

Review article: Research methods in epidemiology, III

Epidemiology in search of infectious diseases: methods in outbreak investigation

S R PALMER

From the Communicable Disease Surveillance Centre (Welsh Unit), Cardiff Royal Infirmary, Newport Road, Cardiff CF2 1SZ, UK.

The value of epidemiology in the investigation of incidents of infectious disease in the UK does not seem to be always readily appreciated within microbiology^{1,2} or environmental health.³ Yet because microbial disease is a result of the interaction of the agent, the host and environmental factors⁴ neither laboratory diagnosis nor environmental inspection are of themselves sufficient for successful investigation and control. Simply isolating microbes from the environment, even if appropriate samples are available, cannot distinguish between a reservoir of an agent and the outbreak source or cross contamination from the reservoir or source. Microbiologically plausible explanations of outbreaks may be wrong, but it is often difficult to persuade investigators of the need to test such hypotheses.

On the other hand, whilst epidemiology may have evolved from the study of infectious disease epidemics, academic epidemiology seems in the main to have passed infectious diseases by in the UK. This is regrettable as there is still the need to consolidate the foundations of a section of public health medicine which has refused to disappear.⁵ This review, in following the steps taken in an outbreak investigation,⁶ draws attention to both the usefulness of epidemiology and to regularly recurring methodological problems which may be encountered.

Describing the outbreak

The classical "shoe leather" approach to infectious disease incidents is the detailed investigation of individual cases in the light of the well established microbiological features of the particular infection and a confidence in the possible routes of transmission.⁷ The background incidence of infectious diseases in the UK is usually well known and is sufficiently low not to require sophisticated statistical methods of cluster analysis⁸ to recognise a local

outbreak. However, with newly recognised infections or diseases presumed to be caused by an unknown infectious agent, surveys of disease incidence or laboratory isolation rates may be necessary to confirm that an outbreak has occurred. The methodological problems encountered are similar to those found in investigating clusters of cancers.⁸

The cases first reported in an outbreak are usually only a small and sometimes biased sample which may mislead, rendering control measures ineffective. Identification and definition of the exposed population is necessary to enable thorough case finding to be carried out. Paradoxically, however, a single case which does not fit the usual characteristics of time, place and person may provide the most valuable evidence of the source of infection.⁷

Data from cases are characterised by time, place and person.⁶ The epidemic curve, a plot of the number of cases over time of onset of illness, is the most useful and immediate means of assessing the type of outbreak. In point source outbreaks in which all cases are exposed at a given time, onset of symptoms of all primary cases will cluster within the range of the incubation period. For each organism there is a characteristic range within which infecting dose⁹ and portal of entry, as well as host factors¹⁰ (eg, age, immunosuppression) give rise to individual variability. In outbreaks propagated from person to person the occurrence of cases will be spread over several incubation periods with peaks at intervals of the incubation period. In larger community outbreaks of diseases spread by person to person the epidemic curve is usually smoother and the peaks at the generation time of the new cases less obvious. Despite their almost universal use in outbreak investigations, there has been little formal evaluation of their use in practice. Sartwell in 1950 established that the frequency distribution of incubation periods of infectious diseases of both long and short incubation

period followed a log normal curve.¹¹ He suggested the use of the dispersion factor, the antilogarithm of the logarithmic standard deviation, to measure variation in incubation periods. Such analyses may be used to distinguish between point source and continuing source outbreaks, estimate the data of a common exposure, determine time of maximum infectiousness, help explain pathogenic mechanisms,¹² identify the occurrence of secondary spread, and evaluate control measures. However, Armenian and Lilienfeld have claimed the initial efforts of Sartwell "have not been followed by any systematic review of the data on incubation periods".¹³

