Observational research methods. Research design II: cohort, cross sectional, and case-control studies

C J Mann

....

Cohort, cross sectional, and case-control studies are collectively referred to as observational studies. Often these studies are the only practicable method of studying various problems, for example, studies of aetiology, instances where a randomised controlled trial might be unethical, or if the condition to be studied is rare. Cohort studies are used to study incidence, causes, and prognosis. Because they measure events in chronological order they can be used to distinguish between cause and effect. Cross sectional studies are used to determine prevalence. They are relatively quick and easy but do not permit distinction between cause and effect. Case controlled studies compare groups retrospectively. They seek to identify possible predictors of outcome and are useful for studying rare diseases or outcomes. They are often used to generate hypotheses that can then be studied via prospective cohort or other studies.

....

Cohort, cross sectional, and case-control studies are often referred to as observational studies because the investigator simply observes. No interventions are carried out by the investigator. With the recent emphasis on evidence based medicine and the formation of the Cochrane Database of randomised controlled trials, such studies have been somewhat glibly maligned. However, they remain important because many questions can be efficiently answered by these methods and sometimes they are the only methods available.

The objective of most clinical studies is to determine one of the following—prevalence, incidence, cause, prognosis, or effect of treatment; it is therefore useful to remember which type of study is most commonly associated with each objective (table 1)

....

Dr C J Mann, Department of Accident and Emergency Medicine, Taunton and Somerset Hospital, Taunton, Somerset, UK

Correspondence to:
Dr C J Mann;
tonygood@doctors.org.uk
....

Table 1

Objective	Common design
Prevalence	Cross sectional
Incidence	Cohort
Cause (in order of reliability)	Cohort, case-control, cross sectional
Prognosis	Cohort
Treatment effect	Controlled trial

While an appropriate choice of study design is vital, it is not sufficient. The hallmark of good research is the rigor with which it is conducted. A checklist of the key points in any study irrespective of the basic design is given in box 1.

Every published study should contain sufficient information to allow the reader to analyse the data with reference to these key points.

In this article each of the three important observational research methods will be discussed with emphasis on their strengths and weaknesses. In so doing it should become apparent why a given study used a particular research method and which method might best answer a particular clinical problem.

COHORT STUDIES

These are the best method for determining the incidence and natural history of a condition. The studies may be prospective or retrospective and sometimes two cohorts are compared.

Prospective cohort studies

A group of people is chosen who do not have the outcome of interest (for example, myocardial infarction). The investigator then measures a variety of variables that might be relevant to the development of the condition. Over a period of time the people in the sample are observed to see whether they develop the outcome of interest (that is, myocardial infarction).

In single cohort studies those people who do not develop the outcome of interest are used as internal controls.

Where two cohorts are used, one group has been exposed to or treated with the agent of interest and the other has not, thereby acting as an external control.

Retrospective cohort studies

These use data already collected for other purposes. The methodology is the same but the study is performed posthoc. The cohort is "followed up" retrospectively. The study period may be many years but the time to complete the study is only as long as it takes to collate and analyse the data.

Advantages and disadvantages

The use of cohorts is often mandatory as a randomised controlled trial may be unethical; for example, you cannot deliberately expose people to cigarette smoke or asbestos. Thus research on risk factors relies heavily on cohort studies.

As cohort studies measure potential causes before the outcome has occurred the study can demonstrate that these "causes" preceded the

---

Box 1

Study purpose

The aim of the study should be clearly stated.

Sample

The sample should accurately reflect the population from which it is drawn.
The source of the sample should be stated.
The sampling method should be described and the sample size should be justified.
Entry criteria and exclusions should be stated and justified.
The number of patients lost to follow up should be stated and explanations given.

Control group

The control group should be easily identifiable.
The source of the controls should be explained—are they from the same population as the sample?
Are the controls matched or randomised—to minimise bias and confounding.

Quality of measurements and outcomes

Validity—are the measurements used regarded as valid by other investigators?
Reproducibility—can the results be repeated or is there a reason to suspect they may be a "one off"?
Blinded—were the investigators or subjects aware of their subject/control allocation?
Quality control—has the methodology been rigorously adhered to?

