Disease outbreak investigations — objectives, methods and importance

P. M. STREBEL, H. G. V. KÜSTNER

Summary
In South Africa approximately 100 disease outbreaks or epidemics requiring investigation and control can be anticipated each year. Outbreak investigations play a critical role in determining effective disease control strategies. The application of modern epidemiological methods to the study of disease outbreaks is discussed and problem-solving methodology, which has similarities to the approach used in clinical diagnostics, described. These methods were applied to the investigation of an outbreak of Legionnaire's disease at a Johannesburg teaching hospital.

Outbreaks of epidemics of disease may be the result of exposure to infectious (e.g. micro-organisms) or non-infectious (e.g. toxins or radiation) agents. The primary objective of outbreak investigations is disease control. In addition epidemics are 'natural experiments' and their investigation yields valuable information in respect of sources of exposure, modes of transmission, the infectivity and incubation period of micro-organisms or the toxicity and latent phase of physical and chemical agents. Information gained from outbreak investigations is combined with clinical and microbiological or toxicological data and acts as the basis for public health policy and interventions aimed at disease control.1

In the USA the Centers for Disease Control investigate about 1 000 outbreaks per annum (J. Lyle Conrad — personal communication). Adjusted to a South African population, this implies about 100 outbreaks a year requiring investigation and control. Locally, many individuals and organisations, e.g. laboratory services, hospital infection-control staff, local health authorities, etc., are involved in the investigation of disease outbreaks. With this in mind, we have summarised an approach to the investigation of disease outbreaks, and discuss the scientific basis of this methodology. In a companion article (p. 329) these methods have been applied to the investigation of an outbreak of Legionnaire's disease, which occurred at a Johannesburg teaching hospital.

Objectives
The immediate objective of a disease outbreak investigation is the control of the current epidemic and the removal of the existing threat to health. To achieve this, it is necessary to define the source and mode of transmission of the agent responsible for the outbreak. For example, in the investigation of an outbreak, the objective is to locate and treat the source of exposure, be it fish contaminated with methylmercury as was the case in the epidemic at Minamata Bay,2 or food soiled by a Salmonella typhi carrier.3 Alternatively, if the agent and mode of transmission are known (e.g. measles), it may be more important to determine the reasons for the outbreak.4 Here the reason may involve poor vaccine coverage or decreased efficacy of the vaccine because of a break in the cold-chain. The investigation may only occur after the outbreak is over. However, the knowledge gained from the investigation acts as a basis for effective control strategies to prevent future epidemics.

An approach to outbreak investigations
Basic problem-solving methodology is used to determine the source of an epidemic. This can be summarised in five steps, which are described in a teaching manual compiled by the Centers for Disease Control, Atlanta, Georgia, USA.5 The steps need not be followed sequentially but must each be addressed during the course of the investigation.

1. Definition of a case
Case criteria determine which patients qualify as a case for the purposes of the epidemiological investigation. Case definitions may be 'rigorous' or 'loose', but ideally incorporate as many of the true cases, while excluding as many of the false cases as possible. Rigorous case definitions have the advantage of excluding nearly all the non-cases but may exclude cases supplying important clues to the source of infection. On the other hand, a loose case definition will tend to dilute the association between exposure and risk of disease by including many non-cases. For these reasons epidemiological and clinical case criteria are rarely identical.

Although the epidemiologist is seldom responsible for the management of individual cases it is useful to see and examine a patient. This allows for clinical confirmation of the disease in question and facilitates decisions on the practical aspects of case definition and case finding.