Developing a hypothesis

The traditional distinction between descriptive (hypothesis generating) and analytical (hypothesis testing) studies has been criticised by Rothman,¹⁴ a criticism which holds true in outbreak investigation. At the simplest level epidemiological investigation may be described in two stages: firstly, generating a hypothesis of causation from descriptive data analysed by time, place and person, and secondly, testing that hypothesis by a case-control or cohort study. But in almost all outbreaks competing a priori hypotheses are voiced at the outset by public health investigators, patients, employers, food producers, etc. Some of these hypotheses may be excluded by the "descriptive" data. These descriptive data in turn may stimulate alternative hypotheses. Indeed, as in clinical investigation of an individual patient, this process of hypothesis generation and refutation goes on continuously as new cases are being interviewed, and challenges the rigid dichotomy of descriptive versus analytic. For example, in an outbreak of *Salmonella ealing*, cases were mainly infants who were discovered to have been fed on a single brand of milk formula.¹⁵ Comparison with a group of well infants showed a significant difference in use of this brand. Was this comparison testing the hypothesis, or could such a design only firm up the hypothesis—the hypothesis seemingly was generated on the same set of cases as it was "tested"? In such outbreaks a priori hypotheses are formulated on the prior knowledge of the epidemiology of salmonellas and, in this case, on the observation of the age distribution of cases. The search for a common milk formula fed to the infants followed a prediction based on the deduction that the most likely vehicle of infection in infants was formula feed.

In some outbreaks the descriptive epidemiology suggests hypotheses for which there are no biological explanations^{16 17} (eg, a cluster of three cases of aplastic anaemia in children¹⁸) or which go against the accepted microbiological theory (eg, an outbreak of shingles¹⁹⁻²¹). In the practice of infectious disease

control, the framework of proven microbiological causes and modes of transmission enables statistical data to be interpreted confidently. Once outside known causal pathways statistical associations are difficult to interpret and public health action, like the clinical diagnosis of an individual patient, usually is taken on the basis of clinical experience and instinct.

Hypothesis testing

When hypotheses are formally tested the epidemiological techniques usually used are retrospective cohort and case-control studies. In certain outbreaks, eg, food poisoning outbreaks in institutions or following receptions, it is usually possible to identify retrospectively and follow up the whole population to identify simultaneously exposures and disease status. (This itself may lead to biases which are described below.) Comparison of attack rates in exposed and unexposed (eg, food specific attack rates) or exposures in cases and controls will often implicate the source. Paradoxically, when a vehicle is universal, eg, tap water, it may be impossible to demonstrate an association with illness unless there is a dose response (which is less likely if the vehicle of infection is heavily contaminated). In addition to universal problems of matching, confounding and appropriate analysis there are regularly recurring problems of design and implementation in outbreak investigations.

SAMPLING BIAS

Sampling may not be required at all as the size of the population at risk or the number of cases is usually small. However the corollary of this is that the power of possible studies in outbreak investigations cannot usually be controlled by the investigators. Nevertheless, the speed at which such studies can be done by experienced investigators, and the relatively low cost of the studies, as well as the potential benefits of demonstrating an association between infection and a possible source, mean that studies of low power will usually still be worth carrying out.

Diagnostic suspicion bias²² in case selection is always a problem when a possible vehicle of infection is widely publicised. For example, people who have eaten a food believed to be the source of an outbreak of botulism may present with neurological symptoms unrelated to botulism. When laboratory confirmation of diagnosis is not possible the only safeguard is to choose cases which meet the strictest clinical case definition (and therefore who are at the most severe end of the disease spectrum).

In case control studies the selection of controls always poses a problem of practicality versus validity. Sampling frames commonly used to select controls

include general practitioner age/sex registers, hotel and reception guest lists, family members of cases, neighbours of cases, and persons investigated by the laboratory but who were negative for the disease in question. When diagnostic access bias²² is considered potentially important (eg, cases of enteric infections are only recognised in the main if they have general practitioners who regularly recommend faecal sampling), it is important to match for neighbourhood of residence, or use patients investigated by laboratories but found to be negative, since failure to do so may lead to a spurious association with, say, a restaurant in the neighbourhood of the "keen" general practitioner. However, using general practitioner lists may result in overmatching. Obtaining controls from general practitioner registers is laborious and slow, and in our experience about 5–10% of general practitioners are reluctant to cooperate.

Increasingly, acquaintances nominated by cases are used as controls¹⁵ although the representativeness of such groups and possible biases they may introduce have not been evaluated systematically. It is possible, for example, that if cases know that a particular exposure is the source of the outbreak their nomination of controls will be influenced. Additionally, the case may know the exposure history of several acquaintances and reveal this with their names to the interviewer, so possibly influencing the interviewer's choice of controls. A protocol to avoid this may be derived, although in practice there is the danger of a conflict of priorities when local investigators, eg, environmental health officers with their own local responsibilities for outbreak investigation, are called upon to carry out the national protocol.