Completeness

Compliance—did all patients comply with the study?
Drop outs—how many failed to complete the study?
Deaths
Missing data—how much are unavailable and why?

Distorting influences

Extraneous treatments—other interventions that may have affected some but not all of the subjects.
Confounding factors—Are there other variables that might influence the results?
Appropriate analysis—Have appropriate statistical tests been used?

Validity

All studies should be internally valid. That is, the conclusions can be logically drawn from the results produced by an appropriate methodology. For a study to be regarded as valid it must be shown that it has indeed demonstrated what it says it has. A study that is not internally valid should not be published because the findings cannot be accepted.
The question of external validity relates to the value of the results of the study to other populations—that is, the generalisability of the results. For example, a study showing that 80% of the Swedish population has blond hair, might be used to make a sensible prediction of the incidence of blond hair in other Scandinavian countries, but would be invalid if applied to most other populations.

---

outcome, thereby avoiding the debate as to which is cause and which is effect.

A further advantage is that a single study can examine various outcome variables. For example, cohort studies of smokers can simultaneously look at deaths from lung, cardiovascular, and cerebrovascular disease. This contrasts with case-control studies as they assess only one outcome variable (that is, whatever outcome the cases have entered the study with).

Cohorts permit calculation of the effect of each variable on the probability of developing the outcome of interest (relative risk). However, where a certain outcome is rare then a prospective cohort study is inefficient. For example, studying 100 A&E attenders with minor injuries for the outcome of diabetes mellitus will probably produce only one patient with the outcome of interest. The efficiency of a prospective cohort study increases as the incidence of any particular outcome increases. Thus a study of patients with a diagnosis of deliberate self harm in the 12 months after initial presentation would be efficiently studied using a cohort design.

Another problem with prospective cohort studies is the loss of some subjects to follow up. This can significantly affect the outcome. Taking incidence analysis as an example (incidence = cases/per period of time), it can be seen that the loss of a few cases will seriously affect the numerator and hence the calculated incidence. The rarer the condition the more significant this effect.

Retrospective studies are much cheaper as the data have already been collected. One advantage of such a study design is the lack of bias because the outcome of current interest was not the original reason for the data to be collected. However, because the cohort was originally constructed for another purpose it is unlikely that all the relevant information will have been rigorously collected.

Retrospective cohorts also suffer the disadvantage that people with the outcome of interest are more likely to remember certain antecedents, or exaggerate or minimise what they now consider to be risk factors (recall bias).

Where two cohorts are compared one will have been exposed to the agent of interest and one will not. The major disadvantage is the inability to control for all other factors that might differ between the two groups. These factors are known as confounding variables.

A confounding variable is independently associated with both the variable of interest and the outcome of interest. For example, lung cancer (outcome) is less common in people with asthma (variable). However, it is unlikely that asthma in itself confers any protection against lung cancer. It is more probable that the incidence of lung cancer is lower in people with asthma because fewer asthmatics smoke cigarettes (confounding variable). There are a virtually infinite number of potential confounding variables that, however unlikely, could just explain the result. In the past this has been used to suggest that there is a genetic influence that makes people want to smoke and also predisposes them to cancer.

The only way to eliminate all possibility of a confounding variable is via a prospective randomised controlled study. In this type of study each type of exposure is assigned by chance and so confounding variables should be present in equal numbers in both groups.

Finally, problems can arise as a result of bias. Bias can occur in any research and reflects the potential that the sample studied is not representative of the population it was drawn from and/or the population at large. A classic example is using employed people, as employment is itself associated with generally better health than unemployed people. Similarly people who respond to questionnaires tend to be fitter and more motivated than those who do not. People attending A&E departments should not be presumed to be representative of the population at large.

How to run a cohort study

If the data are readily available then a retrospective design is the quickest method. If high quality, reliable data are not available a prospective study will be required.

The first step is the definition of the sample group. Each subject must have the potential to develop the outcome of interest (that is, circumcised men should not be included in a cohort designed to study paraphimosis). Furthermore, the sample population must be representative of the general population if the study is primarily looking at the incidence and natural history of the condition (descriptive).

