Infectious diseases considered to be of public health importance are often reported to a central register as part of national or regional surveillance programmes. This allows monitoring of trends and provides a baseline of knowledge allowing for further study, institution of control measures, and evaluation of the effect of these measures.

Such surveillance systems are incomplete compared to the true incidence of infection occurring in the surveyed population. For many of their functions this lack of complete reporting efficiency is not important, in particular if it remains stable over time. In response to increases in the incidence of an infection, either in the form of an apparent outbreak, or of a slower sustained rise, specific studies are usually undertaken to investigate the cause of the increased incidence. This often uses those cases reported routinely in the surveillance system as a starting point.

As a result analytical studies in this setting usually involve the use of case-control methodology. Some problems associated with the use of this method in this setting are the subject of this paper as well as an exploration of possible adaptations of the method. In particular the problem of obtaining unbiased reference exposure information for comparison with reported cases in a surveillance system with low reporting efficiency is emphasized. The example of salmonellosis is used.

Our aim is to highlight the very great selection bias that results when cases from inefficient surveillance systems (such as enteric pathogen surveillance with a reporting efficiency of less than 10%) are compared to controls which are not selected by a similar system. We also suggest a development of case-control methodology to compensate for this problem.

**Case-case comparisons to study causation of common infectious diseases**

Noel McCarthy\(^a,b\) and Johan Giesecke\(^a\)

---

**Background**

Analytical studies of reportable infectious diseases often use the small minority of cases detected through surveillance systems. For many diseases, notification of cases represents a non-random selection process. Apparent differences in exposure histories may be due to biases involved in the surveillance system selection of cases compared to randomly selected controls. In addition, differential recall between cases and controls may occur. One way to avoid these problems is to compare cases with another group of cases with a different disorder selected by a similar surveillance system, although this can introduce new biases.

**Methods**

In infectious diseases cases with the same disease can be divided into aetologically meaningful subgroups by subtyping the pathogen. Exposure history can then be compared between these subgroups.

**Results**

Several biases are removed. The control group composed of other cases does not represent the exposure history of the study base but differs from it in a predictable and useful way. People considered as controls will have a higher incidence of general predisposing factors than the general population. Analysis is limited to factors associated with exposure to the infecting agent.

**Conclusions**

Case-case comparison is a development of case-control methodology made possible by laboratory typing techniques. These comparisons allow a more refined but more focused analysis of the association of some exposures with infection. Determination of how exposure to the infectious agent occurred is more efficient and unbiased than in standard case-control studies but general factors determining whether disease occurs after an infectious exposure can not be studied.

**Keywords**

Epidemiological methods, epidemiological research design, bias (epidemiology), selection bias

Accepted 21 October 1998

---

\(^a\) Swedish Institute for Infectious Disease Control, Stockholm, Sweden.

\(^b\) European Programme for Intervention Epidemiology Training.

Reprint requests to: Dr Noel McCarthy, Bedfordshire Health Authority, Charter House, Alma St, Luton LU1 2PL, UK.
Selection Steps for Cases Detected by Surveillance Systems

Even in Western health care systems reporting efficiency for important diseases with relatively mild symptoms can be low. The process which leads to some incident infections being reported and some not is not random. We will use the example of salmonellosis to illustrate the nature and extent of this process, based mainly on an assessment of the issue by Chalk and Blaser. They used a step by step analysis to describe how incident cases fail to be diagnosed and reported.

Following infection only about 50% of cases have symptoms, depending on host factors, prior antibiotic and possibly other therapy, infecting dose and virulence of the infecting organism. Only a percentage of the symptomatic cases will then seek medical attention. The personality, medical history, geographical and social factors determining this are not well described.

If the patient seeks health care, the doctor may take a faecal specimen for culture depending on usual practice, perhaps seasonal factors, and on assessment of the patient (which may include patient expectations, age, medical history, occupation and medical insurance coverage as well as other factors). The specimen is subsequently processed with a result returned depending on the sensitivity of the local laboratory system.

The doctor may or may not report a positive diagnosis to the public health authorities in a clinic-based reporting system while the isolating laboratory may or may not report (or forward) the isolated strains in a laboratory-based system. Clearly, the above steps do not represent a random loss of cases, but a highly selective process. The extent of loss of infected individuals at each stage is estimated from published data. The summation of these losses leads to the estimate that 1–2.5% of incident salmonella infections are reported. That this may still be an overestimate is suggested from large national outbreaks such as a recently reported one from the US, in which estimated reporting efficiency was under 0.3% for symptomatic cases, implying about 0.1% for infections.