2. Is it an outbreak?
An outbreak or epidemic may be defined as the occurrence in a community or region of cases of an illness clearly in excess of expectancy.6 Although authoritative authors do not draw a strict distinction between 'epidemics' and 'outbreaks',7,8 it is useful to regard an epidemic as a disease involving large numbers of people which lasts for months to years and covers a wide geographical area. On the other hand, outbreaks affect fewer people, are of shorter duration and tend to be localised to a particular place. The Centers for Disease Control define an outbreak as 'the occurrence of two or more cases which are epidemiologically related'.9

The term 'the occurrence of cases . . . clearly in excess of expectancy' is imprecise and the question of the presence of an outbreak can only be addressed after consideration of the background incidence of disease. This is easily accomplished if good records are kept at the hospital, laboratory or local health authority. In incidences of a newly discovered causative agent or new diagnostic facilities, an 'outbreak' may be partly or wholly artefactual, i.e. due to increased diagnostic effort.
A question often raised is: 'Is the cluster of cases large enough to warrant an investigation?' This depends both on the nature of the disease and on the number of cases. An isolated case can be investigated in respect of the source. However, in general the larger the outbreak the more likely one is to find a cause and the more significant the public health implications of this finding.

3. Characterisation by person, time and place

This step involves the basic description of the problem. It helps to define who is at risk of contracting the disease and thereby acts as the seeding ground for generating hypotheses regarding the source of exposure.

**Person.** Patient details, such as age, sex, socio-economic status, occupation, etc. need to be documented as they may offer valuable clues regarding the source of exposure. A line listing or case register is a useful way to summarise these data. Additional information related to the disease in question may be necessary. For example, in a measles outbreak one would need to know the vaccination status and age at vaccination of the patients.\(^4\)

**Time.** The epidemic curve is a histogram of the frequency of cases by time of onset of disease. This representation of the data supplies useful clues in respect of the time and duration of exposure to the causative agent. In Fig. 1 the relationship between the mode of the epidemic curve, the average incubation period and the time of exposure is shown — if any two of the variables are known, the third can be derived. The shape of the epidemic curve also yields valuable information about the nature of the source of infection. In Fig. 2, the epidemic curve for a common-source outbreak (brief exposure), e.g. water-borne cholera, is compared with the curve arising from a propagated (person-to-person spread) source, e.g. mumps.

**Place.** A spot map is the distribution of cases by place, e.g. residential or work address, plotted on a map of suitable scale. Additional resolution in respect of place is dictated by the circumstances, e.g. by hospital ward in the case of a nosocomial outbreak or by work locale in the case of an occupational exposure. The essence of the exercise is to document where patients were at the time of their estimated exposure and hence to identify case clusters in areas associated with a high risk of exposure.

4. Forming and testing hypotheses

Using all the information available on the outbreak (person, time and place data) and the relevant medical literature, a suggestion as to the source of infection or exposure is made. This suggestion (or hypothesis) is then tested using attack rates. For the hypothesis to be accepted those with exposure must show a significantly higher attack rate than those not exposed. If this is not the case, then the hypothesis is rejected and new hypotheses are sought and tested.

To compute attack rates it is necessary to determine the population at risk of disease (denominator) as well as the number of cases which arise from this population (numerator). For example, in the Legionnaire's disease outbreak referred to above (also see p. 329 this issue), ward-specific attack rates were calculated. All patients in the relevant ward during the period of the outbreak were in the denominator. In the numerator were Legionnaires' disease cases who were in the same ward during their incubation period. By comparing ward-specific attack rates it was possible to determine wards in which the risk of acquiring Legionella infection was high.

5. Communication of the findings

This is the action or intervention phase which is crucial if preventive medicine is to work. Having identified the source of the outbreak (or reason for the outbreak), the most appropriate preventive and control measures must be implemented immediately. This may involve decontamination of the source and/or measures to prevent disease in high-risk groups. It is at this stage — the translation of the study findings into action — that even the best investigation can fail. Because it is often not the administrator or local health authority who is responsible for the implementation of the control measures, the study findings should be presented in a clear and convincing fashion with recommendations which are readily transformed into action.

A clinical analogy

The methods used in an outbreak investigation have a close analogy in the realm of clinical medicine. Both disciplines use a problem-orientated approach. The clinical documentation of patient details by means of history-taking and a physical examination is analogous to the characterisation of the outbreak by person, time and place — both form the database from which subsequent causal hypotheses arise. In a clinical setting the development of a differential diagnosis, with special investigations to confirm the diagnosis, is similar to the hypothesis-generating and testing phase of outbreak methodology. Where the clinician may use laboratory tests or radiological techniques to confirm his diagnosis, the epidemiologist uses attack rates.