If however the aim is to analyse the relation between predictor variables and outcomes (analytical) then the sample
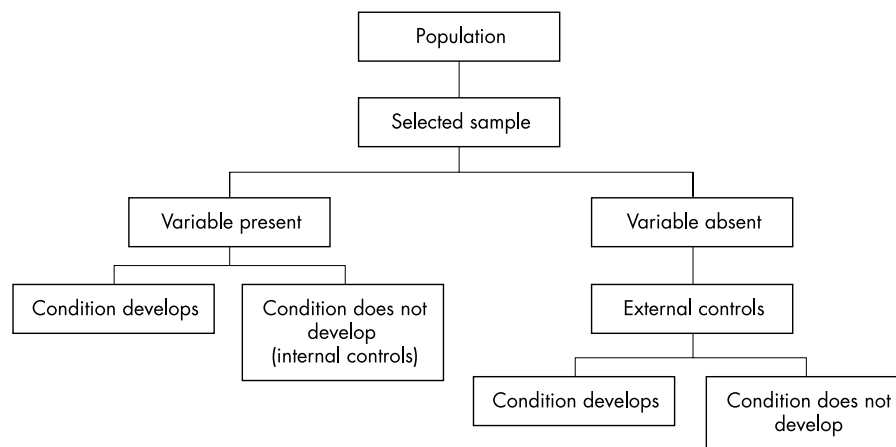
Figure 1 Study design for cohort studies.

```
                        ┌──────────────┐
                        │  Population  │
                        └──────┬───────┘
                        ┌──────┴────────┐
                        │ Selected sample│
                        └──────┬────────┘
            ┌──────────────────┴──────────────────┐
    ┌───────┴────────┐                     ┌───────┴────────┐
    │ Variable present│                     │ Variable absent │
    └───────┬────────┘                     └───────┬────────┘
     ┌──────┴──────┐                        ┌──────┴───────┐
┌────┴────┐ ┌──────┴──────┐           │ External controls│
│Condition│ │Condition does│          └──────┬───────┘
│develops │ │not develop   │            ┌─────┴──────┐
└─────────┘ │(internal     │      ┌─────┴────┐ ┌─────┴──────┐
            │controls)     │      │Condition │ │Condition does│
            └─────────────┘       │develops  │ │not develop   │
                                  └──────────┘ └─────────────┘
```

Key points

Cohort studies
- Cohort studies describe incidence or natural history.
- They analyse predictors (risk factors) thereby enabling calculation of relative risk.
- Cohort studies measure events in temporal sequence thereby distinguishing causes from effects.
- Retrospective cohorts where available are cheaper and quicker.
- Confounding variables are the major problem in analysing cohort studies.
- Subject selection and loss to follow up is a major potential cause of bias.

Table 2

Questionnaire	Interview
Cheap	Expensive
Low response rate	High response rate
Large sample size	Smaller sample size

```
                ┌──────────────┐
                │  Population  │
                └──────┬───────┘
                ┌──────┴────────┐
                │ Selected sample│
                └──────┬────────┘
    ┌─────────┬────────┴────────┬─────────┐
┌───┴────┐┌───┴────┐┌──────────┴┐┌────────┴──┐
│Risk    ││Risk    ││No risk    ││No risk    │
│factor  ││factor  ││factor     ││factor     │
│Disease ││No      ││Disease    ││No disease │
│present ││disease ││present    ││           │
└────────┘└────────┘└───────────┘└───────────┘
```

Figure 2 Study design for cross sectional studies

should contain as many patients likely to develop the outcome as possible, otherwise much time and expense will be spent collecting information of little value.

Each variable studied must be accurately measured. Variables that are relatively fixed, for example, height need only be recorded once. Where change is more probable, for example, drug misuse or weight, repeated measurements will be required.

To minimise the potential for missing a confounding variable all probable relevant variables should be measured. If this is not done the study conclusions can be readily criticised. All patients entered into the study should also be followed up for the duration of the study. Losses can significantly affect the validity of the results. To minimise this as much information about the patient (name, address, telephone, GP, etc) needs to be recorded as soon as the patient is entered into the study. Regular contact should be made; it is hardly surprising if the subjects have moved or lost interest and become lost to follow up if they are only contacted at 10 year intervals!