The true study base in these situations is thus not the general population, but rather that group in the general population who would have been notified as cases, had they become infected. This group represents less than 2.5% of the total population and differs systematically from the other 97.5%. Population controls are therefore not comparable to cases in such studies and it is unreasonable to assume that differences in exposures observed between population controls and reported cases are associated with becoming infected rather than the selection process that has occurred. Figure 1 illustrates these different groups of people.

Effect of case selection

The importance of the process involved in the selection of the small proportion of all incident infections which gives rise to notified cases is dependent on two factors: (a) whether the nature and extent of the selection process is sufficient to result in the notified cases being materially different from the group of infected individuals as a whole in terms of exposures, and (b) whether the notified cases will tend to give information concerning exposures in a way that is different from those incident cases which are not diagnosed and reported.

Concerning point (a), it is not obvious that each step in the selection process will differentially affect the prevalence of exposures in the selected cases compared to those not selected. However, the overall selection process is very great with over 95% of incident infections being lost. Some of the factors involved in this loss of cases such as occupation, social status and health seeking behaviour are likely to be related to differences in important exposures. Exposures such as diet and usage of medicines are among the most obvious ones that may be biased by this process. It therefore seems probable that important selection bias will result and cases obtained by a surveillance system will differ materially from those not chosen in terms of exposure history.

Regarding point (b) the type of individual who will present to a doctor and have a specimen taken when they get a diarrhoeal illness may have a different quality of recall than an individual who does not do so despite similar symptoms. In this way the reported cases can have a different quality of recall compared to unreported cases.

Considerations in Selection of Controls

The study base for reported cases is thus a secondary study base which should be considered in control selection. It is a very distinctive study base which is not likely to be well represented by the overall population. In addition to these issues related to case selection and the resulting study base other factors must be considered when selecting controls in this setting. In particular recall differences between cases and controls and the study question being addressed may affect the choice of controls.

Recall bias

Differential recall between cases and controls as a result of the presence of disease is often mentioned as a problem in reports of case-control studies. The importance of this bias is sometimes seen to be slightly overemphasized. To the extent that it is present, it will add to concerns for validity already raised by the other biases mentioned above.

Study question

Finally, in addition to considering how to reduce biases, controls should ideally be selected to maximize efficiency in answering the specific study question being asked.
are known to be involved in the causation of a disease other than the exposure to be studied it is desirable to have similar levels of these other important risk factors among the case and control group.\textsuperscript{13} In practice this is usually difficult except in large randomized controlled trials. In observational epidemiology the known factors are more usually measured and adjustment is made for them in analysis.

In the case of infectious diseases the two broad groups of exposures are the host factors that resulted in the individual becoming infected following exposure to a causal agent on the one hand, and the way in which the individual became exposed to that agent on the other. Separation of these groups would greatly simplify study of aetiology, especially since many of the host factors may be unknown and so cannot be accounted for in analysis.

Problems with ‘Standard’ Control Groups
Several different types of control groups have been used to supply referent information for cases ascertained though communicable disease surveillance registers.

Population controls
Population controls are preferred if the population is the study base and are sometimes used in case-control studies of reportable infectious diseases.\textsuperscript{6} However, as outlined above the study base for cases identified through a surveillance system with low reporting efficiency is far removed from the general population. As such population-based controls are not ideal.

Friend controls
Friend or case nominated controls are sometimes used in this setting.\textsuperscript{4} They can be identified by the cases and are perhaps more likely to agree to take part in a study than members of the general population. There is no evidence that choosing them will accurately mirror the selection process involved in infected individuals becoming reported cases, though it may have some benefit in this regard due to friends tending to have some similarities in social position, profession, lifestyle and use of medical services.\textsuperscript{14} However, this partial replication of the case selection process will be achieved at the expense of a risk of overmatching for relevant exposures due to friends’ similarities.\textsuperscript{14} In addition, the use of friend controls introduces an inherent potential bias due to gregarious people having a greater chance of selection.\textsuperscript{15,16} Overall they are therefore a poor choice.\textsuperscript{14–16} Neighbourhood controls lie somewhere between friend and population controls in terms of benefits and drawbacks.\textsuperscript{15}

Physician nominated controls
Physician selected controls\textsuperscript{5} also allow for some of the selection process between infection and reporting to be matched but again with the risk of overmatching on other factors and a risk that the physicians’ method of choosing another of their patients as a control will itself be biased. New unmeasurable biases therefore replace those aspects of case selection bias that have been removed.

Overview
It is very difficult to select valid controls for cases derived from infectious disease surveillance systems with low reporting efficiency. Those methods which have been used are open to criticism. Modifications of the case-control method should therefore be explored, if this can lead to more valid comparisons.