In both a clinical and epidemiological setting, a 'spot diagnosis' is a risky business. Sound practice requires that the correct diagnosis or hypothesis is one which best explains the available information. The more closely the source for the outbreak is defined the more selective and cost-effective the control measures will be. This point is illustrated by the investigation of a nosocomial outbreak of Legionnaire's disease reported in the *Journal of the American Medical Association*.\(^8\) Despite the isolation of *Legionella pneumophila* in the hospital drinking-water, the epidemiological investigation showed the hospital cooling tower to be the source of the outbreak. Treatment of the cooling tower only was performed and the outbreak ceased. The expense of treating the entire hospital water system was avoided, indicating the importance of the dictum of 'diagnosis before treatment' in a public health setting.

**Discussion**

Epidemiology is a comparative discipline and in the conduct of disease outbreak investigations attention is given to those who did not become ill as well as to those who did. The objective is to compare diseased individuals with those not diseased and to
contrast individuals exposed to a factor with those not exposed. Any differences in attack rates or exposure rates are then tested for statistical significance. This approach identifies populations or groups at high risk of contracting the disease and hence suggests causes for the disease. In this respect, the methodology used in acute infectious disease epidemiology is no different from chronic disease epidemiology. For example, in the investigation of an outbreak of food poisoning one would contrast those with disease (cases) and those without disease (controls) with respect to their consumption of different foodstuffs. This is the design of a case-control study. Similarly, an analysis of the risk of disease in those eating a suspect foodstuff compared with those not eating the food is a longitudinal cohort study. In the latter instance the sampling direction is from exposure to outcome with only the duration of follow-up being different; short (hours – days) in acute infectious diseases and long (months – years) in the case of a chronic disease.

The most characteristic feature of the methodology used in an outbreak investigation is the hypothesis-generating and testing phase. As in any scientific pursuit, the formulation of the hypothesis is a creative process. It springs from the consideration of available information with respect to both the current problem and previous experience with similar problems. On the basis of a general statement that exposure to a disease-producing agent (either infectious or non-infectious) can result in disease, the following hypothesis is put forward: if X is acting as a source of the agent then individuals exposed to this source should have a higher attack than an appropriate non-exposed group. This is an example of deductive logic – going from the general statement to the formulation and testing of a specific hypothesis. A major strength of this logical approach is the falsifiability of the hypothesis or, in Popperian terms, the concept of critical refutation. Via the process of generating and testing hypotheses, one arrives at the best hypothesis. The characteristics of a good hypothesis are: (i) it explains more of the available data than previous hypotheses; (ii) it makes more precise predictions than previous hypotheses; and (iii) it suggests new tests of its validity, not suggested by older hypotheses. An example to illustrate this point can be taken from the Johannesburg Legionella outbreak. It is not sufficient to argue that *L. pneumophila* was isolated from the hospital water and cases occurred in the hospital, the hospital water was the source of the outbreak. The key facts of the outbreak must support the hypothesis. In this case the location of cases during the incubation period was not associated with either culture-positive water sources or outlets and therefore the hypothesis was rejected. A better hypothesis for the source of the outbreak proved to be patient ventilators, in that it explained more of the available data, made more precise predictions and suggested new tests of its validity, such as an experiment to determine whether patient ventilators can be infected with, and transmit, *Legionella* organisms.

In addition, a statistically significant association between exposure and outcome need not be causal. It needs to be evaluated for bias and/or confounding, as well as biological plausibility, before it can be accepted. Taking the Johannesburg Legionnaire's disease outbreak as an example again, the finding of an association between ventilation and Legionnaire's disease may have been due to confounding by a third variable, e.g. chronic obstructive pulmonary disease (COPD). The suggestion here is that COPD makes a person more likely to be ventilated and is also a risk factor for Legionnaire's disease. If

<table>
<thead>
<tr>
<th>COPD</th>
<th>Attack</th>
<th>Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Non-case</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Not ventilated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Non-case</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

*Fishers exact test, P = 0.005.*

**REFERENCES**