Beware, follow up is usually easier in people who have been exposed to the agent of interest and this may lead to bias.

Examples
There are many famous examples of Cohort studies including the Framingham heart study,[2] the UK study of doctors who smoke[3] and Professor Neville Butler's studies on British children born in 1958.[4] A recent example of a prospective cohort study by Davey Smith *et al* was published in the *BMJ*[5] and a retrospective cohort design was used to assess the use of A&E departments by people with diabetes.[6]

CROSS SECTIONAL STUDIES
These are primarily used to determine prevalence. Prevalence equals the number of cases in a population at a given point in time. All the measurements on each person are made at one point in time. Prevalence is vitally important to the clinician
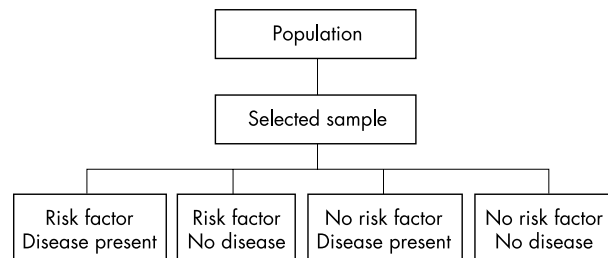
because it influences considerably the likelihood of any particular diagnosis and the predictive value of any investigation. For example, knowing that ascending cholangitis in children is very rare enables the clinician to look for other causes of abdominal pain in this patient population.

Cross sectional studies are also used to infer causation.

At one point in time the subjects are assessed to determine whether they were exposed to the relevant agent and whether they have the outcome of interest. Some of the subjects will not have been exposed nor have the outcome of interest. This clearly distinguishes this type of study from the other observational studies (cohort and case controlled) where reference to either exposure and/or outcome is made.

The advantage of such studies is that subjects are neither deliberately exposed, treated, or not treated and hence there are seldom ethical difficulties. Only one group is used, data are collected only once and multiple outcomes can be studied; thus this type of study is relatively cheap.

Many cross sectional studies are done using questionnaires. Alternatively each of the subjects may be interviewed. Table 2 lists the advantages and disadvantages of each.

Any study with a low response rate can be criticised because it can miss significant differences in the responders and non-responders. At its most extreme all the non-responders could be dead! Strenuous efforts must be made to maximise the numbers who do respond. The use of volunteers is also problematic because they too are unlikely to be representative of the general population. A good way to produce a valid sample

would be to randomly select people from the electoral role and invite them to complete a questionnaire. In this way the response rate is known and non-responders can be identified. However, the electoral role itself is not an entirely accurate reflection of the general population. A census is another example of a cross sectional study.

Market research organisations often use cross sectional studies (for example, opinion polls). This entails a system of quotas to ensure the sample is representative of the age, sex, and social class structure of the population being studied. However, to be commercially viable they are convenience samples—only people available can be questioned. This technique is insufficiently rigorous to be used for medical research.

How to run a cross sectional study

Formulate the research question(s) and choose the sample population. Then decide what variables of the study population are relevant to the research question. A method for contacting sample subjects must be devised and then implemented. In this way the data are collected and can then be analysed

Advantages and disadvantages

The most important advantage of cross sectional studies is that in general they are quick and cheap. As there is no follow up, less resources are required to run the study.

Cross sectional studies are the best way to determine prevalence and are useful at identifying associations that can then be more rigorously studied using a cohort study or randomised controlled study.

The most important problem with this type of study is differentiating cause and effect from simple association. For example, a study finding an association between low CD4 counts and HIV infection does not demonstrate whether HIV infection lowers CD4 levels or low CD4 levels predispose to HIV infection. Moreover, male homosexuality is associated with both but causes neither. (Another example of a confounding variable).

Often there are a number of plausible explanations. For example, if a study shows a negative relation between height and age it could be concluded that people lose height as they get older, younger generations are getting taller, or that tall people have a reduced life expectancy when compared with short people. Cross sectional studies do not provide an explanation for their findings.