Alternative Methods

Case-crossover
In the case-crossover method\textsuperscript{17} cases act as their own controls and are therefore matched on all chronic factors, allowing the study to focus on exposures acting over a short time period. It would appear to be particularly suited to the study of infectious diseases with known and relatively short incubation times. An exposure history is taken from the cases to cover at least two time periods. One period (one incubation time before the onset of disease) refers to the time when the exposures leading to disease occurred. Other periods are compared to this. The case-crossover method has already been applied to infectious disease epidemiology once with good results when Dixon showed that it could select risk factors for haemorrhagic fever with renal syndrome more efficiently than a standard case-control study.\textsuperscript{18} In some settings it may not be efficient, however, particularly when surrogate exposures are used.

For example in the investigation of foodborne outbreaks the surrogate exposure ‘foodstuff’ is used. For an outbreak due to transient contamination of a food product the case-crossover technique will not function well since those who were exposed to the contaminated product during their case exposure time will on average also be more exposed to the product during their control exposure time when the same ‘exposure’ was not associated with any alteration in risk. The same may be true for exposure to, for example, an intermittently contaminated air-conditioning system leading to legionellosis, or for any disease resulting from transient contamination of a drinking water supply.

Case-other disease comparison
Use, as controls, of individuals with a disease considered to have a similar impact on recall to the one being studied has been attempted aiming to avoid recall bias. For instance, to control for maternal recall bias in studies of congenital abnormalities investigators have compared one group of congenital abnormalities with another. This approach must be treated cautiously because (a) knowledge of different hypotheses regarding causes for different abnormalities known to mothers means that recall would still be differentially biased, and (b) exposures relevant to the control abnormality mean that these cases do not represent the real exposure experience of the study base.\textsuperscript{14,19} When comparing one cancer group to another these same problems arise although it is sometimes felt that the advantages of altering recall bias outweigh the disadvantages of the introduced selection bias.\textsuperscript{20}

An improvement on this strategy has been to gather information from cases before diagnosis is complete and then treat as controls those who do not fulfil the case diagnosis, thus decreasing the possibility of recall bias. This approach was evaluated in relation to Sudden Infant Death Syndrome\textsuperscript{21} where cases later being diagnosed with other causes of death were considered as controls and compared to the cases which had the Sudden Infant Death Syndrome diagnosis.
confirmed. A population control group was also studied. The findings were that the control group of children who died from other diseases differed markedly from the population-based controls and that this difference was much more important than the relatively smaller differences in sensitivity and specificity of recall of exposure variables as compared to medical records.

It has thus been difficult to avoid increased selection bias when attempting to reduce recall bias by using cases of different diseases as controls. Moreover, in terms of impact on accuracy of estimated risk, deviations from the true risk due to the presence of recall bias can be relatively unimportant in particular if exposure prevalence is low in the control population. This is because if few controls had an exposure the capacity for poor recall on their part to materially affect the result is limited. In contrast systematic differences between the selected controls (who all have a specific disease) and the true study base can have far more impact. This is because both the exposures leading to that disease and all other factors associated with those exposures will be over represented in the ‘control’ group compared to the true study base. This means that the usefulness of comparing cases to cases of another disease is in general rather limited.

**Case-Case Studies for Infectious Diseases**

The problems in comparing cases of one disease to cases of another arise from the fact that the diseases are different. As such the sufferers of each disease differ from the study base in terms of exposures relevant to their disease and also have their recall influenced by their disease and prevailing beliefs relating to it. In addition, if case ascertainment is incomplete and the cases are detected using different surveillance systems, then different surveillance artefacts may also be an important cause of selection bias.

**Aetiologically relevant subgroups**

All of these problems disappear if cases of disease X are compared with themselves. This is not generally useful since there are also no real differences between the groups. In the case of infectious diseases, however, it is often possible to divide all cases of disease X into several subgroups. They remain cases of disease X, with the same predisposing host factors, the same case ascertainment method, the same prevailing attitudes to causation affecting recall by cases. They differ by the subtype of organism which is causing the infection. Since different subtypes of, for example, salmonella affect different foodstuffs to different extents and since outbreaks of infectious diseases usually arise from a single subtype, it is often possible to define epidemiologically meaningful groups of cases by typing of the infecting organisms. For example *Salmonella enteritidis* is common in eggs and poultry but not in other livestock. *Salmonella typhimurium* is more common in pork. Detailed subtyping involving phage typing, antibiotic resistance patterns, and plasmid profiles allows identification of single clones which can then be considered highly likely to have a common source. In this way, cases likely to share a common exposure (it could turn out to be different foods produced from a single batch of contaminated eggs for example) can be defined and compared to other cases who have a different infecting subtype.

