population 394 000; 172 860 in Dundee) and north east Fife (population 68 260). Routine testing for HIV-I and HIV-II antibody in all subjects at necropsy was initiated in April 1990 as a health and safety measure. Seropositive cases were not subjected to necropsy except for compelling legal reasons. The identity of seropositive subjects was disclosed to the Department of Public Health only for its confidential records.

In all, 500 subjects, identified by case number, were tested for HIV-I and HIV-II antibody. A further 11 cases were not tested because of simple omission, container leakage, or unavailability of blood in nine (one known to the police as a homosexual) and because a test result was available immediately before death in two (both intravenous drug misusers; one was seropositive).

Samples were tested by enzyme linked immunosorbent assay (ELISA) using recombinant HIV (Abbot Laboratories). Positive results were confirmed by a second ELISA (Wellcome Diagnostics) and further tests as necessary.

The 500 subjects comprised 31 children under 16 and 469 adults (mean age 58, range 16-95); 301 were male. In all, 457 adults had no known risk factors and all proved seronegative. Of the remaining 12, 11 were known drug misusers (seven suspected intravenous drug misusers) and one was suspected of being so by association. The information sources were the police in nine cases, hospital records in three, and general practitioners in four.

Seven of the 12 suspected drug misusers were thought to be positive for HIV-I antibody. On testing, four were seropositive; this information was in hospital records in two cases, general practitioner records in one, and police records in all four. Of the three cases in which the drug misusers were seronegative, two had been thought to be seropositive by the police and one by both the police and the general practitioner; in this last case there was suspicion that the subject had been impersonated to obtain HIV testing and possibly drugs. Cross checking against the Department of Public Health's confidential register of HIV positive people disclosed no omissions or errors. Two of the 31 children had known risk factors: both their fathers were positive for HIV antibody and the mother of one was an intravenous drug misuser. Both children were seronegative.

Of the total 511 subjects in this necropsy series, 13 were known drug misusers—nine probably of intravenous drugs—and five of them were positive for HIV antibody. No other subjects tested positive.

### Comment

HIV-I was first identified in Tayside early in 1984 and spread rapidly among intravenous drug misusers;