Rare conditions cannot efficiently be studied using cross sectional studies because even in large samples there may be no one with the disease. In this situation it is better to study a cross sectional sample of patients who already have the disease (a case series). In this way it was found in 1983 that of 1000 patients with AIDS, 727 were homosexual or bisexual men and 236 were intrvenous drug abusers.[6] The conclusion that individuals in these two groups had a higher relative risk was inescapable. The natural history of HIV infection was then studied using cohort studies and efficacy of treatments via case controlled studies and randomised clinical trials.

Examples

An example of a cross sectional study was the prevalence study of skull fractures in children admitted to hospital in Edinburgh from 1983 to 1989.[7] Note that although the study period was seven years it was not a longitudinal or cohort study because information about each subject was recorded at a single point in time.

A questionnaire based cross sectional study explored the relation between A&E attendance and alcohol consumption in elderly persons.[9]

A recent example can be found in the *BMJ*, in which the prevalence of serious eye disease in a London population was evaluated.[10]

Key points

Cross sectional studies

- Cross sectional studies are the best way to determine prevalence
- Are relatively quick
- Can study multiple outcomes
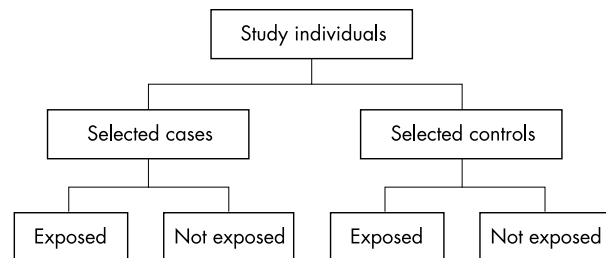- Do not themselves differentiate between cause and effect or the sequence of events

Figure 3 Study design for case-control studies.

CASE-CONTROL STUDIES

In contrast with cohort and cross sectional studies, case-control studies are usually retrospective. People with the outcome of interest are matched with a control group who do not. *Retrospectively* the researcher determines which individuals were exposed to the agent or treatment or the prevalence of a variable in each of the study groups. Where the outcome is rare, case-control studies may be the only feasible approach.

As some of the subjects have been deliberately chosen because they have the disease in question case-control studies are much more cost efficient than cohort and cross sectional studies—that is, a higher percentage of cases per study.

Case-control studies determine the relative importance of a predictor variable in relation to the presence or absence of the disease. Case-control studies are retrospective and cannot therefore be used to calculate the relative risk; this a prospective cohort study. Case-control studies can however be used to calculate odds ratios, which in turn, usually approximate to the relative risk.

How to run a case-control study

Decide on the research question to be answered. Formulate an hypothesis and then decide what will be measured and how. Specify the characteristics of the study group and decide how to construct a valid control group. Then compare the "exposure" of the two groups to each variable.

Advantages and disadvantages

When conditions are uncommon, case-control studies generate a lot of information from relatively few subjects. When there is a long latent period between an exposure and the disease, case-control studies are the only feasible option. Consider the practicalities of a cohort study or cross sectional study in the assessment of new variant CJD and possible aetiologies. With less than 300 confirmed cases a cross sectional study would need about 200 000 subjects to include one symptomatic patient. Given a postulated latency of 10 to 30 years a cohort study would require both a vast sample size and take a generation to complete.

In case-control studies comparatively few subjects are required so more resources are available for studying each. In consequence a huge number of variables can be considered. This type of study is therefore useful for generating hypotheses that can then be tested using other types of study.

This flexibility of the variables studied comes at the expense of the restricted outcomes studied. The only outcome is the

presence or absence of the disease or whatever criteria was chosen to select the cases.

The major problems with case-control studies are the familiar ones of confounding variables (see above) and bias. Bias may take two major forms.

Sampling bias
The patients with the disease may be a biased sample (for example, patients referred to a teaching hospital) or the controls may be biased (for example, volunteers, different ages, sex or socioeconomic group).

Observation and recall bias
As the study assesses predictor variables retrospectively there is great potential for a biased assessment of their presence and significance by the patient or the investigator, or both.