**Effects of comparing groups of cases with the same disease**

**Benefits**

The total group of reported cases is available for random or systematic sampling. All cases came through the selection processes involved in becoming a reported case following infection. This removes any potential for bias between the compared groups being caused by the selection process of the surveillance system. This is true for selection biases affecting either their actual exposure history, or in how it is remembered. In addition the differential recall bias occurring when cases are compared to healthy controls is removed by this method. The possibility to simultaneously reduce recall as well as selection bias, in contrast to the more usual situation of having to compromise between the two, is a particularly attractive feature of this approach.

**Drawbacks**

Two problems arise. The first affects the validity of the comparison. Each ‘control-case’ will have one exposure which differs systematically and undesirably from the ideal study base, the exposure which led to their being infected. This could be helped if cases with a broad range of subtypes are included in the control group, since the wide variety of different exposures involved should dilute each other. Random selection of controls from cases due to a broad mix of other serotypes would then be little affected by each individual aetiologically relevant exposure. Alternatively, data can be adjusted for known causative exposures, thus reducing this problem. The second problem is that compared to a healthy control group these ‘control-cases’ will differ in terms of general risk factors for the studied disease. In the case of salmonellosis these factors include medicines which lower gastric acidity and so increase the probability of an infective dose of ingested organisms surviving passage through the stomach for example. This can be seen as a selection bias which renders the study blind to any factor which affects the risk of salmonellosis with all serotypes (general risk factors). This will restrict the range of exposures which can be studied. However, this is predictable and is the only negative effect of this selection bias. Moreover the same selection effect can be carried with some advantages as outlined below.

**Restriction and refinement of analysis**

As described under ‘Drawbacks’ this method isolates exposure to the infecting agent from other aetiological factors. These are essentially host susceptibility factors which affect the risk of succumbing to salmonellosis of any type including age, medicines used, chronic diseases and other unknown factors. Case-case comparison does not allow study of these factors, since they will have similar distributions in case and control groups. Considering the causal pie model of causation of Rothman, the case-case comparison isolates the slice of the causal pie that represents how a case became exposed to the infecting strain from other all slices representing factors, known and unknown, which cause a general predisposition to salmonellosis. In this sense it is similar to the case-crossover design, although case-case comparison does not involve such individual matching as occurs in case-crossover. Instead it frequency matches for other aetiological factors, both known and unknown, when comparing the case group with the control group. It creates groups that are similar for risk factors other than the particular infectious exposure without any individual matching. Also compared to
case-crossover studies it need not be susceptible to time trends in exposure if cases and ‘control-cases’ are recruited over the same time period.\textsuperscript{23,24}

Practical considerations

A considerable real-life advantage with case-case studies for notifiable infectious diseases is that most surveillance centres already routinely collect quite detailed information on exposures for sporadic cases, and maybe even more so in outbreak situations. There thus exist databases on exposure readily available in such centres, which could be used directly when a new outbreak is encountered. Such case-case studies could be undertaken very much faster than an ordinary case-control study.

Naturally there are also practical drawbacks with this method. In small populations it may be difficult to have a large group of controls. Another criticism could be that not all subtypes of an organism are created equal. For instance, some clones may carry virulence plasmids which make them more likely to cause disease following infection than others and therefore cause them to be reported more than those causing less disease. However, this should not seriously affect the validity of the results obtained. A third problem in using historical cases as controls is that exposures such as dietary habits, change over time, and one can probably not go several years back.

Conclusion

This article has considered the impact of low reporting efficiency on analytical studies using cases derived from surveillance registers. It has emphasized the great potential for selection bias that can result. It has considered alternative approaches with different advantages and disadvantages, in particular the possibility of case-case comparisons made feasible by microbiological typing techniques. This approach offers unique opportunities in terms of simultaneously controlling both recall and selection biases involved in routine case-control methodology. The different selection bias that results is predictable and acts to isolate a particular slice of the causal pie from other causes. It will thus not help in determining general risk factors but will allow more rapid and efficient determination of the sources of infections. It could offer a fast and inexpensive alternative to regular case-control studies, and surveillance centres holding large databases on exposures in previous cases should test its practicability.

The many validity problems identified for standard case-control methodology in this area should make us open to exploring new techniques and adaptations of old ones. In addition, despite lots of activity in the field of foodborne illnesses their incidence in Western countries is not falling.\textsuperscript{24} If sufficiently clear and reliable information is available from current methods to guide control measures why is there no improvement? This should discourage complacency and encourage the exploration of alternative approaches.

References