Misclassification of controls can be avoided by asking all the individuals in the cohort or control group about symptoms over the appropriate time period so that those who may be unrecognised cases can be excluded or reclassified. The feasibility of serotesting, swabbing or otherwise testing controls to exclude or reclassify asymptomatic infected persons should always be considered. However, few well people are willing to give a sample of faeces, and obtaining blood samples from well children is difficult and may be considered by some clinicians to be unethical.

MEASUREMENT BIASES

In all but the smallest outbreaks data are collected by questionnaire, but often practicalities require groups to be interviewed by a combination of telephone, mail, face to face and proxy interviews. The sensitivity and specificity of questionnaires is seldom considered. In one study food consumption recalled by people 2–3

days after luncheon was compared with that observed at the time and recorded on video tape.²³ Only four of 32 attendees made no errors in reporting. The sensitivity of the food consumption histories questionnaire was 87.6% and the specificity 96.1%. Thirteen percent of the respondents reported eating "dummy" food items listed on the questionnaire but not served at the luncheon. Such errors in recall can be reduced by providing background details of events, good questionnaire design, and making use of other sources to check data such as diaries, menus, discussion with relatives, etc. Validation of patient histories is often not possible in practice due to the nature of the exposures, the lag between exposure and investigation and the urgency of investigation. However, random errors in recall are unlikely to give rise to false associations between illness and, for example, a food item. Rather they will reduce the power of a study to identify the true vehicle of infection. For some types of data, such as history of immunization, patient or parent recall are particularly inaccurate and validation of vaccine history from medical records is required.²⁴

Loss of cases or controls from the study because of refusal to be interviewed or failure to trace patients is much less a problem in outbreak investigation since the affected population is usually eager to know the cause, and response rates should approach 100%. Throughout the course of their illness cases will have gone over in their minds possible exposures and their recall may be biased by their own preconceptions or by press and media speculations.

Cases will often have been interviewed on many occasions (by general practitioners, environmental health officers, community physicians) before an analytical study is carried out, and this may have introduced bias from suggestions made by interviewers, as well as prompting a more detailed recall. Family members may have suggested likely sources of infection to the cases. Patients may have anticipated the view of the interviewer if the suspected cause of the outbreak has been learned from local media reports and they may alter their answers accordingly. These biases can often be anticipated and can sometimes be assessed in analysis. For example, in a suspected waterborne outbreak of giardiasis in Reno, Nevada, considerable local publicity occurred before the case-control study was conducted.²⁵ However, because of the long incubation period of giardia many new cases became ill several days after the water supply was treated and so media attention persisted and water was still considered suspect. However, increased water consumption was only associated with cases exposed before the water supply was treated. Bias from media reports was likely to have affected later cases also but no association was found.

A major methodological problem is the bias which may result from the interviewer knowing the disease status of the person and having his own suspicion or prejudice as to the source. Selective histories may be taken in which the patient with salmonella is only asked in detail about a particular food. Blinded interviewing is rarely possible since resources are limited (most outbreaks are investigated by only one or two people), and indeed exposure history is usually obtained at the same time as disease history. Professionalism in the technique of interviewing, and use of a structured questionnaire are the only safeguards. Controls, who anyway have no special reason to recall certain exposures, are usually interviewed several days after cases, which adds to problems of accuracy and detail of recall. To help overcome this problem, questionnaires may include memory jogging points related to notable events, etc. It is important that the controls are interviewed as carefully and thoroughly as the cases were.