Overcoming sampling bias
Ideally the cases studied should be a random sample of all the patients with the disease. This is not only very difficult but in many instances is impossible because many cases may not have been diagnosed or have been misdiagnosed. For example, many cases of non-insulin dependent diabetes will not have sought medical attention and therefore be undiagnosed. Conversely many psychiatric diseases may be differently labelled in different countries and even by different doctors in the same country. As a result they will be misdiagnosed for the purposes of the study. However, in reality you are often left studying a sample of those patients who it is possible to recruit. Selecting the controls is often a more difficult problem.

To enable the controls to represent the same population as the cases, one of four techniques may be used.

(1) A convenience sample—sampled in the same way as the cases, for example, attending the same outpatient department. While this is certainly convenient it may reduce the external validity of the study.

(2) Matching—the controls may be a matched or unmatched random sample from the unaffected population. Again the problems of controlling for unknown influences is present but if the controls are too closely matched they may not be representative of the general population. "Over matching" may cause the true difference to be underestimated.

The advantage of matching is that it allows a smaller sample size for any given effect to be statistically significant.

(3) Using two or more control groups. If the study demonstrates a significant difference between the patients with the outcome of interest and those without, even when the latter have been sampled in a number of different ways (for example, outpatients, in patients, GP patients) then the conclusion is more robust.

(4) Using a population based sample for both cases and controls. It is possible to take a random sample of all the patients with a particular disease from specific registers. The control group can then be constructed by selecting age and sex matched people randomly selected from the same population as the area covered by the disease register.

Overcoming observation and recall bias
Overcoming retrospective recall bias can be achieved by using data recorded, for other purposes, before the outcome had occurred and therefore before the study had started. The success of this strategy is limited by the availability and reliability of the data collected. Another technique is blinding where neither the subject nor the observer know if they are a case or control subject. Nor are they aware of the study hypothesis. In practice this is often difficult or impossible and only partial blinding is practicable. It is usually possible to blind the subjects and observers to the study hypothesis by asking spurious questions. Observers can also be easily blinded to the case or control status of the patient where the relevant observation is not of the patient themselves but a laboratory test or radiograph.

> **Key points**
>
> **Case-control studies**
> - Case-control studies are simple to organise
> - Retrospectively compare two groups
> - Aim to identify predictors of an outcome
> - Permit assessment of the influence of predictors on outcome via calculation of an odds ratio
> - Useful for hypothesis generation
> - Can only look at one outcome
> - Bias is an major problem

Blinding cases to their case or control status is usually impracticable as they already know that they have a disease or illness. Similarly observers can hardly be blinded to the presence of physical signs, for example, cyanosis or dyspnoea.

As a result of the problems of matching, bias and confounding, case-control studies, are often flawed. They are however useful for generating hypotheses. These hypotheses can then be tested more rigorously by other methods—randomised controlled trials or cohort studies.

Examples
Case-control studies are very common. They are particularly useful for studying infrequent events, for example, cot death, survival from out of hospital cardiac arrest, and toxicological emergencies.

A recent example was the study of atrial fibrillation in middle aged men during exercise.[11]

USING DATABASES FOR RESEARCH (SECONDARY DATA)
Pre-existing databases provide an excellent and convenient source of data. There are a host of such databases and the increasing archiving of information on computers means that this is an enlarging area for obtaining data. Table 3 lists some common examples of potentially useful databases.

Such databases enable vast numbers of people to be entered into a study prospectively or retrospectively. They can be used to construct a cohort, to produce a sample for a cross sectional study, or to identify people with certain conditions or outcomes and produce a sample for a case controlled study. A recent study used census data from 11 countries to look at the relation between social class and mortality in middle aged men.[12]

Advantages and disadvantages
These type of data are ordinarily collected by people other than the researcher and independently of any specific hypothesis. The opportunity for observer bias is thus diminished. The use of previously collected data is efficient and comparatively inexpensive and moreover the data are collected in a very standardised way, permitting comparisons over time and between different countries. However, because the data are collected for other purposes it may not be ideally suited to the testing of the current hypothesis, additionally it may be incomplete. This may result in sampling bias. For example, the electoral roll depends upon registration by each individual. Many homeless, mentally ill, and chronically sick people will not be registered. Similarly the notification of certain communicable diseases is a statutory responsibility for doctors in the UK: while it is probable that most cases of cholera are reported it is highly unlikely that most cases of food poisoning are.

Causes and associations
Because observational studies are not experiments (as are randomised controlled trials) it is difficult to control many external variables. In consequence when faced with a clear and significant association between some form of illness or

Table 3

Demographic and lifestyle data	Census, General Household Survey, social trends, annual reports
Morbidity data	GP morbidity statistics, communicable disease surveillance, hospital inpatient inquiry, hospital activity analysis, cancer registration
Mortality data	OPCS mortality statistics
Health services data	Immunisation data, screening levels, district and annual reports, confidential inquiry reports.
Specific databases	OPCS survey of disability, OPCS longitudinal study, abortion data, congenital abnormalities, workmens compensation data
Other	Social Security statistics, commissions of inquiry

Table 4

The strength of the association	The greater the magnitude of the association the greater the likelihood that it is causal
Consistency	If the association is observed at different times, places and by different researchers it is more credible
Specificity	The more specific the disease and the groups of people affected the greater the likelihood of causality
Temporal relation	Does the suspected cause precede the disease
Biological gradient	Is there a dose–response type relation
Biological plausibility	Do the findings fit with plausible biological and disease mechanisms
Coherence of the evidence	The cause and effect interpretation should fit with other known facts regarding the natural history and biology of the disease
Prevention	Does avoidance or removal of the cause decrease the incidence of the disease
Reasoning by analogy	Does the evidence mirror or match another cause and effect relation

cause of death and some environmental influence a judgement has to be made as to whether this is a causal link or simply an association. Table 4 outlines the points to be considered when making this judgement.[13]

None of these judgements can provide indisputable evidence of cause and effect, but taken together they do permit the investigator to answer the fundamental questions "is there any other way to explain the available evidence?" and is there any other more likely than cause and effect?"

SUMMARY
Qualitative studies can produce high quality information but all such studies can be influenced by known and unknown confounding variables. Appropriate use of observational studies permits investigation of prevalence, incidence, associations, causes, and outcomes. Where there is little evidence on a subject they are cost effective ways of producing and investigating hypotheses before larger and more expensive study designs are embarked upon. In addition they are often the only realistic choice of research methodology, particularly where a randomised controlled trial would be impractical or unethical.

Cohort studies look forwards in time by following up each subject
- Subjects are selected before the outcome of interest is observed
- They establish the sequence of events
- Numerous outcomes can be studied
- They are the best way to establish the incidence of a disease
- They are a good way to determine causes of diseases
- The principal summary statistic of cohort studies is the relative risk ratio
- If prospective, they are expensive and often take a long time for sufficient outcome events to occur to produce meaningful results

Cross sectional studies look at each subject at one point in time only

- Subjects are selected without regard to the outcome of interest
- Less expensive
- They are the best way to determine prevalence
- Quick
- The principal summary statistic of cross sectional studies is the odds ratio
- Weaker evidence of causality than cohort studies
- Inaccurate when studying rare conditions

Case-control studies look back at what has happened to each subject
- Subjects are selected specifically on the basis of the outcome of interest
- Cheap
- Efficient (small sample sizes)
- Produce odds ratios that approximate to relative risks for each variable studied
- Prone to sampling bias and retrospective analysis bias
- Only one outcome is studied

GLOSSARY OF TERMS
Bias
The inclusion of subjects or methods such that the results obtained are not truly representative of the population from which it is drawn

Blinding
The process by which the researcher and or the subject is ignorant of which intervention or exposure has occurred.

Cochrane database
An international collaborative project collating peer reviewed prospective randomised clinical trials.

Cohort
Is a component of a population identified so that one or more characteristic can be studied as it ages through time.

Confounding variable

A variable that is associated with both the exposure and outcome of interest that is not the variable being studied.

Control group

A group of people without the condition of interest, or unexposed to or not treated with the agent of interest.