Conclusion

Do epidemiological studies in outbreak investigations have to be "quick and dirty" or can they be "quick and clean"? They must be quick. There is a need to develop and evaluate epidemiological methods in outbreak investigation so that they are both practical and sound. Epidemiological standards for the practice of the newly re-established discipline in public health medicine—communicable disease control—must be developed so that attributed causes of outbreaks are not left to be a "matter of opinion".²⁶

References

- ¹ Anonymous. Stanley Royd: the epidemiological lesson. *Br Med J*, 1986; **292**: 644–5.
- ² Powell DEB. Salmonella enteritidis. *Lancet* 1989; i: 99.
- ³ North R. *A failure of government. A report on the risk of food poisoning from eggs and egg products*. Bradford: Foodguard Publications, 1989.
- ⁴ Brachman PS. The control of infectious diseases. In: Holland WW, Detels R, Knox G, eds. *Oxford textbook of public health*, Vol 2. Oxford: Oxford University Press, 1985: 8–18.
- ⁵ Evans AS. The eradication of communicable diseases: myth or reality? *Am J Epidemiol* 1985; **122**: 199–207.
- ⁶ Gregg MB. The principles of an epidemic field investigation. In: Holland WW, Detels R, Knox G, eds. *Oxford textbook of public health*, Vol 2. Oxford: Oxford University Press, 1985: 284–99.
- ⁷ Galbraith NS. The application of epidemiological methods in the investigation and control of an acute episode of infection. In: Holland WW, Detels R, Knox G, eds. *Oxford textbook of public health*, Vol 4. Oxford: Oxford University Press, 1985: 3–21.
- ⁸ Openshaw S, Craft AW, Charlton M, Birch JM. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet* 1988; i: 272–3.
- ⁹ Blaser MJ, Newman LS. A review of human salmonellosis 1. Infective dose. *Rev Infect Dis* 1982; **4**: 1096–106.
- ¹⁰ Khoury MJ, Flanders WD, Greenland S, Adams MJ. On the measurement of susceptibility in epidemiologic studies. *Am J Epidemiol* 1989; **129**: 183–90.
- ¹¹ Sartwell PE. The distribution of incubation periods of infectious disease. *Am J Hyg* 1950; **51**: 310–18.
- ¹² Sartwell PE. The incubation period and the dynamics of infectious disease. *Am J Epidemiol* 1966; **83**: 204–16.
- ¹³ Armenian HK, Lilienfeld AM. Incubation period of disease. *Epidemiol Rev* 1983; **5**: 1–15.
- ¹⁴ Rothman KJ. *Modern epidemiology*. Boston, Toronto: Little Brown, 1986.
- ¹⁵ Rowe B, Begg NT, Hutchinson DN, et al. Salmonella ealing infections associated with consumption of infant dried milk. *Lancet* 1987; ii: 900–3.
- ¹⁶ Schwartz RS, Callen JP, Silva J. A cluster of Hodgkin's disease in a small community. *Am J Epidemiol* 1978; **108**: 19–26.
- ¹⁷ Gordis L. Challenges to epidemiology in the next decade. *Am J Epidemiol* 1988; **128**: 1–9.
- ¹⁸ Morgan GJ, Palmer SR, Onions D, Anderson M, Cartwright RA, Bentley DP. A cluster of three cases of aplastic anaemia in children. *Clin Lab Haematol* 1988; **10**: 29–32.
- ¹⁹ Palmer SR, Caul EO, Donald DE, Kwantes W, Tillett H. An outbreak of shingles? *Lancet* 1985; ii: 1108–11.
- ²⁰ Peto TEA, Gilks CF, Juel-Jensen BE. Clusters of shingles. *Lancet* 1985; ii: 1433.
- ²¹ Palmer SR, Tillett H. Shingles clusters. *Lancet* 1986; i: 273–4.
- ²² Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; **32**: 51–63.
- ²³ Decker MD, Booth AL, Dewey MJ, Fricker RS, Hutcheson RH, Schaffner W. Validity of food consumption histories in a foodborne outbreak investigation. *Am J Epidemiol* 1986; **124**: 859–63.
- ²⁴ Orenstein WA, Bernier RH, Dondero TJ, et al. Field evaluation of vaccine efficacy. *Bull WHO* 1985; **63**: 1055–68.
- ²⁵ Navin TR, Juranek DD, Ford M, Minedew DJ, Lippy EC, Pollard RA. Case control study of waterborne giardiasis in Reno, Nevada. *Am J Epidemiol* 1985; **122**: 269–75.
- ²⁶ House of Commons Agriculture Committee. *Salmonella in eggs. First Report*, Vol 2 (Minutes of evidence and Appendices). London: HMSO, 1989.

Accepted for publication July 1989