False positive

A test result that suggests that the subject has a specific disease or condition when in fact the subject does not.

Incidence

Is a rate and therefore is always related either explicitly or by implication to a time period. With regard to disease it can be defined as the number of *new* cases that develop during a specified time interval.

Latency

A period of time between exposure to an agent and the development of symptoms, signs, or other evidence of changes associated with that exposure.

Matching

The process by which each case is matched with one or more controls, which have been deliberately chosen to be as similar as the test subjects in all regards other than the variable being studied.

Observational study

A study in which no intervention is made (in contrast with an experimental study). Such studies provide estimates and examine associations of events in their natural settings without recourse to experimental intervention.

Odds ratio

The ratio of the probability of an event occurring to the probability of non-occurrence. In a clinical setting this would be equivalent to the odds of a condition occurring in the exposed group divided by the odds of it occurring in the non-exposed group.

Prevalence

Is not defined by a time interval and is therefore not a rate. It may be defined as the number of cases of a disease that exist in a defined population at a specified point in time.

Randomised controlled trial

Subjects are assigned by statistically randomised methods to two or more groups. In doing so it is assumed that all variables other than the proposed intervention are evenly distributed between the groups. In this way bias is minimised.

Relative risk

This is the ratio of the probability of developing the condition if exposed to a certain variable compared with the probability if not exposed.

Response rate

The proportion of subjects who respond to either a treatment or a questionnaire.

Risk factor

A variable associated with a specific disease or outcome.

Validity—internal

The rigour with which a study has been designed and executed—that is, can the conclusion be relied upon?

Validity—external

The usefulness of the findings of a study with respect to other populations.

Variable

A value or quality that can vary between subjects and/or over time

REFERENCES

1 **Fowkes F**, Fulton P. Critical appraisal of published research: introductory guidelines. *BMJ* 1991;**302**:1136–40.
2 **Lerner DJ**, Kannel WB. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26 year follow-up of the Framingham population. *Am Heart J* 1986;**111**:383–90.
3 **Doll R**, Peto H. Mortality in relation to smoking. 40 years observation on female British doctors. *BMJ* 1989;**208**:967–73.
4 **Alberman ED**, Butler NR, Sheridan MD. Visual acuity of a national sample (1958 cohort) at 7 years. *Dev Med Child Neurol* 1971;**13**:9–14.
5 **Davey Smith G**, Hart C, Blane D, *et al*. Adverse socioeconomic conditions in childhood and cause specific mortality: prospective observational study. *BMJ* 1998;**316**: 1631–5.
6 **Goyder EC**, Goodacre SW, Botha JL, *et al*. How do individuals with diabetes use the accident and emergency department? *J Accid Emerg Med* 1997;**14**:371–4.
7 **Jaffe HW**, Bregman DJ, Selik RM. Acquired immune deficiency in the US: the first 1000 cases. *J Inf Dis* 1983;**148**:339–45.
8 **Johnstone AJ**, Zuberi SH, Scobie WH. Skull fractures in children: a population study. *J Accid Emerg Med* 1996;**13**:386–9.
9 **van der Pol V**, Rodgers H, Aitken P, *et al*. Does alcohol contribute to accident and emergency department attendance in elderly people? *J Accid Emerg Med* 1996;**13**:258–60.
10 **Reidy A**, Minassian DC, Vafadis G, *et al*. *BMJ* 1998;**316**:1643–7.
11 **Karjaleinen**, Kujala U, Kaprio J, *et al*. *BMJ* 1998;**316**:1784–5.
12 **Kunst A**, Groenhof F, Mackenbach J. *BMJ* 1998;**316**:1636–42.
13 **Hill AB**, Hill ID. *Bradford Hills principles of medical statistics.* 12th edn. London: Edward Arnold, 1991.

EMJ

Observational research methods. Research design II: cohort, cross sectional, and case-control studies

C J Mann

Updated information and services can be found at:

http://emj.bmj.com/content/20/1/54

These include:

References	This article cites 11 articles, 8 of which you can access for free at: **http://emj.bmj.com/content/20/1/54#BIBL**
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes